The University Of Sheffield.

Department Of Economics

Arbitrary Inflation in Fractional Models

Sarah Brown, Daniel Gray, William Greene, Mark N. Harris

Sheffield Economic Research Paper Series

SERPS no. 2023006

ISSN 1749-8368

28 February 2023

Arbitrary Inflation in Fractional Models

Sarah Brown[a], Daniel Gray[a,*], William Greene[b], Mark N. Harris[c]

[a] *Department of Economics, University of Sheffield, UK*
[b] *Stern Business School, New York University, USA*
[c] *School of Accounting, Economics and Finance, Curtin University, Australia*

Abstract

The analysis of subjective probabilities, that are fractional or share variables by definition, is becoming increasingly widespread in both economics and the social sciences in general. To avoid nonsensical predictions, empirical predictions for such variables must respect the fact that they are necessarily bounded on the $0-1$ (or, $0-100$, for percentage-type responses) interval. In addition, where the response variable of interest corresponds to a self-report on a fixed scale, individuals are often drawn to particular focal-point responses, resulting in distinct spikes in the empirical distribution. In this paper, we suggest a simple model that accounts for all of the nuances of such data, including its fractional and bounded nature as well as arbitrary inflation at such focal-points (which may appear at any point in the interval and are highly likely at the endpoints). We estimate our model using data drawn from the Italian Survey of Income and Wealth relating to an individual's subjective marginal propensity to consume (MPC).

Keywords: Focal-points, Fractional models, Inflated outcomes, Subjective probabilities, Marginal propensity to consume
JEL: D84; C34; E21

February 28, 2023

[*] Corresponding author

Email addresses: sarah.brown@sheffield.ac.uk (Sarah Brown), d.j.gray@sheffield.ac.uk (Daniel Gray), wgreene@stern.nyu.edu (William Greene), Mark.Harris@curtin.edu.au (Mark N. Harris)

1. Introduction and Background

Subjective probability questions are now widespread in many large secondary data sources; see, for example, the Health and Retirement Survey (HRS) and the CentER data, amongst many others.[1] Subjective probabilistic questions have been shown to contain valuable information about a wide range of domains, including individual health outcomes and survival probabilities, financial expectations and expectations about firm performance, and have been discussed from a range of disciplines, including, health, economics and political science, see for example Hurd (2009) and Altig et al. (2022) amongst many others. However, there remain significant challenges in modeling responses to these types of questions, given the problems around focal point responses, rounding and the fractional nature of the underlying distribution. This paper contributes to the growing literature related to modeling subjective probabilities, as well as making a more general contribution by proposing a new, intuitive econometric model, that captures the key artefacts of this type of data. We then apply this to an economic application, which explores an individual's subjective marginal propensity to consume. How individuals respond to subjective probability questions, and the likelihood of focal point or rounded responses, is dependent on the characteristics of the respondent. When answering a survey question, Schwarz and Oyserman (2001) identify five distinct steps in the response process and argue that increasing the difficulty of one or more of these steps can increase the probability of focal point answers or non-response.[2] Moreover, Manski and Molinari (2010) suggest that a rounded response may capture partial knowledge or represents an attempt to simplify a response. Generally, responses to subjective probability questions are grouped at multiples of 5 or 10 (depending on the question), with a large number of responses at the extremes of the distribution (0% and 100%) in addition to 50%. As argued by Hurd (2009), 50% could capture uncertain responses, rather than individuals responding a true 50%. Individuals who fail to understand the question (or the concept of probability) may tend to choose the middle of the scale resulting in a disproportionate number of responses at this point (de Bruin et al., 2000). Several studies, predominantly analysing the HRS data, emphasize the large number of focal answers and show that the likelihood of such answers is correlated with education and cognitive measures (Hurd et al., 1998).

As a result of these artefacts of the data, the modeling of subjective probability questions has received increasing attention from the econometrics literature. For example, Heiss et al. (2022) exploit a panel finite mixture model that allows for rounding in the responses and unobserved heterogeneity

[1] The HRS asks respondents to give subjective probabilities about the likelihood of living to age 75 and to age 85, leaving a bequest of up to $10,000 and $100,000, receiving an inheritance, working after the ages of 62, 65 and 70 and having a health problem in the next 10 years. Likewise, the CentER data contains subjective probabilities, for example, relating to income expectations and the likelihood of job loss.

[2] The steps they identify are namely: interpreting the question; recalling behaviour; inferring an answer; mapping the answer onto the response format; and editing.

in stock market expectations, whilst Giustinelli et al. (2022) focus on tail and center rounding in individual subjective probability questions relating to a range of outcomes contained in the HRS data, as outlined previously. Further, Kleinjans and Soest (2014) explore subjective probabilities in the HRS allowing for non-response and focal point responses and de Bresser and van Soest (2013) develop a panel data model to explore expectations captured on a percentage scale that incorporates non-response, rounding and focal answers. This model is then applied to a representative sample of Dutch employees and exploits data relating to expectations of the retirement income replacement rate. These studies have attempted to capture the various characteristics of the data, including rounding and focal points, in different ways. Ignoring the inherent properties of such response variables is likely to produce mis-leading results and poor model fit.

We contribute to the existing literature by suggesting a simple approach that accounts for all of the observed nuances of this type of data, including its fractional and bounded nature and also for arbitrary inflation at such focal points (which may appear at any point in the interval, but are particularly likely to encompass the endpoints).[3] We also show how our cross-sectional approach can be easily adapted to a panel data setting and outline the wide range of potential quantities of interest that can be calculated *ex post*, as well as relevant model selection and testing techniques.

The econometric model is illustrated with an application to individual unit survey responses on individuals' marginal propensity to consume, which have been previously analysed in the macroeconomic and public policy literature (Jappelli and Pistaferri, 2014, 2020). Using a wide range of model selection metrics, our newly developed approach out-performs all of the 'competing' models considered, and the summary predicted measures and distributions of our new approach very closely mimic those of the observed sample, in contrast to the other methods considered. Given the rise in the availability of, as well as the interest in, such response variables, we believe that this model will be of widespread use across the social sciences.

The remainder of the paper is structured as follows. Section 2 outlines the newly developed model and model selection criteria, whilst Section 3 outlines various estimates of interest. Section 4 presents the empirical application and finally Section 5 concludes.

2. Methods

2.1. Rounding and Focal Points

A starting point for our approach is Greene et al. (2015), who consider self-reports of self-assessed health (SAH). An initial investigation revealed that the distribution of these responses on a 5-point *likert*-scale appeared to be inflated in the *good* and *very good* outcomes, especially in relation to more

[3] Arbitrary inflation refers to the fact that any outcome in the distribution could be inflated.

objective measures of population health. The approach taken to model SAH, following most of the literature on such 'inflation' models, was a latent class/partial observability one, whereby individuals are 'first' partitioned into two unobserved classes, which are termed 'accurate' and 'inaccurate' respondents. The reason for such an approach, is that the former would be free to choose any outcome on the choice scale, whereas the latter would be drawn, for whatever reason, to any of the hypothesised inflated outcomes.

Following this general approach, let *type* denote an unobserved binary variable indicating the split between the two regimes, $\mathit{type} = I, A$ (inaccurate and accurate, respectively), where *type* is related to the latent variable type^* via the usual mapping: $\mathit{type} = A$ for $\mathit{type}^* > 0$ and $\mathit{type} = I$ for $\mathit{type}^* \leq 0$. As usual, the propensity equation is given by

$$\mathit{type}^* = \mathbf{x}'\beta + \varepsilon, \tag{1}$$

where $\mathbf{x}$ is a k_x vector of covariates, β a vector of unknown coefficients, and ε a (standard-normally distributed) error term.[4]

For the inflation component of the model, this will involve the joint probability of $\mathit{type} = I$ and the particular inflated outcome. Our focus here is primarily on fractional, or share, models such as subjective probabilities. As such variables inherently embody cardinality, any inflation points lying on this scale, in our example, these are predominantly 0, 0.5 and 1, similarly inherit this trait (75% is 25pp greater than 50%); or alternatively could be treated simply as an ordered variable. In light of this, an ordered discrete model would appear to be appropriate here (or a binary model, if there were only two hypothesised inflation outcomes). Two obvious examples here would be an ordered probit (OP) model, with estimated cut-points, or an interval regression with (IR) fixed cut-points (Greene and Hensher, 2010). For ease of exposition, and to facilitate the notation, we will for now assume just three inflation points at $\widetilde{y} = 0, 0.5, 1$, although the extension to more is straightforward. Here, the usual OP/IR approach would be driven by an underlying stochastic latent variable, $\widetilde{y}^*$, of the form

$$\widetilde{y}^* = \mathbf{z}'\gamma + u, \tag{2}$$

with $\mathbf{z}$ being a k_z vector of explanatory variables (with no constant) with unknown weights γ, and u

[4] Note that following the related literature (see Section 2.5), if present in the data, the *type* variable can be expanded to include a 'don't know/unanswered' outcome, to lessen the chances of any endogenous estimation sample selection.

a (standard normal) error term, with the usual mapping of

$$\widetilde{y} = \begin{cases} 0 & \text{if } \widetilde{y}^* \leq \mu_0, \\ 0.5 & \text{if } \mu_0 < \widetilde{y}^* \leq \mu_1, \\ 1 & \text{if } \mu_1 \leq \widetilde{y}^*, \end{cases} \tag{3}$$

where μ are the usual cut-points, either freely estimated (OP) or fixed (0.25, 0.75; IR).[5] However, the flexibility that the former affords would appear to make it preferable as boundaries are free to move around to improve overall model fit.

Note that under the usual assumptions of normality, we have

$$P(type = A|\mathbf{x}) = \Phi(\mathbf{x}'\beta), \tag{4}$$

and

$$\Pr(\widetilde{y}|\mathbf{z}) = \begin{cases} \Pr\left(\widetilde{y} = 0 \left| \mathbf{z}, r = 1\right.\right) = \Phi\left(\mu_0 - \mathbf{z}'\gamma\right) = P(l^i) \\ \Pr\left(\widetilde{y} = 0.5 \left| \mathbf{z}, r = 1\right.\right) = \Phi\left(\mu_1 - \mathbf{z}'\gamma\right) - \Phi\left(\mu_0 - \mathbf{z}'\gamma\right) = P(m^i) \\ \Pr\left(\widetilde{y} = 1 \left| \mathbf{z}, r = 1\right.\right) = \Phi\left(\mathbf{z}'\gamma - \mu_1\right) = P(u^i), \end{cases} \tag{5}$$

where Φ denotes the standard normal cumulative distribution function. Note that we use the terms l^i, m^i, u^i to denote the lower-, mid- and upper-inflation points, respectively. Note that whilst probabilities throughout are conditioned on covariates, for example see equations (4) and (5), for ease of exposition we subsequently omit the conditioning.

