

# Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature

By RACHAEL MEAGER\*

*Expanding credit access in developing contexts could help some households while harming others. Microcredit studies show different effects at different quantiles of household profit, including some negative effects; yet these findings also differ across studies. I develop new Bayesian hierarchical models to aggregate the evidence on these distributional effects for mixture-type outcomes such as household profit. Applying them to microcredit, I find a precise zero effect from the 5th to 75th quantiles, and uncertain yet large effects on the upper tails, particularly for households with business experience. These quantile estimates are more reliable than averages because the data is fat-tailed.*

Financial market expansions in the developing world have the potential to create winners and losers. Increasing access to credit in particular may have heterogeneous effects, both because borrowers differ in their investment opportunities and because of general equilibrium dynamics (Banerjee 2013, Kaboski and Townsend 2011). Proponents of financial interventions such as microcredit claim that the positive impact on high-productivity borrowers justifies continued market expansion; detractors claim that the resulting market "saturation" leads to exploitative lending practices which systematically harm the most vulnerable borrowers (Ahmad 2003, Roodman 2012). Although recent studies have estimated sets of quantile treatment effects to address this concern, existing meta-analyses of microcredit ignored these sets of quantile effects due to a lack of methodology to aggregate them (Meager 2019, Vivalt 2016, Banerjee et al 2015a). In this paper I develop new models to aggregate evidence on distributional treatment effects, and apply them to randomized trials of expanding access to microcredit.

\* The London School of Economics and Political Science, Houghton St, London WC2A 2AE. r.meager@lse.ac.uk. Funding for this research was generously provided by the Berkeley Initiative for Transparency in the Social Sciences (BITSS), a program of the Center for Effective Global Action (CEGA), with support from the Laura and John Arnold Foundation. I am immensely grateful to Ted Miguel and the team at BITSS. I also thank Esther Duflo, Abhijit Banerjee, Anna Mikusheva, Rob Townsend, Victor Chernozhukov, Isaiah Andrews, Tamara Broderick, Ryan Giordano, Jonathan Huggins, Jim Savage, Andrew Gelman, Tetsuya Kaji, Michael Betancourt, Bob Carpenter, Ben Goodrich, Whitney Newey, Jerry Hausman, John Firth, Cory Smith, Arianna Ornaghi, Greg Howard, Nick Hagerty, Jack Liebersohn, Peter Hull, Ernest Liu, Donghee Jo, Matt Lowe, Yaroslav Mukhin, Ben Marx, Reshma Hussam, Yu Shi, Frank Schilbach, David Atkin, Gharad Bryan, Oriana Bandiera, Dean Karlan, Greg Fischer, and the audiences of many seminars for their feedback and advice. I thank the authors of the 7 microcredit studies, and the journals in which they published, for making their data and code public. All my code and data are accessible online at <https://bitbucket.org/rmeager/aggregating-distributional-treatment-effects>. The paper was pre-registered on the OSF at <https://osf.io/tdvc8/> recognizing that pre-registration is different for a meta-analysis.

Microcredit institutions reached 140 million low-income clients with a global loan portfolio worth 124 billion dollars in 2019, and the figure is growing yearly (Microfinance Barometer, 2020). At this scale, negative impacts for even a small subset of borrowers would be concerning, and several governments have curtailed microfinance operations ostensibly for this reason (Microfinance Focus 2011, Banerjee 2013, Breza and Kinnan 2018). Even if microcredit benefits all households, an unequal distribution of gains may affect social and political institutions (Acemoglu and Robinson 2008, Acemoglu et al 2015). Several randomized trials find evidence of negative effects at lower quantiles of household business profits, but others find zero or positive impact there and on higher quantiles of the distribution (Augsburg et al. 2015, Attanasio et al. 2015, Banerjee et al. 2015b, Crepon et al. 2015, Angelucci et al. 2015, Tarozzi et al. 2015, Karlan and Zinman 2011). Forming a broad consensus on the distributional impact of microcredit is difficult given the lack of power to estimate these effects (Leon and Heo 2009) and the possibility of substantial differences in effects across studies, often referred to as concerns about generalizability or "external validity". Yet despite these concerns, academics and policymakers need to understand the typical impact of microcredit, especially given this the potential for harm (Schicks 2013).