Equations (4)-(5) form the basis of our proposed inflation approach. We note that there is a suite of *hurdle* and *double-hurdle* models, developed to address the build-up of '*zero*' observations, where the response variable is a continuous variable, but with only a single inflation point; that is, typically with a non-zero probability mass at zero (Cragg, 1971; Jones, 1989; Smith, 2003). Methods have also been developed to similarly account for excess zeros in count data variables (Mullahy, 1986; Heilbron, 1989; Lambert, 1992; Greene, 1994; Pohlmeier and Ulrich, 1995; Mullahy, 1997) and finally, building on these developments, there has been a recent rise in the development of so-called inflated-OP models (Harris and Zhao, 2007; Bagozzi and Mukherjee, 2012; Brooks et al., 2012; Brown et al., 2020; Sirchenko, 2020). However, an important gap in this literature lies in the growing instances of fractional, or share, continuous variables with arbitrary inflation. We suggest an appropriate approach, couched in the general set-up described above, which will simultaneously account for arbitrary inflation as well as respecting the $0-1$ (or $0-100$ if expressed as a percentage) nature of the response variable.

[5] Note that if an IR approach is taken, one can now identify the scale of u; Greene and Hensher (2010).

2.2. A Poisson-based Approach

The Poisson *pseudo*-maximum likelihood ($PPML$) estimator has long been put forward as a robust, and simple, approach for data characterised by $y \geq 0$, with potentially excess zeros (Santos Silva and Tenreyro, 2006, 2011; Motta, 2019, for example). Thus, the $PPML$ approach appears well-suited for our purposes, and moreover, the issue of non-integers can be straightforwardly handled by a simple scaling of the original fractional data. For example, in our application, where fractional responses are recorded to 2 decimal places, this means that $0 \leq y \times 100 \leq 100$ and where all $(y \times 100)$ transformed variables are integers.

Although the $PPML$/Poisson estimator is defined for $y \geq 0$, we also require $y \leq 1(100)$. However, it is easy to consider a *top-coded/censored* Poisson (TCP) version (Terza, 1985). The usual Poisson density is given by

$$f(y;\lambda) = \frac{\exp(-\lambda)\,\lambda^{y}}{y!} = f(P), \quad y = 0, 1, \ldots \tag{6}$$

with

$$\lambda = \exp\left(\mathbf{w}'\delta\right), \tag{7}$$

where $\mathbf{w}$ is a vector of covariates driving the mean of the Poisson process, λ. Note that throughout we use $f(P)$ to denote the (standard) Poisson density.

Defining the upper censoring point as u^{i}, where here $u^{i} = 100$ (for the re-scaled response variable), then to ensure that the 'counts' are appropriately bounded, define an indicator $d^{u^{i}}$ as

$$d^{u^{i}} = \begin{cases} 1 \text{ if } y = u^{i} \\ 0 \text{ otherwise,} \end{cases} \tag{8}$$

then the top-censored Poisson density is

$$f^{*}(y;\lambda,u^{i}) = \{f(P)\}^{(1-d^{u^{i}})}\{1 - F(u^{i}-1)\}^{d^{u^{i}}}, \tag{9}$$

where $1 - F(u^{i}-1) = 1 - \sum_{j=0}^{u^{i}-1} f(j) = f(P^{u^{i}})$; that is, 1 minus the sum of all preceding count probabilities $(0 - 99)$.

Our conjecture is that such an approach will be of most use for response variables in the $[0, 1]$ range. Thus, this necessitates evaluation of the Poisson-type probabilities of the form presented in equation (9). Due to the presence of $y!$ inherent in these, it may be necessary to avoid the likely resultant numerical overflow issues by consideration of the *log* of these probabilities, as it is these which then directly enter the log-likelihood function as,

$$lnP(y) = -V + ylnV - ln(y!), \quad y = 0, 1, \ldots u^{i} - 1 \tag{10}$$

where $V = exp(\mathbf{w}'\delta)$ and $ln(y!)$ is evaluated using Stirling's formula. Similarly to avoid overflows, the TCP probability for the top-censored value of u^i can be computed using the regularised lower incomplete gamma function such that

$$lnP(y = u^i) = \int_0^V \frac{exp^{-t} t^{(u^i - 1)}}{\Gamma(u^i)} dt. \tag{11}$$

We now take the top-censored $PPML$ estimator as a starting point (or the TCP model) and combine with equations (4)-(5) to allow for the necessary inflation processes. The overall likelihood here will consist of four separate components: one each for all of the inflated outcomes ($y = 0, 50, 100$) and then one for the remaining observed outcomes. Maintaining the assumption that the two unobserved classes described above, correspond to 'accurate' and 'inaccurate' respondents, then the latter can freely choose any observed outcome, conditional on $type = A$. By defining an indicator, d^{n^i} (for 'non-inflated') such that $d^{n^i} = 1 \times [y \neq l^i, m^i, u^i]$, then the density for these non-inflated observations, $f(d^{n^i})$, is simply the usual TCP of equation (9) weighted by $Pr(type = A)$, such that

$$f(d^{n^i}) = [\Phi(\mathbf{x}'\beta) f^*(P)]. \tag{12}$$

Next, consider the lower-inflated observations, such that $y = l^i = 0$. For these, their density is a combination of that corresponding to the inflation process generating the l^i's with that from the usual count process generating the same. Define an indicator for these observations as $d^{l^i} = 1 \times [y = l^i]$, then $f(d^{l^i})$ will be given by

$$f(d^{l^i}) = [\Phi(\mathbf{x}'\beta) f^*(P) + \Phi(-\mathbf{x}'\beta) P(l^i)], \tag{13}$$

recalling that $P(l^i)$ is the OP probability of the lower inflation point from equation (5).

Analogously, define $d^{m^i} = 1 \times [y = m^i]$ as an indicator for the middle-inflated observations, their density will be given by

$$f(d^{m^i}) = [\Phi(\mathbf{x}'\beta) f^*(P) + \Phi(-\mathbf{x}'\beta) P(m^i)], \tag{14}$$

where $P(m^i)$ is from equation (5). Finally, for the top-inflated observations, we have

$$f(d^{u^i}) = [\Phi(\mathbf{x}'\beta) f^*(P^{u^i}) + \Phi(-\mathbf{x}'\beta) P(u^i)] \tag{15}$$

recalling that $f(P^{u^i})$ is the TCP density for the upper censored observations.

Given these components, and noting that $d^{n^i} + d^{l^i} + d^{m^i} + d^{u^i} \equiv 1$, then the full likelihood for an individual is

$$\mathcal{L}_i = [f_i(d^{n^i}) \times d^{n^i}] + [f_i(d^{l^i}) \times d^{l^i}] + [f_i(d^{m^i}) \times d^{m^i}] + [f_i(d^{u^i}) \times d^{u^i}], \tag{16}$$

which can be maximised in the usual fashion to estimate all the parameters of the model. We term this model the fractional inflated-*pseudo* Poisson ($FIPP$) model.

Note that implicitly independence has been assumed across all of the inherent components of the $FIPP$ model which may, or may not, be a tenable assumption. We next move on to relaxing this assumption.

2.3. Extensions to Panel Data

As in our empirical example presented in Section 4, researchers will often have panel data to hand. The approach outlined above can be straightforwardly extended to allow for such, and thereby allow for the presence of unobserved individual heterogeneity in each equation. As is usual in the panel data data literature, we can simply augment equations (1), (2) and (7), respectively, as

$$type^*_{it} = \mathbf{x}'_{it}\beta + \alpha_i^{type} + \varepsilon_{it}, \tag{17}$$

$$\widetilde{y}^*_{it} = \mathbf{z}'_{it}\gamma + \alpha_i^{y} + u_{it} \text{ and} \tag{18}$$

$$\lambda_{it} = \exp\left(\mathbf{w}'_{it}\delta + \alpha_i^{\lambda}\right), \tag{19}$$

where α^j represent the $j = 1, 2, 3$ unobserved effects for the three respective equations. Due to the usual *incidental parameters problem* (Neyman and Scott, 1948) invariably it would appear appropriate to treat these as random draws from a multivariate distribution, and, as is common, we will assume normality (Greene, 2018), such that

$$\alpha \sim MVN(\mathbf{0}, \mathbf{\Omega}) \text{ where} \tag{20}$$

$$\mathbf{\Omega} = \begin{pmatrix} \sigma^2_{type} & \sigma_{type,y} & \sigma_{type,\lambda} \\ \sigma^2_{y,type} & \sigma^2_{y} & \sigma_{y,\lambda} \\ \sigma_{\lambda,type} & \sigma_{\lambda,y} & \sigma^2_{\lambda} \end{pmatrix}. \tag{21}$$

The model implies certain restrictions regarding Ω. Individuals are implicitly split by the *type* equation, $\sigma_{y,\lambda} = \sigma_{\lambda,y} = 0$ and, moreover, Ω is symmetric, such that $\sigma^2_{y,type} = \sigma^2_{type,y}$ and so on.

This addition is advantageous as it allows one to correlate the appropriate equations, whereas it is not obvious how to achieve this across all equations in the cross-sectional setting. Note that whilst all covariates are *it* indexed, there is no restriction that this needs to be the case, and all could vary in any/all dimension(s). Note that equation (17) is the most flexible approach in the context of panel data: individuals are able to change *type* over time; the restricted version would only accommodate heterogeneity across individuals, but not over time, such that *types* do not change over time.[6]

On the other hand, such an addition does complicate estimation. We evaluate the (log-)likelihood equation via simulation techniques using 500 Halton draws. Define $\mathbf{v}_i$ as a vector of standard normal random variates, which enter the model generically as $\omega_i = \mathbf{\Gamma v}_i$, such that for a single draw of $\mathbf{v}_i$, $\omega_\mathbf{i} = \left(\alpha_i^{type}, \alpha_i^y, \alpha_i^\lambda\right)$; where $\mathbf{\Gamma} = chol\left(\Sigma\right)$ such that $\Sigma = \mathbf{\Gamma\Gamma}'$. Conditioned on $\mathbf{v}_i$, the sequence of T_i outcomes for household i are independent, such that the contribution to the likelihood function for a group of t observations is defined as the product of the sequence $\mathcal{L}_{it}$; equation (16). The unconditional log-likelihood is found by integrating out these innovations such that

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{N} \log \int_{\omega_\mathbf{i}} \prod_{t=1}^{T_i} \left(\mathcal{L}_{it} \mid \omega_\mathbf{i}\right) f(\omega_\mathbf{i}) d\omega_i, \tag{22}$$

where all parameters of the model are contained in θ. Since $\mathbf{v}_i$ is a vector of independent standard normal variables, where $\phi(.)$ denotes the standard normal probability density function, the joint density is similarly the product of standard normals, yielding

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{N} \log \int_{\mathbf{v_i}} \prod_{t=1}^{T_i} \left(\mathcal{L}_{it} \mid \omega_\mathbf{i}\right) \prod_{k=1}^{K} \phi(v_{ik}) dv_{ik}, \tag{23}$$

where $k = 1, 2, \ldots, K$ indexes the various stochastic unobserved effects in the model. It is possible to evaluate the expected values in the integrals by simulation. In practice this involves drawing $r = 1, \ldots, R$ variates of $\mathbf{v}_i$ from the multivariate standard normal population and the simulated log likelihood function is constructed as follows

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{N} \log \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} \left(\mathcal{L}_{it} \mid \omega_\mathbf{ir}\right). \tag{24}$$

Note that any concerns of correlations between these unobserved effects and observed heterogeneity, can be handled by the usual inclusion of ‘Mundlak’ variables (Mundlak, 1978).