The main contribution of this paper is a method to aggregate the distributional impact of microcredit that addresses concerns about generalizability even in the presence of certain data features which – though widespread in economics – substantially complicate the use of existing econometric approaches. I work within the Bayesian hierarchical modelling framework because it specifies the potential for treatment effect heterogeneity as a parameter of interest in its own right, and uses that parameter to adjust the uncertainty about the typical impact across settings and likely impact in new settings (Rubin 1981, Gelman et al 2004, Gelman and Pardoe 2006). This approach is well-established in statistics and is increasingly used for evidence aggregation in economics (Chib and Greenberg 1995, Dehejia 2003, Hsiang, Burke and Miguel 2013, Vivalt 2016, Bandiera et al 2017, Meager 2019). Partial pooling models such as Empirical Bayes have also been used to borrow power across regions or sub-units when studying large geographic areas, or to combine multiple estimates within a study (Hull 2018, Chetty and Hendren 2018, Chetty, Friedman and Rockoff 2014). Yet within the hierarchical framework there are no established tools to aggregate distributional effects.<sup>1</sup>

In applying the Bayesian hierarchical framework to the distributional effects of microcredit, I confront several challenges that require a substantially new, tailored modelling approach. The first issue is that for the microcredit data, even

<sup>1</sup>Even outside of the hierarchical framework, the economics literature on external validity and generalizability has focused on inference across different types of average effects, such as extrapolating the LATE to the ATE within similar settings, or adjustments based on correlations between observable and unobservable covariates (Heckman, Tobias, and Vytalil 2001, Angrist 2004, Angrist and Fernandez-Val 2010, Bertanha and Imbens 2014, Allcott 2015, Dehejia, Pop-Eleches and Samii 2015, Gechter 2015, Athey and Imbens 2016, Andrews and Oster 2018).

the estimation of quantiles themselves is not straightforward because the business outcomes – profit, revenues and expenditures – have large spikes at zero which invalidate the quantile estimator’s classical asymptotics (Mosteller, 1946). The sampling variance of the quantiles is no longer Gaussian, and more troublingly, is often estimated to be zero in the sample due to the large number of ties. Obtaining zero standard errors causes practical problems for evidence aggregation because the general treatment effect is typically estimated using a weighted average of the effects from each study, where the *inverse* standard error enters the weight. Yet even if one does not face this problem, there is a second challenge: aggregating using potentially quite different weights at different quantiles can introduce quantile crossing – the situation in which the point estimate of, say, the 50th quantile lies above the point estimate of the 60th quantile – in the aggregate quantile effects. Thus, the inherent monotonicity constraint on the quantile function can be violated during the aggregation exercise even if it holds in the individual studies.

My approach addresses both of these problems by directly modelling the distributions of household outcomes using a flexible set of parametric mixture models. I specify treatment effects on the parameters of the tailored density function and partially pool information on these effects across studies by placing a hierarchical structure on these treatment effects. Building these models relies on contextual economic knowledge of the variables at hand. For example, when we sample households in rural villages in India or Mexico and ask about their business profit, we ought to expect a spike at zero because not all households own businesses and nor do they necessarily operate their businesses in every period, often being engaged in seasonal agricultural labour for months at a time. In this context, household business variables are produced by an extensive-margin decision (to operate a business or not) followed by an intensive margin decision (how much to spend or invest once in operation) – the spikes at zero are part of the data-generating process that can and should be modelled explicitly.

To capture these data features, I develop a set of mixture probability density function (PDF) models, each of which has a point mass or “spike” of households at zero and a continuous tail distribution or “slab” on the rest of the real line. I specify potential treatment effects on the probabilities that households find themselves in the tails versus the spike, and effects on features of the tails such as means and variances. Economic theory and prior data suggests that profits, expenditures, revenues and even consumer durables spending outcomes tend to have fatter tails than Normal distributions, so I consider models with either log-normal, Pareto, or a mixture of these two distributions for the continuous portions of these variables (Piketty 2015, Gabaix 2008, Roy 1950). I analytically compute and report the quantile treatment effects implied by the underlying parameters of the tailored PDFs by employing the method of Castellaci (2012); the posterior uncertainty on these transformed effects is automatically provided within the Bayesian framework.