2.4. Model Selection and Testing Issues

The suggested approach does not (obviously) nest alternative approaches in the traditional sense of parameter restrictions. Moreover, this is similarly true across different choices of the hypothesised inflation-points. An obvious tool here appears to be information criteria (*IC*). In particular, the

[6] We do not entertain such a restricted model in our empirical application, as there is a relatively long time between waves, such that it would appear overly restrictive to not allow individuals to change ‘types’ over such a long period. We also note that if individuals do not change types over time, the restricted approach is likely to better identify the types.

Bayesian Information Criterion (BIC/SC) (Schwarz, 1978), appears appropriate given that it has been shown to be the preferred criterion across a wide range of scenarios (see, for example, Gannon et al., 2014) and can also be shown to be consistent in the sense that $\Pr\left(M^{true}\right) \to 1$ as $N \to \infty$, where M^{true} is the true model. In addition to IC metric(s), the $Vuong$ test for non-nested models can also be used (Vuong, 1989). Due to the potential large differences in model sizes across different potential competitor models, we suggest using the 'BIC' correction factor in this, as proposed in Vuong (1989).

The standard $Vuong$ test (for example, Greene, 2018) for comparing two competing models $j = 1, 2$ is based on m_i the individual differences in the two log-likelihoods, such that

$$m_i = ln\left(\frac{f_1\left(y_i|\tilde{x}_i\right)}{f_2\left(y_i|\tilde{x}_i\right)}\right), \tag{25}$$

where f_j are the respective likelihoods from the two $j = 1, 2$ competing models; the $Voung$ test is then

$$V = \frac{\sqrt{n}\bar{m}}{s_m}, \tag{26}$$

where n is the sample size and $\bar{m}$ and s_m are the sample average and standard deviation of m_i, respectively. The test has a limiting standard normal distribution, with values of $|V| < 1.96$ being indeterminate, whereas large positive (negative) values favour model 1 (2). As noted, we propose the BIC corrected version of this test (Vuong, 1989), which is simply

$$m_i^c = m_i + \left(k_2 - k_1\right)\frac{ln\left(n\right)}{2n}, \tag{27}$$

where k_j refers to the number of estimated parameters in model j. With a number of competing models, as opposed to a strict pairwise comparison, we follow the suggestion of Durand et al. (2022) in that an appropriate model selection metric amongst these is that model with the most favoured number of pairwise selections.

2.5. Comparison with Previous Literature

As noted above, this approach most closely follows that of Greene et al. (2015); however, it is also related to other streams of literature that are concerned with both misclassification of the dependent variable (see, for example, Hausman et al., 1998), partial observability (for example, Poirier, 1980) and so-called 'inflation' models (for example, Harris and Zhao, 2007). Similarly, Farrell et al. (2011) consider the impact of cigarette pack-sizes leading to observed spikes at these counts, on the daily number of the cigarettes consumed by smokers.

Another closely related paper is Kleinjans and Soest (2014), who also consider the modeling of subjective probabilities. Their approach, which builds on the so-called 'rounding' methods of Heitjan and Rubin (1990), involves a combination of three equations, specifically probabilistic models of: a) true, but unobserved, responses (the actual statistical model here is not stated, although it appears to be of an *interval regression* form); b) an ordered probit model for rounding processes in multiples of 1, 5, 10, 25 and 50 (as only integer values are considered, all responses have to be rounded to at least a multiple of 1); and c) a multinomial-type approach for the choice between rounding, item non-response (INR) and 'complete uncertainty', or 'focal answers' (which they define as 50%). Based on their underlying assumptions, the final model they estimate is a very highly specified one, with essentially separate models/specifications for *all* observed choice outcomes. In turn, this paper is closely related to Hudomiet et al. (2011), but additionally incorporating focal answers and item non-response (or 'don't knows').

Compared to the approach of Kleinjans and Soest (2014), the MPC data used in our application (see Section 4 below) has no INR. However, our approach can easily be adapted to such by augmenting the 'first stage' binary Probit equation for accurate/inaccurate (rounded) response - equation (4) - to additionally include an outcome relating to INR, and the appropriate statistical model, following Kleinjans and Soest (2014), would be of the multinomial logit (MNL) form. By specifying an OP model for the inflated outcome variables, with flexible boundary points, our suggested approach also explicitly allows for rounding between these several 'focal point' observations. Moreover, this approach does not constrain equal Euclidean distance between such neighbouring points.

Note that we do not consider an augmented approach for 'complete uncertainty' as in Kleinjans and Soest (2014), (although, again our 'first-stage' equation could easily be augmented to account for such), as it seems asymmetrical to treat this single outcome differentially, as it appears rather arbitrary to pick the 50% observation as this single 'focal point'. In our set-up, as opposed to Kleinjans and Soest (2014), the approach can explicitly, probabilistically, identify the stylised individuals who tend to favour the inflated outcomes - which, dependent upon the particular application, could be extremely valuable for policy making. Finally, the current set-up is much simpler than that of Kleinjans and Soest (2014); for example in the latter, 'true' responses require explicit specification of every single observed outcome $y_i \in (0, 100)$, whereas in the former, all that is required is the generic Poisson form of equation (6), a simple function of y_i. The simplicity of the current approach here, is evident in a comparison of equation (16), with the likelihood of Kleinjans and Soest (2014), as given by their equation (8).

3. *Ex Post* Quantities of Interest

Post estimation, several quantities of interest are available. Note that for the panel variants, these can either be computed at $E(\alpha) = \mathbf{0}$, or the same draws used as in estimation.[7]

[7] The latter would appear to be preferable and is indeed, the approach followed below.

3.1. Probabilities

To account for the arbitrary (in the sense of any outcome can be such) item inflation, all models encompass the respondent type equation; equation (4). Post estimation, with $\hat{\beta}$ in hand, these can be straightforwardly computed as either $\Phi(\bar{x_i}'\hat{\beta})$ or $\overline{\Phi(x_i'\hat{\beta})}$, and standard errors obtained by the usual Delta method. Depending on the application at hand, these could be extremely policy relevant, as they give an aggregate estimate of the scale of such reporting effects in the data.

The unconditional empirical distribution of a fractional variable of interest with multiple, and potentially arbitrary, inflation is likely to be very non-standard (for example, as is evidenced by our empirical example below). Therefore, an appropriate visual metric by which to determine the performance of such inflated approaches would appear to be a comparison of actual *versus* estimated densities. In light of this, we suggest a comparison of the sample proportions of the observed outcomes, compared to the $FIPP$ probabilities for $j = 0, \dots J$ of the form:

$$P_{ij}^{IPP} = f_i(d^{n^i}) + f_i(d^{l^i}) + f_i(d^{m^i}) + f_i(d^{u^i}), \tag{28}$$

averaged over individuals.

Again, with all relevant parameter estimates in-hand, it is also possible to compute all probabilistic elements of the likelihood function of equation (16). For example, we can evaluate the TCP in isolation: that is, ‘purged’ of inflation/reporting effects; or the marginal inflation probabilities corresponding to the OP probabilities evaluated in isolation; or the joint probabilities of these along with the accurate reporting probabilities.

Following the more traditional latent class literature (see, for example, Greene, 2018), we can also consider *posterior* probabilities, which here essentially answer the question: given all of the observed data, what is the probability that any particular observation lies in a particular respondent-type class. For example, the posterior probability of an accurate respondent ($type = A$) will be

$$P\left(type_i = A|x_i, w_i, y_i\right) = \frac{f\left(y_i|type = A, w_i\right) P\left(type_i = A|x_i\right)}{f\left(y_i\right) = \mathcal{L}_i}. \tag{29}$$

Definitionally, the posterior probability of an inaccurate respondent ($type = I$) will be $1 - P\left(type_i = A|x_i, w_i, y_i\right)$, and will be the sum of the individual posteriors for each of the inflated outcomes.

3.2. Expected Values

Several expected values (EVs) can be considered, and can be based on the underlying EVs of the parent model. Thus, here these will be (for the top-censored Poisson model):

$$EV(TCP) = u^i - \sum_{j=0}^{u^i-1} f(j)(u^i - j). \tag{30}$$

We consider two forms of EVs: *purged* and *overall* (or probability-weighted). The former simply essentially evaluates the above expressions based on the estimated parameters from the full model. In essence, these can be regarded as $\widehat{EV}$s purged of any reporting effects.

Overall EVs, on the other hand, take into account all of the reporting effects. Generically, these will be given by

$$EV(overall) = P(A)EV(purged) + P(I)P(l^i)l^i + P(I)P(m^i)m^i + P(I)P(u^i)u^i. \tag{31}$$

However, several candidates are available for both $P(A)$ (or $P(I)$) and $P(l^i, m^i, u^i)$. These could correspond to either prior or posterior probabilities and/or predicted 0/1 outcomes, based on the maximum probability rule.

A final form of EV we consider, is to explicitly combine the joint *quasi*-continuous EVs of the underlying Poisson form, with the discrete ones from the inflation processes along with the aforementioned posterior probabilities. Explicitly, we estimate posterior probabilities for all of the inflated outcomes, and of overall ‘accurate’ reporting. Using the usual 0.5 cut-off rule (based on these posterior probabilities), individuals predicted to be ‘accurate’ reporters, were assigned their $TCP - EV$s, whereas the ‘inaccurate’ reporters were assigned the EV corresponding to the maximum probability of the inflated outcomes.

3.3. Partial Effects

Once the probabilistic and EV quantities have been defined, it is straightforward to evaluate partial effects of all, or a subsection, of these, by differentiating these with respect to covariates of interest. Firstly, define $\tilde{\beta}$, $\tilde{\gamma}$ and $\tilde{\delta}$ as the zero-inflated counterparts to β, γ and δ, respectively, to ensure equal parameter vector length, the analytical partial effects of $EV(overall)$, equation (31) are

$$\begin{aligned}\frac{\partial EV(y|\mathbf{x},\mathbf{z},\mathbf{w})}{\partial(\mathbf{x},\mathbf{z},\mathbf{w})} &= \Phi(\mathbf{x}'\beta)\left[\sum_{j=0}^{u^i-1} f(j)(j-u^i)(j-\lambda)\right]\tilde{\delta} \qquad (32)\\ &+ \left[u^i - \sum_{j=0}^{u^i-1} f(j)(u^i-j)\right]\phi(\mathbf{x}'\beta)\tilde{\beta}\\ &+ \phi(\mathbf{x}'\beta)(-\tilde{\beta})\left[\sum_{inf=0}^{u^i} P(OP_j)j\right]\\ &+ \Phi(\mathbf{x}'\beta)\left[\sum_{j=0}^{u^i}\left[\phi(\mu_{j-1}-\mathbf{z}'\gamma)-\phi(\mu_j-\mathbf{z}'\gamma)\right]j\right]\tilde{\gamma}, \qquad (33)\end{aligned}$$

where $P(OP_j)$ are the OP probabilities for inflation points $j = (0, 50, 100)$, but where generalisations of different sets of such are obvious. Analytical derivatives for components of the generic $EV(overall)$ are obtained by evaluation of the relevant parts of equation (33). These analytical derivatives are also used for the Delta method in the computation of the standard errors of these quantities.