Applying these models to seven randomized trials of expanding access to microcredit, I find a precise zero effect on household outcomes from the 5th to 75th percentiles. Above the 75th percentile, there is substantial probability of a large positive impact on most outcomes, but there is greater uncertainty around this effect due to heterogeneity within and across studies. There are no generalizable negative quantile treatment effects. These patterns hold regardless of whether one uses lognormal or Pareto tails, although the lognormal fits this data best. Moreover, my analysis shows that the tails of profit, revenues and business expenditures are so heavy that estimated average treatment effects and use of Gaussian asymptotics for averages are unreliable (Koenker and Basset 1978, Mosteller 1946). The majority of the right tail impact of microcredit and the heterogeneity across studies occurs within the group of households who had previous business experience.<sup>2</sup> The resulting potential increase economic inequality even within the group of households who previously operated businesses implies that the social welfare effects of microcredit are likely to be complex.

The models presented in this paper have broad applicability in economics beyond the microcredit literature. As a complement to the traditional approach in economics of writing review articles to summarize literatures, formal exercises to synthesize evidence improve power and prevent undue focus on the most extreme effects (Rubin 1981). Yet in economics, as in social science more broadly, contextual heterogeneity across study settings often makes it difficult to combine the evidence because the generalizability of the evidence is unknown (Allcott 2015, Bisbee et al 2016, Pritchett and Sandefur 2015). The hierarchical approach is broadly appropriate as it incorporates uncertainty about the heterogeneity in effects across studies, and provides some indication of the extent to which extrapolation across settings is appropriate. Moreover, there are many policy settings in which quantiles are implicated in policy directly, often because welfare or tax policies are explicitly concerned with distributional effects. My models are especially relevant when the data contains discrete point masses due to extensive margin decisions, and studies for which the outcome data is likely to be fat-tailed.

## I. Data and Context

To aggregate the evidence on the quantile treatment effects of microcredit, I use data from seven studies which meet the following inclusion criteria: the main intervention must be an expansion of access to microcredit either at the community or individual level, the assignment of access must be randomized, and the study must be published before February 2015 (the period of my literature search). The selected studies are: Angelucci et al. 2015, Attanasio et al. 2015, Augsburg et al. 2015, Banerjee et al. 2015b, Crepon et al. 2015, Karlan and Zinman 2011, and Tarozzi et al. 2015, six of which were published in a special

<sup>2</sup>In Appendix D I pursue a bounding exercise to show that the precise and generalizable impact at zero is unlikely to be due to low take-up of loans.

issue of the *American Economics Journal: Applied Economics*.<sup>3</sup> I restrict the sample to randomized controlled trials (RCTs) because they typically have high internal validity for estimating causal effects, and they all consider a reasonably uniform and simple "expansion" treatment in this case.<sup>4</sup>

I analyse six outcomes linked to the claim that offering households more credit on more favourable terms should stimulate entrepreneurship (Morduch 1999, Yunus 2006, Roodman 2012). Because microfinance institutions (MFIs) offer lower interest rates relative to informal moneylenders, poor entrepreneurs may be able to start new businesses or grow their existing businesses, increasing their business expenditures, revenues and ultimately profits (Yunus 2006). Households could then increase their consumption in the medium and long run. Yet even households without business investment opportunities may use microloans to shift spending away from "temptation goods" (pleasurable yet harmful expenditures) and towards durable goods (Roodman 2012, Banerjee 2013). This could happen if microcredit increases a household's expectation of escaping poverty in the future, or solves a self-control problem (Banerjee and Mullainathan 2010, Banerjee 2013). In all cases I analyze the effect of expanding access itself, called the Intention to Treat Effect in the original studies (Banerjee et al 2015b). I do not pursue an instrumental variables strategy as network links and potential general equilibrium effects within villages means the Stable Unit Treatment Value Assumption (SUTVA) is likely to be violated at the household level (Banerjee 2013, Kinnan and Townsend 2012, Breza 2012).<sup>5</sup>

Despite the restrictive inclusion criteria, the selected studies still differ substantially in their implementations and local contexts. They cover seven different countries, they have different partner NGOs, offering similar but not identical loan contract structures with different interest rates and loan sizes, and they differ in terms of their randomization units - five randomized at the community level and two at the individual level - with various encouragement and sampling designs (see Appendix C for details). Given this heterogeneity across studies, heterogeneity in the resulting effects seems likely. However, the 95% confidence intervals of the quantile effects do overlap across most of the studies, suggesting meaningful

<sup>3</sup>I focus on expanding access to microcredit because this is the intervention closest to the policy of subsidizing microfinance institutions (MFIs) or promoting interventions under the general umbrella of "microcredit". Other RCTs of microfinance tend to randomly vary certain characteristics of the loans themselves, which allows researchers to understand the impact of these features of the loans but complicates the inference on the general impact of the standard microcredit model (Field et al 2013). Karlan and Zinman 2009 expands access to consumer credit, but microcredit is often considered categorically different to consumer credit; see Banerjee 2013 for a deeper discussion of this.

<sup>4</sup>While the set of studies here may not be a representative sample of all possible microcredit interventions, there is little reason to suspect publication bias in this literature. The papers here published a variety of results most of which were null. This leads to less risk of classical publication bias in which only "significant" results appear in the literature. However, it is still possible that these studies are not representative of the world. To address this issue requires substantially more structure and has thus far been ignored in the meta-analysis literature. This may be because such an exercise requires the development of aggregation techniques that can account for differential types of studies in a more complex way than a simple meta-regression, which fails to share information across the study types.

<sup>5</sup>To investigate the role of take-up in this context, I pursue a bounds analysis explained further in Appendix D.

similarities across settings. In this context, where the generalizability of the evidence across settings is unclear ex-ante, the Bayesian hierarchical framework is an appropriately cautious way to proceed with evidence aggregation.

The open data policies of the *American Economics Journal: Applied Economics* and *Science* allow me access to the microdata from all of these experiments, such that I can standardize which quantiles I compute across studies and can construct each underlying variable in a uniform manner across studies. The variables were measured in different currencies, in different years, and over different time periods (this matters because these are all flow variables). I standardize all measurements to be USD PPP in 2009 dollars over a two-week period. Business variables require further standardization: to capture the potential for microcredit to allow individuals to open new businesses or to switch to operating any existing seasonal businesses throughout the year, households with no business or missing business data have profits imputed as zero. This was the decision made by the original authors of many of the seven studies, because the business creation channel is closely tied to the central claims of Yunus (2006), and dropping the missing values can lead to underestimating the business creation effects of microcredit if there are any. I employ this strategy throughout to business expenditures and revenues as well.<sup>6</sup> I also construct all profit variables in my analysis using reported revenues and expenditures data to minimize recall bias or rounding biases. Other than standardizing the construction of variables as much as possible, I have conformed to the decisions made by the original authors.<sup>78</sup>

Household and study-level covariates may play some role in determining heterogeneity in the quantile treatment effects, but there are limitations to pursuing a covariates analysis in this literature. Only three of the microcredit RCTs collected comprehensive individual-level baseline surveys. One pre-treatment variable was recorded at endline in all studies due to its theoretical importance: a binary indicator that a household had previous experience operating a business (Banerjee et al 2015b). Although covariates at the study level may also predict variation in effects across context, there at least seven such covariates and only seven studies, so conventional regression analysis will be overfitted and misleading. It is still useful to aggregate the evidence without conditioning on covariates, as this permits an understanding how much unconditional heterogeneity there is; if there is little or no variation across settings, further analysis is a less pressing concern for

<sup>6</sup>While it would be ideal to examine effects on other variables such as income and assets, the measurement and definition of those variables differed across the studies to such an extent that it is unclear how to aggregate them. This issue was noted in Meager (2019) and in my pre-registration: <https://osf.io/tdvc8/>.

<sup>7</sup>I have used the entire sample available in the online data sets except in Ethiopia: this study contained a cross-randomized family planning treatment. I use only the pure control and the pure microcredit samples, which is the conservative choice given that we do not know how microcredit interacts with family planning (the study estimates a very imprecise interaction).

<sup>8</sup>I do not winsorize any of the variables because most of the original studies did not winsorize themselves. However Augsburg et al (2015) found that winsorizing outliers sometimes made results statistically significant when they were not significant in the full sample. If the extreme values do not change the point estimate but increase the uncertainty, winsorising them may underestimate the true uncertainty.

future work.