4. Empirical Application: Self-Reported Marginal Propensity to Consume

4.1. Background

Here we illustrate the above methods with an application based on data relating to the marginal propensity to consume (MPC) of Italian households. The MPC is defined as the proportion of additional household disposable income allocated to spending/consumption. The MPC is an important concept in economics, particularly macroeconomics, as it is used to calculate potential multiplier effects of fiscal injections. Recent literature has sought to explore the "response heterogeneity" of the MPC; that is, whether the consumption of different households responds differently to the same stimulus. Gelman (2021) outlines two broad views in which consumption heterogeneity arises; one relates to circumstances (*e.g.*, current financial position) and the other relates to household characteristics (*e.g.*, behavioural traits and preferences). A recent focus of this literature has been to explore how the MPC changes across the income and wealth distributions, see, for example, Kaplan and Violante (2014), Carroll et al. (2017) and Fisher et al. (2020) amongst many others. Generally, these studies find that those in the lower parts of the income and wealth distributions display a higher MPC.

To elicit an individual's MPC, we follow Jappelli and Pistaferri (2014) and Jappelli and Pistaferri (2020), who analyse self-reported consumption propensities of Italian households, whereby the survey asks households hypothetical questions relating to the percentage they would spend or save from a cash windfall (for example from a lottery). More recently, there has been a growing body of work that elicits quantitative spending responses from similar survey questions, see, for example, Graziani et al. (2016), Christelis et al. (2019) and Ameriks et al. (2020).[8]

4.2. The Data

We draw on longitudinal data from the Italian Survey of Income and Wealth (SHIW), which is a representative sample of approximately 8,000 Italian households. The SHIW contains comprehensive data relating to Italian households' assets and income, in addition to a broad range of demographic and socio-economic characteristics. This data has been extensively used in the area of household finance, see for example, Jappelli and Pistaferri (2014), Paiella and Pistaferri (2017) and Guiso et al. (1992) amongst many others. In our analysis, we focus on the 2010 and 2016 waves of the survey and on the household head, that is, the person most knowledgeable about the family's finances.[9]

Importantly for our application, the 2010 and 2016 waves of the SHIW contain a question which captures the household's subjective MPC; that is, it asks the household head, or the person most knowledgeable about the family's finances, how much they would consume out of an unexpected transitory income change. Specifically, the survey asks: *"Imagine you unexpectedly receive a reimbursement equal to the amount your household earns in a month. How much of it would you save and how much would you spend? Please give the percentage you would save and the percentage you would spend."* Figure 1 presents the distribution of responses for the 2010 (7,950 observations) and 2016 (7,415 observation) waves, separately, whilst Table 1 presents the associated summary statistics. From both Figure 1 and Table 1, it is clear that the distribution of the MPC is remarkably constant across the two years: for example, the average MPC in 2010 (2016) is 47.6% (46.7%), whilst it is 47.2% in the pooled sample. A key characteristic of these distributions is the mass of responses at 0%, 50% and 100%, which together account for 64.9% of responses, in addition to smaller inflation points at multiples of ten. This characteristic of the MPC distribution is in line with other subjective probability type questions, for example captured in the HRS, and informs the methodological approach taken.

In our empirical specifications, we control for a broad range of head of household and household characteristics based on Jappelli and Pistaferri (2014). Specifically, we control for a range of demographic variables including age dummy variables, gender, marital status, education, family size, in addition to city size and residence in the South of the country. We also control for household income and wealth via quintile dummy variables, in addition to indicators for holding positive debt and for being unemployed. Summary statistics for the estimation sample are presented in Table 1.

[8] An alternative method of recovering the MPC is to calculate the changes in income and consumption between pairs of periods over time, allowing the estimation of the Euler equation for how consumption changes with respect to changes in income, see, for example, inter alia Oh and Reis (2012) and Baker (2018).

[9] In our analysis, we analyse the two waves both pooled and as panel, and determine the preferred model using a range of model specification tests.

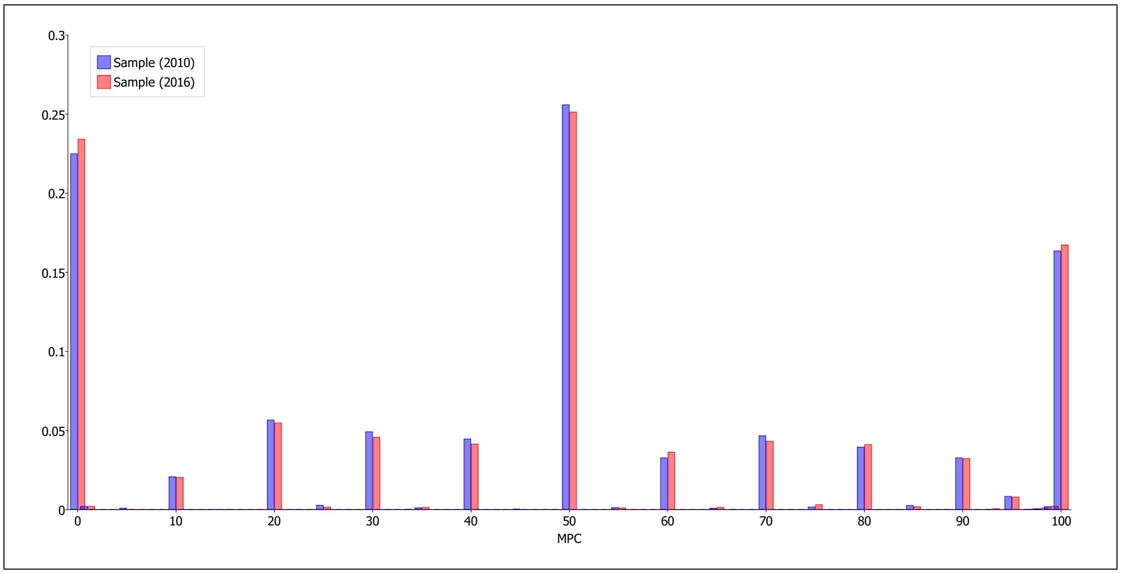

Figure 1: Distribution of sample proportions by year

In order to identify the model, we include a range of exclusion restrictions in the *accurate reporting equation*. Specifically, we include five variables which reflect aspects relating to the interviewer's perceptions about the interviewee's understanding about the interview process and their financial position (relating to both financial assets and income). We also include the duration of the interview and the individual's level of financial literacy, as captured by an index, which is the sum of correct answers to three questions relating to inflation, interest rates and risk. The summary statistics relating to these five variables are presented in Table 1.

4.3. Results

Initially, we discuss the performance of the $FIPP$ model relative to a range of alternatives, considering a range of predicted probabilities and more formal tests of model fit. We then go on to explore the effects of the covariates on the overall MPC, in addition to a range of distinct partial effects estimated within our $FIPP$ model. Finally for reference, we compare our results to a standard Tobit specification, the common approach in the literature when exploring subjective MPC, to highlight the value of our newly developed modeling approach.

Table 2 presents the summary of the expected values and the model information criteria for a range of models, namely, censored Tobit and censored Poisson models, in addition to our $FIPP$ model with three spikes at 0, 50 and 100 (named the primary) and the full $FIPP$ model with 11 spikes (corresponding to zero and multiples of ten) for both pooled cross-section and panel variants.[10] It is clear from the BIC statistics that the full model is clearly preferred compared to the alternative

[10] For reference, we also present the coefficients relating to the primary specification with three inflation points in Table 8.

models, in both the cross-section and panel specifications, and that the panel variant of the full model is preferred to the cross-section model. With regard to the average predicted values, all the models give a good approximation of the sample average, however the implied standard deviations of the distributions are superior in the $FIPP$ models. For example, considering the predicted distributions of the $Tobit$ model and the $FIPP$ model, both give an expected value close to the sample average (with predicted values of 46.7 compared to 47.2 for the sample). However, the implied standard deviation of the predicted values is significantly different. The Tobit model gives a standard deviation of 11.1 for the prediction whilst the $FIPP$ model better captures the full range of the sample, giving a standard deviation of 34.9, compared to the sample standard deviation of 35.2. Overall, this highlights that despite the Tobit model being able to capture the expected value of the distribution relatively well, it is inferior in terms of capturing the distribution of the responses. Furthermore, Table 3 presents the pairwise $Vuong$ tests for goodness of fit for our full $FIPP$ model against nested alternatives. The test statistics reveal once again that the full panel $FIPP$ model is the preferred specification amongst the competing alternatives.

Table 4 presents the summary probabilities of the inflated points for the pooled and panel variants of the model. For brevity, we only consider the panel variant. Focusing on the "Overall" column for the panel model presented in Table 4, demonstrates that our approach is able to capture the overall inflation points observed in the underlying data. For example, considering the inflation points at 0, 50 and 100, the predicted (observed) probabilities are 0.230 (0.230), 0.253 (0.254) and 0.167 (0.166), respectively. This close prediction is also observed for the other inflation points. Moreover, Table 4 reveals the predicted probabilities of the distinct parts of the model, that is, the marginal inflation probabilities corresponding to the OP probabilities evaluated in isolation and the joint probabilities of these augmented with the accurate reporting probabilities.

Visually, Figures 2 and 3 show the estimated densities of the panel variants of the primary and full models, respectively. This highlights the close match between the sample distribution and the predicted distribution for the $FIPP$ model. Overall, this highlights the superior ability of the $FIPP$ model of being able to capture the unique features of the data, compared to the standard approaches adopted in the literature.

4.3.1. Independent Variables

Prior to exploring the partial effects, we observe that the random effects included in the model, and the covariance between them, are statistically significant at conventional levels, as presented in Table 7, further advocating our panel approach. We now turn our attention to the influence of the individual covariates in the $FIPP$ model. We initially consider the overall partial effects, and subsequently explore the effects of the covariates in the distinct parts of the model, namely, the OP

inflation model, the *Poisson* model and the accurate (non-focal point) reporting model.[11] The overall partial effects presented in Table 5 reveal, in line with Jappelli and Pistaferri (2014, 2020), that a range of demographic and socio-economic characteristics influence the self-reported MPC. Relative to being aged 60 years or above, younger individuals report a higher MPC. More educated individuals, as captured by the years of education, and family size are positively associated with the self-reported MPC. For example, an additional year of education is associated with a 0.2 percentage point increase in the MPC, whilst an additional family member in the household is associated with a 1.8 percentage point higher MPC. The residence of the individual, in terms of the location of the household and also the size of the city, are statistically significant determinants of the self reported MPC. Residents in the South of Italy report a higher MPC, whilst those who live in smaller towns report a lower MPC, highlighting the importance of the geographical variation in the MPC across Italy. Interestingly, there is significant heterogeneity of the MPC across the wealth and income distributions, in line with Jappelli and Pistaferri (2014, 2020). Those in the lowest quintile of the income and wealth distributions report significantly higher MPCs. For example, for income and wealth, compared to being in the top quintile, those in the lowest quintiles report a higher MPC of 7.0 and 11.4 percentage points, respectively.

A significant advantage of our modeling approach is that it allows us to explore the partial effects in each of the estimated equations, that is the accurate reporting equation, the fractional *Poisson* equation and the OP inflation equation. The partial effects relating to the accurate reporting equation are presented in Table 5. Intuitively, this captures an individual's propensity to not report a multiple of ten, i.e. the specified inflation points, and reveals those types of individual who are more likely to provide non-focal point responses. Considering the statistically significant results reveals that those who have a larger family are more likely to provide non-focal point answers, whilst those lower in the income distribution are also more likely to report non-focal answers. Those who are unemployed are also more likely to report in this manner whereas homeowners are more likely to report a focal MPC response - that is, an inflation point. Our exclusion restrictions also appear to be appropriate, with three out of the five variables having a statistically significant impact on the individual's reporting behaviour. Interviewee understanding is positively related to non-focal reporting; whilst financial understanding and the duration of the interview are positively associated with reporting a focal answer. For example, a one point increase in the financial understanding score, measured on a 10 point scale, is associated with a 0.9 percentage point lower likelihood of reporting a non-focal point response. Further, a 1% increase in the duration of the interview is associated with a 3.4 percentage point

[11] Throughout the discussion of the results we use the term focal response to capture inaccurate responses, that is multiples of 10, and non-focal responses to capture accurate responses as outlined in Section 2.

increase in the probability of reporting a focal answer.[12]

In terms of the size of these effects, the partial effects are presented in Table 5. Comparing the overall effect to the partial effect in the fractional model, the level of financial assets has a larger impact in the fractional model. For example, compared to being in the 5th quintile of the asset distribution those in the 1st quintile report a higher MPC of 20 percentage points in the fractional model compared to 11 for the overall effect. In general, the results indicate that financial assets have twice the impact for a non-focal reporter compared to the overall effect. Being resident in the South of Italy is associated with a 8 point higher MPC, compared to other regions.

Table 6 presents the partial effects of the OP inflation equation, and these are interpreted in the usual way.[13] Compared to being in the 5th quintile of the asset distribution, being in the 1st quintile is negatively associated with reporting 0 (5.9 percentage points lower), whilst it is associated with an increased likelihood of reporting an MPC of 90 and 100 by 1.3 and 1.1 percentage points, respectively. Compared to being in a large city, being resident in a city with a population of 20,00 people or less, is positively associated with reporting an MPC of 0, that is, 14.7 percentage points more likely to report 0, and inversely related to reporting the highest MPCs, for example, 3.1 and 2.6 percentage points lower for 90 and 100, respectively.

Overall the discussion above highlights the flexibility of the $FIPP$ model and the wide array of parameter estimates available in the model, which may be of particular interest depending on the application of the model. The discussion above demonstrates that different independent variables have differential impacts across the distinct parts of the model.

4.4. The Tobit Model

For comparison with the existing literature, we briefly consider the results relating to a random effects $Tobit$ model. The results presented are in line with Jappelli and Pistaferri (2014, 2020) and so are only briefly discussed. In this setting, a range of demographic and socio-economic characteristics influence the self-reported MPC and generally these results are consistent with the overall partial effects captured in the $FIPP$ model. For example, those higher in the income and wealth distributions report a lower MPC. For example, being in the 1st quintile of the wealth distribution relative to the 5th quintile is associated with a 12 percentage points higher MPC. A small number of covariates have a differential impact in terms of statistical significance compared to the $FIPP$ model, including for example, marriage and financial literacy. In addition, the magnitudes of the overall partial effects

[12] We have repeated the analysis including only the subset of statistically significant identifying variables and obtain similar results.

[13] Additional partial effects relating to the joint inflation probabilities, that is the marginal OP probabilities augmented with the accurate reporting predictions, are easily recovered from the model. However, for brevity, we omit these partial effects.

appear quite different - potentially over - or under- estimating the economic importance of some variables, for example, the effect of unemployment, on self-reported MPC. As outlined above, the key limitation of the $Tobit$ model relates to the model not capturing the underlying distribution of the individual responses.

In summary, considering the predicted probabilities, the random effects Tobit model reveals that the means are very close to the sample mean and those of the $FIPP$ model. This suggests that despite explicitly accounting for inflated responses, which clearly have an important role in the model in terms of capturing the underlying distribution, failing to account for these elements does not induce a large bias on the estimated means. This finding echos the results of Kleinjans and Soest (2014), who explore a range of subjective probability responses in the HRS. Focusing on the specific covariates, we find that signs and significance levels are very similar between the random effects $Tobit$ model and the overall partial effects in the $FIPP$ model, but some of the sizes of the effects are different across the two models. Overall, the $Tobit$ model fails to capture the more nuanced effects of the independent variables in the distinct parts of the $FIPP$ model.

5. Conclusion

This paper has developed a new method for modeling subjective probability questions, which are characterised by being bounded between 0-100 and display rounding and focal answers. Responses to such questions have been increasingly used to elicit information from individuals and are found to be predictive of a range of behaviours, including health outcomes and survival probabilities, financial expectations and expectations about firm performance. We have developed a fractional inflated-*pseudo Poisson* ($FIPP$) model, which accounts for both the inflation of specific values and the fractional nature of the responses.

The $FIPP$ model characterises two types of individuals; accurate reporters who report a refined point probability response, and inaccurate reporters who report focal answers, in our application multiples of ten. This approach allows us to recover a range of interesting and economically relevant partial effects and predicted values. Overall, this is a flexible approach that can be applied in the modeling of any subjective or self reported response data captured on the unit interval, but is ideally suited to modeling subjective probabilities.

We explore individual level data on the subjective MPC using our model and compare our approach with a $RE\ Tobit$ model, which is commonly used in the existing literature, and a censored *Poisson* model. The $RE\ Tobit$ and censored *Poisson* models, which fail to account for rounded reporting behavior, give similar partial effects at the average, in terms of the signs and significance levels to our newly developed model. In contrast, the predicted distribution of the $FIPP$ model better captures the sample distribution, in addition to revealing a more nuanced picture of the influence of the independent variables in each part of the proposed model. Overall, our results are in line with Kleinjans and Soest (2014), that is, standard models that fail to account for inflation and rounded responses, such as the *RE Tobit* model, are suitable for researchers who wish to explore the effects of covariates at the average, but this however may not be suitable for those researchers who wish to better capture the unique artefacts of the underlying sample.

References

Altig, D., Barrero, J.M., Bloom, N., Davis, S.J., Meyer, B., Parker, N., 2022. Surveying business uncertainty. Journal of Econometrics 231, 282–303.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M.D., Tonetti, C., 2020. Long-term-care utility and late-in-life saving. Journal of Political Economy 128, 2375–2451.

Bagozzi, B.E., Mukherjee, B., 2012. A mixture model for middle category inflation in ordered survey responses. Political Analysis 20, 369–386.

Baker, S.R., 2018. Debt and the response to household income shocks: Validation and application of linked financial account data. Journal of Political Economy 126, 1504–1557.

de Bresser, J., van Soest, A., 2013. Survey response in probabilistic questions and its impact on inference. Journal of Economic Behavior and Organization 96, 65–84.

Brooks, R., Harris, M.N., Spencer, C., 2012. Inflated ordered outcomes. Economics Letters 117, 683–686.

Brown, S., Harris, M., Spencer, C., 2020. Modelling category inflation with multiple inflation processes: Estimation, specification and testing1. Oxford Bulletin of Economics and Statistics 82, 1342–1361.

de Bruin, W.B., Fischhoff, B., Millstein, S.G., Halpern-Felsher, B.L., 2000. Verbal and numerical expressions of probability:"It's a fifty–fifty chance". Organizational Behavior and Human Decision Processes 81, 115–131.

Carroll, C., Slacalek, J., Tokuoka, K., White, M.N., 2017. The distribution of wealth and the marginal propensity to consume. Quantitative Economics 8, 977–1020.

Christelis, D., Georgarakos, D., Jappelli, T., Pistaferri, L., Van Rooij, M., 2019. Asymmetric consumption effects of transitory income shocks. The Economic Journal 129, 2322–2341.

Cragg, J., 1971. Some statistical models for limited dependent variables with applicationto the demand for durable goods. Econometrica 39, 829–44.

Durand, R., Greene, W., Harris, M., Khoo, J., 2022. Heterogeneity in speed of adjustment using finite mixture models. Economic Modelling 107, 105713.

Farrell, L., Fry, T., Harris, M., 2011. A pack a day for 20 years: Smoking and cigarette pack sizes. Applied Economics 43, 2833–2842.

Fisher, J.D., Johnson, D.S., Smeeding, T.M., Thompson, J.P., 2020. Estimating the marginal propensity to consume using the distributions of income, consumption, and wealth. Journal of Macroeconomics 65, 103218.

Gannon, B., Harris, D., Harris, M., 2014. Threshold effects in nonlinear models with an application to the social capital-retirement-health relationship. Health Economics 23, 1072–1083.

Gelman, M., 2021. What drives heterogeneity in the marginal propensity to consume? Temporary shocks vs persistent characteristics. Journal of Monetary Economics 117, 521–542.

Giustinelli, P., Manski, C.F., Molinari, F., 2022. Tail and center rounding of probabilistic expectations in the Health and Retirement Study. Journal of Econometrics 231, 265–281.

Graziani, G., Van Der Klaauw, W., Zafar, B., 2016. Workers' spending response to the 2011 payroll tax cuts. American Economic Journal: Economic Policy 8, 124–59.

Greene, W., 1994. Accounting for Excess Zeros and Sample Selection in Poisson and NegativeBinomial Regression Models. Working Paper EC-94-10. Stern School of Business, New York University. Stern School of Business, New York University.

Greene, W., 2018. Econometric Analysis 8e. Eighth ed., Pearson, New Jersey, USA.

Greene, W., Harris, M., Hollingsworth, B., 2015. Inflated responses in measures of self-assessed health. American Journal of Health Economics 1, 461–493.

Greene, W., Hensher, D., 2010. Modeling Ordered Choices. Cambridge University Press.

Guiso, L., Jappelli, T., Terlizzese, D., 1992. Earnings uncertainty and precautionary saving. Journal of Monetary Economics 30, 307–337.

Harris, M., Zhao, X., 2007. A zero-inflated ordered probit model, with an application to modelling tobacco consumption. Journal of Econometrics 141, 1073–1099.

Hausman, J.A., Abrevaya, J., Scott-Morton, F.M., 1998. Misclassification of the dependent variable in a discrete-response setting. Journal of Econometrics 87, 239–269.

Heilbron, D., 1989. Generalized Linear Models for Altered Zero Probabilities and Overdispersionin Count Data. Technical Report. University of California. University of California, San Francisco.