The main analysis that follows is therefore accounting for all sources of heterogeneity including differences in MFI policies, study design, measurement protocols, and contextual factors in the local economies. I do not attempt to separate out these different kinds of heterogeneity both because it is substantially challenging with 7 data points and because it is unclear if they can or should be separated. MFI policies, study design and even measurement protocols may all be endogenously determined by the MFI and researchers in response to the different contextual factors. Indeed, a team of researchers and a single MFI applying the exact same decision rules in which they optimally determine say interest rates or measurement protocols conditional on contextual factors would generically make different decisions in different contexts. Adjusting for the "measurement protocol" before applying the Bayesian model could therefore create differences across settings where none exist in the raw data. I therefore refrain from adjusting for any sources of heterogeneity a priori in the analysis that follows.

## II. Methodology

### A. Bayesian Hierarchical Models

#### HIERARCHICAL MODELS

Consider a body of evidence consisting of  $K$  studies indexed by  $k$ , each of which provides some  $k$ -specific data  $\mathcal{Y}_k$  about a given policy intervention. The  $K$  data sets taken together form one large data set, denoted  $\mathcal{Y} = \{\mathcal{Y}_k\}_{k=1}^K$ . Each study setting has a site-specific parameter of interest  $\theta_k \in \Theta_k$ , which could be univariate (e.g. the average treatment effect), or multivariate (e.g. the entire set of quantile treatment effects). The full data in each site  $k$  consists of  $N_k$  households, summing to  $N$  households in the total combined sample of all settings.

Suppose that the analyst wishes to learn about the expected value of these  $\{\theta_k\}_{k=1}^K$  parameters across different settings, and to learn about the uncertainty she should have due to unobserved differences across settings. This implies a need for inference on the average value of these parameters, sometimes called the "hypermean" and denoted  $\theta = E[\theta_k]$  with the expectation taken across the study settings. Accounting for uncertainty across settings permits inference not just to unobserved households in the existing sites but to unobserved study sites themselves, and thus, such a procedure could generate inference applicable beyond the current set of studies to future policy settings if such extrapolation is warranted.

One can learn about  $\theta$  using the evidence on  $\{\theta_k\}_{k=1}^K$ , but the optimal learning procedure depends on the heterogeneity or dispersion of  $\{\theta_k\}_{k=1}^K$  around  $\theta$ , denoted  $\Sigma_\theta$  (Rubin 1981, Gelman et al. 2004). This  $\Sigma_\theta$  describes the strength or weakness of the relationship between any  $\theta_k$  and the general parameter  $\theta$ : if the dispersion  $\Sigma_\theta$  is small, then  $\theta_k$  is close to  $\theta$  and provides a strong signal of the value of  $\theta$ . By the same logic, when  $\Sigma_\theta$  is small then  $\theta$  is a strong predictor of

$\theta_{K+1}$  for some future setting.<sup>9</sup> Here  $\Sigma_\theta$  parameterizes a notion of generalizability of the evidence contained in  $\mathcal{Y}$  to external settings, which captures the definition of external validity in Allcott (2015) and Dehejia, Pop-Eleches and Samii (2015). If  $\Sigma_\theta = 0$ , then  $\theta$  is a perfect predictor of  $\theta_{K+1}$ ; if not, there will be some extrapolation error which grows large as the parameter  $\Sigma_\theta$  grows large.

Joint estimation of  $\theta$  and  $\Sigma_\theta$  is the core challenge of aggregation across studies. Hierarchical models approach this problem by defining a set of parameters at the site level,  $\{\theta_k\}_{k=1}^K$ , a set of parameters at the population level,  $\theta$ , and a relationship between them (Efron and Morris 1975, Rubin 1981, Gelman et al. 2004). The "lower level" of the model describes the dependence between the data and local parameters in site  $k$ :

$$(1) \quad \mathcal{Y}_k \sim f(\cdot|\theta_k) \forall k.$$

The "upper level" of the model describes the potential for statistical dependence between local parameters and general parameters (also called "hyperparameters") via some likelihood function  $\psi(\cdot)$ , which contains the hypervariance  $\Sigma_\theta$  either implicitly or explicitly depending on the specific model. While  $\psi(\cdot|\theta, \Sigma_\theta)$ , this second argument is often implicit and thus notationally suppressed. This upper level "general" or "parent" distribution is then denoted:

$$(2) \quad \theta_k \sim \psi(\cdot|\theta) \forall k.$$

A hierarchical likelihood contains both levels:

$$(3) \quad \mathcal{L}(\mathcal{Y}|\theta) = \prod_{k=1}^K f(\mathcal{Y}_k|\theta_k) \psi(\theta_k|\theta).$$