Heiss, F., Hurd, M., van Rooij, M., Rossmann, T., Winter, J., 2022. Dynamics and heterogeneity of subjective stock market expectations. Journal of Econometrics 231, 213–231.

Heitjan, D.F., Rubin, D.B., 1990. Inference from coarse data via multiple imputation with application to age heaping. Journal of the American Statistical Association 85, 304–314.

Hudomiet, P., Kézdi, G., Willis, R.J., 2011. Stock market crash and expectations of American households. Journal of Applied Econometrics 26, 393–415.

Hurd, M.D., 2009. Subjective probabilities in household surveys. Annual Review of Economics 1, 543–562.

Hurd, M.D., McFadden, D., Gan, L., et al., 1998. Subjective survival curves and life cycle behavior. Inquiries in the Economics of Aging 259, 305.

Jappelli, T., Pistaferri, L., 2014. Fiscal policy and MPC heterogeneity. American Economic Journal: Macroeconomics 6, 107–36.

Jappelli, T., Pistaferri, L., 2020. Reported MPC and unobserved heterogeneity. American Economic Journal: Economic Policy 12, 275–97.

Jones, A., 1989. A double-hurdle model of cigarette consumption. Journal of Applied Econometrics 4, 23–39.

Kaplan, G., Violante, G.L., 2014. A model of the consumption response to fiscal stimulus payments. Econometrica 82, 1199–1239.

Kleinjans, K., Soest, A.V., 2014. Rounding, focal point answers and nonresponse to subjective probability questions. Journal of Applied Econometrics 29, 567–585.

Lambert, D., 1992. Zero inflated poisson regression with an application to defects in manufacturing. Technometrics 34, 1–14.

Manski, C.F., Molinari, F., 2010. Rounding probabilistic expectations in surveys. Journal of Business and Economic Statistics 28, 219–231.

Motta, V., 2019. Estimating poisson pseudo-maximum-likelihood rather than log-linear model of a log-transformed dependent variable. RAUSP Management Journal 54, 508–518.

Mullahy, J., 1986. Specification and testing of some modified count data models. Journal of Econometrics 33, 341–365.

Mullahy, J., 1997. Heterogeneity, excess zeros and the structure of count data models. Journal of Applied Econometrics 12, 337–350.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

Neyman, J., Scott, E.L., 1948. Consistent estimates based on partially consistent observations. Econometrica 16, 1–32.

Oh, H., Reis, R., 2012. Targeted transfers and the fiscal response to the great recession. Journal of Monetary Economics 59, S50–S64.

Paiella, M., Pistaferri, L., 2017. Decomposing the wealth effect on consumption. Review of Economics and Statistics 99, 710–721.

Pohlmeier, W., Ulrich, V., 1995. An econometric model of the two-part decision-making process in the demandfor health care. Journal of Human Resources 30, 339–361.

Poirier, D.J., 1980. Partial observability in bivariate probit models. Journal of Econometrics 12, 209–217.

Santos Silva, J., Tenreyro, S., 2006. The Log of Gravity. The Review of Economics and Statistics 88, 641–658.

Santos Silva, J., Tenreyro, S., 2011. Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. Economics Letters 112, 220–222.

Schwarz, G., 1978. Estimating the dimensions of a model. Annals of Statistics 6, 461–464.

Schwarz, N., Oyserman, D., 2001. Asking questions about behavior: Cognition, communication, and questionnaire construction. The American Journal of Evaluation 22, 127–160.

Sirchenko, A., 2020. A model for ordinal responses with heterogeneous status quo outcomes. Studies in Nonlinear Dynamics & Econometrics 24, 20180059.

Smith, M., 2003. On dependency in double-hurdle models. Statistical Papers 44, 581–595.

Terza, J., 1985. A tobit-type estimator for the censored poisson regression model. Economics Letters 18, 361 – 365.

Vuong, Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–334.

Appendix

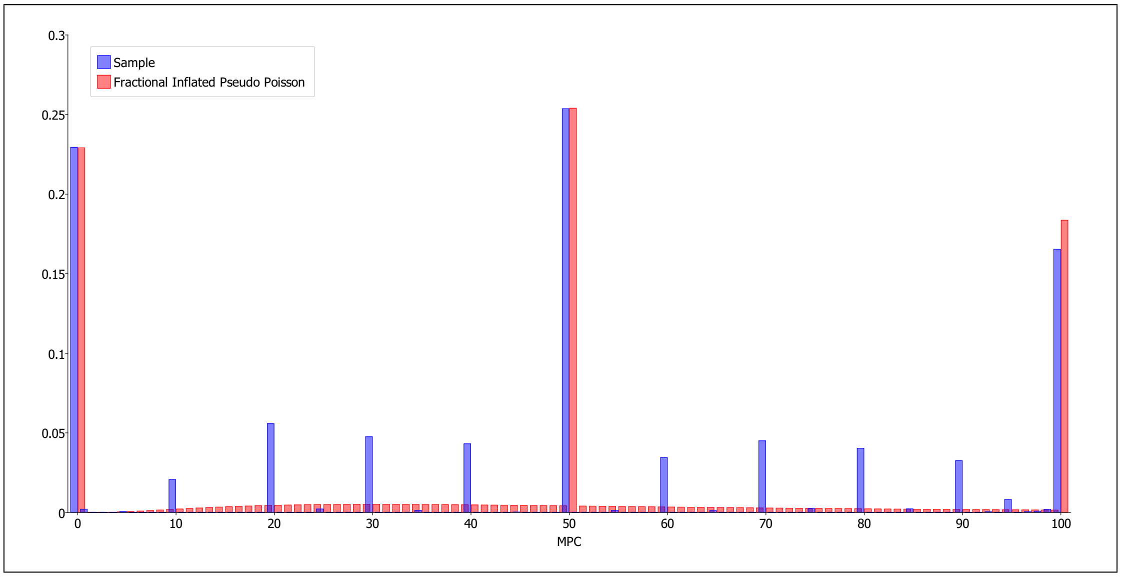

Figure 2: Estimated densities and sample proportions: Primary model (3 inflation points)

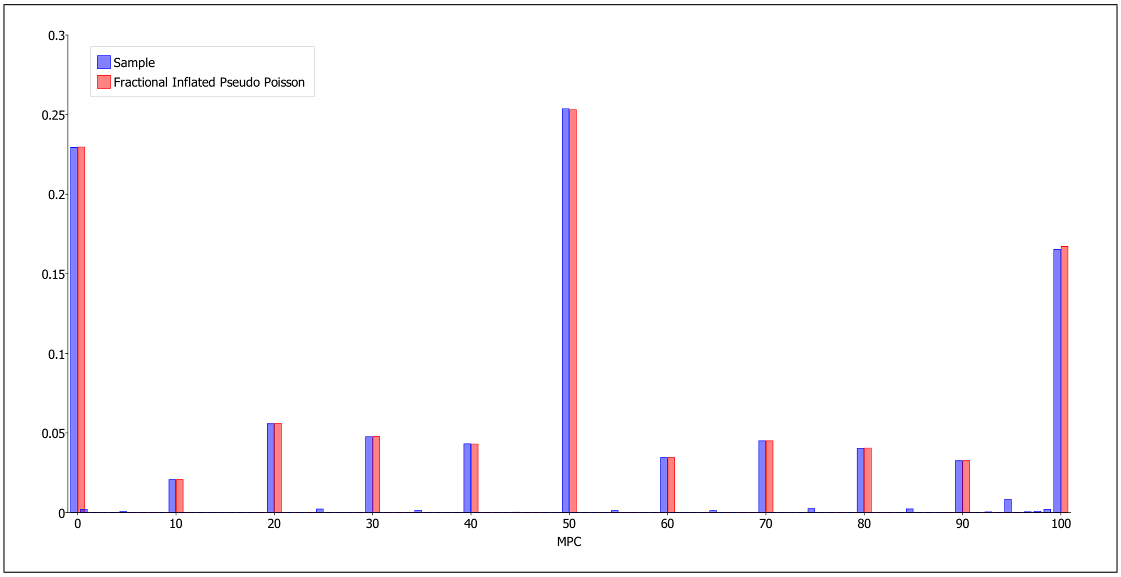

Figure 3: Estimated densities and sample proportions: Full model (11 inflation points)

Table 1: Summary Statistics

Variable	Description	Mean (s.d.)	[Min., Max.]
Dependent Variable			
MPC	Imagine you unexpectedly receive a reimbursement equal to the amount your household earns in a month. How much of it would you save and how much would you spend? Please give the percentage you would save and the percentage you would spend. Measured on 0-100 scale.	47.22 (35.22)	[0,100]
Independent Variables			
Age	*Omitted category: Aged 61 and above.*		
Age 18-30	= 1 if head of household is aged 18-30, 0 otherwise.	0.03	[0,1]
Age 31-45	= 1 if head of household is aged 31-45, 0 otherwise.	0.17	[0,1]
Age 46-60	= 1 if head of household is aged 46-60, 0 otherwise.	0.30	[0,1]
Male	= 1 if male, 0 if female.	0.54	[0,1]
Married	= 1 if married, 0 otherwise.	0.58	[0,1]
Years of Education	Years of education.	9.63 (4.52)	[0,20]
Family Size	Number of individuals in the household.	2.36 (1.24)	[0,12]
Resident in South	= 1 if residence is located in the South of the country, 0 otherwise.	0.33	[0,1]
City of Residence Size	*Omitted category: City population above 500,000.*		
City Size: $< 20,000$	= 1 if located in a city with population less that 20,00.0	0.25	[0,1]
City Size: $20,000 - 40,000$	= 1 if located in a city with population between 20,000 and 40,000.	0.18	[0,1]
City Size: $40,000 - 500,000$	= 1 if located in a city with population between 40,000 and 500,000.	0.48	[0,1]
Financial Assets	*Omitted category: 5^{th} quintile of the financial asset distribution.*		
Financial Asset Quintile: I	= 1 if in 1^{st} quintile of the asset distribution, 0 otherwise.	0.20	[0,1]
Financial Asset Quintile: II	= 1 if in 2^{nd} quintile of the asset distribution, 0 otherwise.	0.22	[0,1]
Financial Asset Quintile: III	= 1 if in 3^{rd} quintile of the asset distribution, 0 otherwise.	0.19	[0,1]
Financial Asset Quintile: IV	= 1 if in 4^{th} quintile of the asset distribution, 0 otherwise.	0.20	[0,1]
Household Income	*Omitted category: 5^{th} quintile of the income distribution.*		
Income Quintile: I	= 1 if in 1^{st} quintile of the income distribution, 0 otherwise.	0.26	[0,1]
Income Quintile: II	= 1 if in 2^{nd} quintile of the income distribution, 0 otherwise.	0.22	[0,1]
Income Quintile: III	= 1 if in 3^{rd} quintile of the income distribution, 0 otherwise.	0.20	[0,1]
Income Quintile: IV	= 1 if in 4^{th} quintile of the income distribution, 0 otherwise.	0.17	[0,1]
Positive Debt	=1 if has outstanding financial liabilities, 0 otherwise.	0.22	[0,1]
Unemployed	= 1 if head of household is unemployed, 0 otherwise.	0.04	[0,1]
Homeowner	= 1 if home is owned by the household, 0 otherwise.	0.71	[0,1]
2016	= 1 if observations is in 2016, 0 if 2010 observation.	0.48	[0,1]
Exclusion Restrictions			
Interviewee Understanding	Based on *How do you rate the respondent's level of understanding of the questions.* Measured on a 10-point scale.	8.11 (1.62)	[1,10]
Financial Literacy	Index of financial literacy. It is the sum of correct answer to 3 questions relating to interest rates, inflation and risk.	1.53 (1.01)	[0,3]
Financial Understanding	Based on: *How do you rate the reliability of the information on forms of saving and financial investment provided by the respondent?* Measured on a 10-point scale.	7.86 (1.72)	[0,10]
Income Understanding	Based on: *How do you rate the reliability of the information on income provided by the respondent?* Measured on a 10-point scale.	8.01 (1.68)	[0,10]
Ln(Interview Duration)	Natural logarithm of interview duration.	3.87 (0.41)	[2.08,5.52]
Observations	15,365		

Table 2: Summary Expected Values, EV(s), and BICs

				$FIPP$ (Full)			$FIPP$ (Primary)		
	Sample	$Tobit$	$Poisson$	Overall	Purged	Joint	Overall	Purged	Joint
					Pooled				
EV	47.22	46.73	47.41	47.21	67.08	47.22	47.21	50.83	47.22
sd/se	(35.22)	(2.35)[11.13]	(0.28)[12.75]	(0.27)[36.17]	(1.54)	(0.04)	(0.27)[36.16]	(0.35)	(0.12)
BIC		$112,278$	$564,589$		$75,691$			$150,963$	
					Panel				
EV	47.22	46.71	42.07	46.72	76.27	47.04	46.71	52.16	41.30
sd/se	(35.22)	(2.35)[11.13]	(0.31)[41.11]	(0.27)[34.95]	(1.18)	(0.23)	(0.27)[34.95]	(1.06)	(0.65)
BIC		$112,271$	$176,259$		$\mathbf{68,152}$			$92,144$	

Notes: Preferred model in **bold**; sd/se standard deviation/standard error of prediction in parentheses (.); Implied sd of predicted distribution in brackets [.].

Table 3: Pairwise $Vuong$ tests

	Alternative Model		
	Poisson	$FIPP$ (primary)	Total
Null Model		Pooled	
$FIPP$ (full)	115.8	56.3	2
$FIPP$ (primary)	91.9	-	1
		Panel	
$FIPP$ (full)	51.8	54.9	2
$FIPP$ (primary)	40.7	-	1

Notes: large positive (negative) values of V favour null (alternative) model.

Table 4: Summary probabilities

Outcome	Sample	Overall	*s.e.*	Joint	*s.e.*	Marg.	*s.e.*
			Pooled				
0	0.230	0.230	(0.003)	0.230	(0.003)	0.238	(0.003)
10	0.021	0.021	(0.001)	0.021	(0.001)	0.022	(0.001)
20	0.056	0.056	(0.002)	0.056	(0.002)	0.058	(0.002)
30	0.048	0.048	(0.002)	0.048	(0.002)	0.050	(0.002)
40	0.043	0.043	(0.002)	0.043	(0.002)	0.045	(0.002)
50	0.254	0.254	(0.004)	0.253	(0.004)	0.263	(0.004)
60	0.035	0.035	(0.001)	0.034	(0.001)	0.035	(0.002)
70	0.045	0.045	(0.002)	0.045	(0.002)	0.046	(0.002)
80	0.041	0.041	(0.002)	0.040	(0.002)	0.042	(0.002)
90	0.033	0.033	(0.001)	0.032	(0.001)	0.033	(0.001)
100	0.166	0.166	(0.003)	0.164	(0.003)	0.169	(0.003)
$P(A)$		0.035	(0.002)				
			Panel				
0	0.230	0.230	(0.003)	0.230	(0.003)	0.270	(0.005)
10	0.021	0.021	(0.001)	0.021	(0.001)	0.025	(0.001)
20	0.056	0.056	(0.002)	0.056	(0.002)	0.067	(0.002)
30	0.048	0.048	(0.002)	0.048	(0.002)	0.057	(0.002)
40	0.043	0.043	(0.002)	0.043	(0.002)	0.052	(0.002)
50	0.254	0.253	(0.003)	0.253	(0.004)	0.311	(0.005)
60	0.035	0.035	(0.001)	0.034	(0.001)	0.043	(0.002)
70	0.045	0.045	(0.002)	0.045	(0.002)	0.057	(0.002)
80	0.041	0.041	(0.002)	0.040	(0.002)	0.051	(0.002)
90	0.033	0.033	(0.001)	0.033	(0.001)	0.042	(0.002)
100	0.166	0.167	(0.003)	0.019	(0.007)	0.026	(0.008)
$P(A)$		0.179	(0.007)				

Notes: "Overall" correspond to the sample probabilities of the full model, the marginal inflation probabilities correspond to the OP probabilities evaluated in isolation and the joint probabilities of these augmented with the accurate reporting probabilities. $P(A)$ reports the proportion of accurate reporters.

Table 5: Partial Effects

	$Tobit$ model	$FIPP$ - Overall	Inaccurate Reporting	$FractionalPoison$
Age 18-30	4.409**	4.091**	0.009	4.496
	(1.783)	(1.656)	(0.019)	(3.259)
Age 31-45	3.347***	3.777***	0.001	4.018**
	(0.946)	(0.876)	(0.011)	(1.767)
Age 46-60	3.214***	2.869***	-0.003	2.556
	(0.753)	(0.703)	(0.009)	(1.617)
Male	-0.286	-0.504	-0.000	1.330
	(0.612)	(0.567)	(0.007)	(1.222)
Married	-0.990	-1.265**	-0.002	-3.444*
	(0.767)	(0.729)	(0.009)	(1.986)
Years of Education	0.300***	0.193**	0.001	0.232
	(0.082)	(0.075)	(0.001)	(0.187)
Family Size	1.791***	1.838***	0.013***	0.022
	(0.330)	(0.308)	(0.004)	(0.695)
Resident in South	9.607***	10.217***	0.010	8.724***
	(0.664)	(0.621)	(0.009)	(1.592)
City Size: $< 20,000$	-10.876***	-11.204***	-0.021	-1.219
	(1.112)	(1.049)	(0.015)	(2.984)
City Size: $20,000 - 40,000$	-8.915***	-9.146***	-0.025*	-2.364
	(1.162)	(1.103)	(0.015)	(3.273)
City Size: $40,000 - 500,000$	-6.829***	-7.321***	-0.007	-0.592
	(1.031)	(0.974)	(0.013)	(2.921)
Income Quintile: I	7.318***	7.028***	0.081***	6.083**
	(1.355)	(1.267)	(0.016)	(3.085)
Income Quintile: II	4.079***	4.242***	0.026*	1.631
	(1.164)	(1.085)	(0.014)	(2.417)
Income Quintile: III	3.804***	3.578***	0.012	-5.649**
	(1.080)	(1.011)	(0.013)	(2.233)
Income Quintile: VI	2.938***	2.352**	0.007	-5.551***
	(1.033)	(0.969)	(0.013)	(2.143)
Financial Asset Quintile: I	12.182***	11.369***	0.111***	20.922***
	(1.131)	(1.058)	(0.013)	(2.501)
Financial Asset Quintile: II	6.727***	6.317***	0.052***	10.617***
	(1.007)	(0.940)	(0.012)	(1.940)
Financial Asset Quintile: III	4.245***	4.470***	0.031**	8.159***
	(1.000)	(0.934)	(0.012)	(1.930)
Financial Asset Quintile: IV	1.924**	1.477*	-0.017	3.602*
	(0.943)	(0.895)	(0.012)	(2.039)
Homeowner	-1.815**	-1.440 **	-0.017**	-0.339
	(0.708)	(0.664)	(0.008)	(1.750)
Positive Debt	-1.414*	-0.902	0.008	6.314***
	(0.738)	(0.681)	(0.009)	(1.350)
Unemployed	5.794***	3.880***	0.044***	1.784
	(1.439)	(1.360)	(0.015)	(2.603)
Financial Literacy	-0.960***	0.278	0.006	
	(0.333)	(0.180)	(0.004)	
Interviewee Understanding	0.208	0.397***	0.008***	
	(0.264)	(0.137)	(0.003)	
Financial Understanding	-0.017	-0.404***	-0.009***	
	(0.299)	(0.155)	(0.003)	
Income Understanding	-1.534***	-0.210	-0.004	
	(0.300)	(0.157)	(0.003)	
Interview Duration	-0.664	-1.583***	-0.034***	
	(0.760)	(0.417)	(0.009)	
2016	-0.857	-0.085	-0.002	7.762 ***
	(0.612)	(0.544)	(0.007)	(1.220)
Observations	15,365	15,365	15,365	15,365

Notes:***, ** and *, denote significance at 1, 5 and 10%, respectively. Standard errors presented in parentheses. Dependent variable is measured on 0-100 scale. Variable definitions and omitted categories are presented in Table 1.

Table 6: Partial Effects of Marginal Inflation Probabilities (se's underneath)

Variable	0	10	20	30	40	50	60	70	80	90	100
Age 18-30	-0.047	-0.002	-0.004	-0.002	-0.001	0.015	0.005	0.008	0.010	0.010	0.008
	(0.021)	(0.001)	(0.002)	(0.001)	(0.000)	(0.007)	(0.002)	(0.004)	(0.004)	(0.005)	(0.004)
Age 31-45	-0.049	-0.002	-0.004	-0.002	-0.001	0.015	0.005	0.009	0.010	0.010	0.009
	(0.011)	(0.000)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Age 46-60	-0.040	-0.002	-0.004	-0.002	-0.001	0.013	0.004	0.007	0.008	0.009	0.007
	(0.009)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Male	0.009	0.000	0.001	0.000	0.000	-0.003	-0.001	-0.002	-0.002	-0.002	-0.002
	(0.007)	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married	0.012	0.001	0.001	0.001	0.000	-0.004	-0.001	-0.002	-0.002	-0.003	-0.002
	(0.009)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Years of Education	-0.002	-0.000	-0.000	-0.000	-0.000	0.001	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Family Size	-0.018	-0.001	-0.002	-0.001	-0.000	0.006	0.002	0.003	0.004	0.004	0.003
	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Resident in South	-0.129	-0.005	-0.012	-0.006	-0.003	0.041	0.014	0.023	0.026	0.028	0.023
	(0.008)	(0.000)	(0.001)	(0.001)	(0.000)	(0.003)	(0.001)	(0.002)	(0.002)	(0.003)	(0.007)
City Size: $< 20,000$	0.147	0.006	0.013	0.007	0.003	-0.046	-0.016	-0.026	-0.030	-0.031	-0.026
	(0.013)	(0.001)	(0.001)	(0.001)	(0.000)	(0.005)	(0.002)	(0.003)	(0.003)	(0.004)	(0.008)
City Size: $20,000 - 40,000$	0.113	0.005	0.010	0.005	0.002	-0.035	-0.012	-0.020	-0.023	-0.024	-0.020
	(0.014)	(0.001)	(0.001)	(0.001)	(0.000)	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)	(0.006)
City Size: $40,000 - 500,000$	0.101	0.004	0.009	0.005	0.002	-0.032	-0.011	-0.018	-0.021	-0.021	-0.018
	(0.012)	(0.001)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)	(0.005)
Income Quintile: I	-0.038	-0.002	-0.003	-0.002	-0.001	0.012	0.004	0.007	0.008	0.008	0.007
	(0.015)	(0.001)	(0.001)	(0.001)	(0.000)	(0.005)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Income Quintile: II	-0.041	-0.002	-0.004	-0.002	-0.001	0.013	0.005	0.007	0.008	0.009	0.007
	(0.013)	(0.001)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Income Quintile: III	-0.052	-0.002	-0.005	-0.002	-0.001	0.016	0.006	0.009	0.011	0.011	0.009
	(0.012)	(0.001)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Income Quintile: VI	-0.037	-0.002	-0.003	-0.002	-0.001	0.012	0.004	0.007	0.008	0.008	0.007
	(0.011)	(0.000)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)
Financial Asset Quintile: I	-0.059	-0.003	-0.005	-0.003	-0.001	0.019	0.007	0.011	0.012	0.013	0.011
	(0.013)	(0.001)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)
Financial Asset Quintile: II	-0.041	-0.002	-0.004	-0.002	-0.001	0.013	0.005	0.007	0.008	0.009	0.007
	(0.011)	(0.000)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
Financial Asset Quintile: III	-0.032	-0.001	-0.003	-0.002	-0.001	0.010	0.004	0.006	0.007	0.007	0.006
	(0.011)	(0.000)	(0.001)	(0.001)	(0.000)	(0.004)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Financial Asset Quintile: IV	-0.028	-0.001	-0.002	-0.001	-0.001	0.009	0.003	0.005	0.006	0.006	0.005
	(0.010)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Homeowner	0.028	0.001	0.002	0.001	0.001	-0.009	-0.003	-0.005	-0.006	-0.006	-0.005
	(0.009)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Positive Debt	-0.024	-0.001	-0.002	-0.001	-0.001	0.008	0.003	0.004	0.005	0.005	0.004
	(0.018)	(0.001)	(0.002)	(0.001)	(0.000)	(0.006)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)
Unemployed	0.009	0.000	0.001	0.000	0.000	-0.003	-0.001	-0.002	-0.002	-0.002	-0.002
	(0.008)	(0.000)	(0.001)	(0.000)	(0.000)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
2016	0.011	0.000	0.001	0.001	0.000	-0.004	-0.001	-0.002	-0.002	-0.002	-0.002
	(0.007)	(0.000)	(0.001)	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Notes: partial effects of the OP equation; standard errors presented in parentheses. Stars are omitted due to space constraints. Variable definitions and omitted categories are presented in table 1.

Table 7: Inflated Poisson Parameter Estimates (Inaccurate, Fractional and Inflation Equations); Panel Full model

Variable	Inaccurate eqn		Fractional eqn		Inflation eqn	
Constant	-0.7384***	(0.192)	3.9654***	(0.328)	–	-
Age 18-30	0.0376	(0.086)	0.2917	(0.211)	0.1566**	(0.070)
Age 31-45	0.0048	(0.048)	0.2606**	(0.115)	0.1606***	(0.036)
Age 46-60	-0.0128	(0.040)	0.1658	(0.106)	0.1329***	(0.029)
Male	-0.0013	(0.031)	0.0863	(0.079)	0.0300	(0.023)
Married	-0.0086	(0.040)	-0.2235*	(0.128)	0.0401	(0.029)
Years of Education	0.0036	(0.004)	0.0151	(0.012)	0.0064*	(0.003)
Family Size	0.0583***	(0.017)	0.0015	(0.045)	0.0585***	(0.013)
Resident in South	0.0456	(0.040)	0.5659***	(0.104)	0.4279***	(0.028)
City Size: $< 20,000$	-0.0937	(0.063)	-0.0791	(0.194)	0.4864***	(0.045)
City Size: $20,000 - 40,000$	-0.1112*	(0.064)	-0.1534	(0.213)	0.3727***	(0.046)
City Size: $40,000 - 500,000$	-0.0323	(0.056)	-0.0384	(0.190)	0.3338***	(0.041)
Financial Asset Quintile: I	0.4882***	(0.060)	1.3573***	(0.170)	0.1968***	(0.044)
Financial Asset Quintile: II	0.2270***	(0.054)	0.6888***	(0.126)	0.1370***	(0.037)
Financial Asset Quintile: III	0.1369**	(0.054)	0.5293***	(0.126)	0.1061***	(0.037)
Financial Asset Quintile: IV	-0.0750	(0.054)	0.2336*	(0.133)	0.0928***	(0.034)
Income Quintile: I	0.3558***	(0.070)	0.3946**	(0.202)	0.1265**	(0.051)
Income Quintile: II	0.1160*	(0.062)	0.1058	(0.157)	0.1370***	(0.043)
Income Quintile: III	0.0517	(0.059)	-0.3665**	(0.145)	0.1729***	(0.040)
Income Quintile: IV	0.0312	(0.057)	-0.3601**	(0.140)	0.1239***	(0.038)
Positive Debt	0.0365	(0.038)	0.4096***	(0.088)	0.0925***	(0.028)
Unemployed	0.1917***	(0.066)	0.1158	(0.168)	0.0798	(0.059)
Homeowner	-0.0768**	(0.035)	-0.0220	(0.113)	0.0282	(0.027)
2016	-0.0079	(0.032)	0.5035***	(0.081)	0.0370*	(0.022)
Interviewee Understanding	0.0372***	(0.013)	-	-	-	-
Financial Literacy	0.0260	(0.017)	-	-	-	-
Financial Understanding	-0.0378***	(0.015)	-	-	-	-
Income Understanding	-0.0196	(0.015)	-	-	-	-
Ln(Duration)	-0.1480***	(0.040)	-	-	-	-
μ_0	-	-	-	-	-0.4809***	(0.077)
μ_1	-	-	-	-	-0.4012***	(0.077)
μ_2	-	-	-	-	-0.1976***	(0.076)
μ_3	-	-	-	-	-0.0336	(0.076)
μ_4	-	-	-	-	0.1106	(0.077)
μ_5	-	-	-	-	1.0459***	(0.081)
μ_6	-	-	-	-	1.2171***	(0.083)
μ_7	-	-	-	-	1.4904***	(0.088)
μ_8	-	-	-	-	1.8373***	(0.103)
μ_9	-	-	-	-	2.3408***	(0.171)
Random Effects	0.1944***	(0.070)	2.7588***	(0.230)	0.1059**	(0.044)
$Cov_{A,F}$	0.6363***	(0.069)				
$Cov_{A,I}$	0.0540***	0.042				

Notes:***, ** and *, denote significance at 1, 5 and 10%, respectively. $Cov_{A,F}$ is covariance between the random effects of the accurate reporting model and the fractional model and $Cov_{A,I}$ is covariance between the random effects of the accurate reporting model and the inflation model. Variable definitions and omitted categories are presented in Table 1.

Table 8: Inflated Poisson Parameter Estimates (Inaccurate, Fractional and Inflation Equations); Panel Primary model

Variable	Accurate eqn		Fractional eqn		Inflation eqn	
Constant	0.4052***	(0.164)	3.6953***	(0.074)		
Age 18-30	0.0463	(0.086)	0.2543***	(0.069)	0.1258	(0.111)
Age 31-45	0.1730***	(0.043)	0.2131***	(0.033)	0.0253	(0.057)
Age 46-60	0.0087	(0.034)	0.0310	(0.024)	0.1504***	(0.043)
Male	-0.0372	(0.027)	-0.0242	(0.025)	0.0115	(0.037)
Married	0.0575*	(0.035)	0.0053	(0.030)	-0.1037	(0.046)
Years of Education	0.0038	(0.004)	0.0027	(0.003)	0.0113***	(0.005)
Family Size	-0.0107	(0.015)	0.0367***	(0.012)	0.0747***	(0.020)
Resident in the South	0.3168***	(0.031)	0.3453***	(0.027)	0.2857***	(0.043)
City Size: $< 20,000$	-0.2943***	(0.051)	-0.1886***	(0.043)	-0.4870***	(0.073)
City Size: $20,000 - 40,000$	-0.2459***	(0.053)	-0.1532***	(0.045)	-0.3923***	(0.075)
City Size: $40,000 - 500,000$	-0.2533***	(0.048)	-0.1623***	(0.040)	-0.2859***	(0.067)
Financial Asset Quintile: I	-0.1905***	(0.051)	0.0505	(0.033)	0.6934***	(0.069)
Financial Asset Quintile: II	0.0556	(0.043)	0.0312	(0.028)	0.4005***	(0.060)
Financial Asset Quintile: III	0.0354	(0.042)	0.0296	(0.027)	0.2264***	(0.057)
Financial Asset Quintile: IV	-0.0650*	(0.039)	-0.0519	(0.026)	0.2218	(0.052)
Income Quintile: I	0.1642***	(0.060)	0.1334***	(0.046)	0.2040***	(0.079)
Income Quintile: II	0.1629	(0.050)	0.0658*	(0.038)	0.0483	(0.065)
Income Quintile: III	0.1093***	(0.046)	-0.0453	(0.036)	0.1577***	(0.060)
Income Quintile: IV	0.0639	(0.044)	-0.0784***	(0.032)	0.1657***	(0.056)
Positive Debt	-0.1361***	(0.033)	-0.0284	(0.023)	-0.0653	(0.042)
Unemployed	-0.1507**	(0.068)	-0.2900	(0.049)	0.5810	(0.083)
Homeowner	-0.0405	(0.032)	0.0228	(0.025)	-0.1025***	(0.042)
2016	-0.1722	(0.027)	-0.0455	(0.014)	0.0675**	(0.033)
Interviewee Understanding	-0.0369***	(0.011)	-	-	-	-
Financial Literacy	-0.0016	(0.014)	-	-	-	-
Financial Understanding	0.0185	(0.013)	-	-	-	-
Income Understanding	-0.0572***	(0.013)	-	-	-	-
Ln(Duration)	0.0512	(0.033)	-	-	-	-
μ_0	-	-	-	-	0.0505	(0.121)
μ_1	-	-	-	-	1.6468***	(0.135)
Random Effects	0.1325***	(0.037)	0.6063***	(0.018)	0.3238***	(0.099)
$Cov_{A,F}$	0.1990***	(0.037)				
$Cov_{A,I}$	0.1465***	(0.037)				

Notes:***, ** and *, denote significance at 1, 5 and 10%, respectively. $Cov_{A,F}$ is covariance between the random effects of the accurate reporting model and the fractional model and $Cov_{A,I}$ is covariance between the random effects of the accurate reporting model and the inflation model. Variable definitions and omitted categories are presented in Table 1.