This likelihood structure nests both the "no pooling" case in which no information is shared across settings, and the "full pooling" aggregation case in which all information is shared fully across settings as if they are indistinguishable. This nesting is possible because the hyperparameters that govern the  $\psi(\cdot)$  function, including  $\Sigma_\theta$ , are estimated rather than imposed. For example, the model may estimate that  $\theta_k \approx \theta_{k'} \forall k, k'$ , and hence that  $\Sigma_\theta = 0$ , if that is supported by the data. This result would recover the full-pooling model's solution, up to a degrees of freedom correction.<sup>10</sup> Of course, the model could detect large dispersion in  $\{\theta_k\}_{k=1}^K$ ; in that case it recovers the no-pooling model's solution. In fact, this model can recover a solution anywhere on the spectrum between these two

<sup>9</sup>Technically the sites must be "exchangeable", this condition is discussed in the final paragraph of this section.

<sup>10</sup>This discussion is in reference to models for estimating the average effects across the different settings, with uncertainty intervals referring to extrapolation to an unobserved setting. The full pooling model has another function: it can serve to estimate the average effect across all the individuals in the existing settings, with uncertainty intervals referring to an extrapolation to an unobserved individual within these given settings. This distinction is discussed further in the results section.



extremes if that intermediate or "partial pooling" solution is most supported by the data. Inference on  $\{\theta_k\}_{k=1}^K$ ,  $\theta$  and  $\Sigma_\theta$  is influenced by the extent of this partial pooling, which is also called "shrinkage" because the  $\{\theta_k\}_{k=1}^K$  estimates are somewhat "shrunk" together when information is shared across settings.

Hierarchical models require that  $\{\theta_k\}_{k=1}^K$  be "exchangeable", such that their joint distribution is invariant to permutation of the indices (Diaconis, 1977). This means the analyst must be ignorant of any definite ordering, dependence or sub-clustering of the parameters *a priori*; this is the case for the microcredit effects as in most applications to social science. Knowledge of contextual differences typically does not inform us *ex ante* of how  $\theta_k$  parameters differ (Rubin 1981).<sup>11</sup> The extensive theoretical literature on microcredit markets does not dictate a specific ordering of effects in this case, which makes exchangeability a reasonable structure (Gelman et al. 2004). Any future site for which  $\theta_{K+1}$  is used to predict the effect must be exchangeable with the sites in the sample; this is a general requirement for any out-of-sample prediction (see for example Allcott 2015).

#### BAYESIAN IMPLEMENTATION

Bayesian inference offers several advantages for the analysis of hierarchical models, particularly when aggregating across a relatively small number of studies, as is typically the case in economics. With seven microcredit studies, estimation of the variation in effects across settings is performed off only seven contexts. Given this low-data environment, incorporating informative Bayesian priors can substantially improve the performance of the hierarchical model by "regularizing" the likelihood problem in the classical sense: introducing new information that constrains the fit of the model, typically trading off bias for variance to reduce overall mean squared error and out-of-sample prediction (Hastie, Tibshirani and Friedman 2009, section 10.2; Chung et al. 2013, 2015).

The inference proceeds by specifying a prior on the unknowns at the highest level,  $\mathcal{P}(\theta)$ , and combining it with the likelihood via Bayes' rule to generate the joint posterior distribution  $f(\theta|\mathcal{Y})$ . The specification of a proper prior distribution ensures that  $f(\theta|\mathcal{Y})$  is a proper probability distribution with desirable decision-theoretic properties such as consistency and admissibility (Berger 2013, Van der Vaart 1998, Efron 1982).

The Bayesian approach also automatically delivers the correct marginal distribution of the treatment effect in a hypothetical future site  $\theta_{K+1}$ . This is often the object of most interest for policymakers, but the distribution of this object must account for the full joint posterior uncertainty on the hyperparameters rather than conditioning on a particular point estimate.<sup>12</sup> The Bayesian approach de-

<sup>11</sup>If an established and verified economic theory dictates that a particular covariate can only be correlated in a certain way with the treatment effects, that can be translated into *conditional* exchangeability by introducing this covariate into the model. It is also possible to build a more complex structure that allows "partial exchangeability" if this is desired; see Albert and Chib (1997) for a discussion of this approach.

<sup>12</sup>The full joint posterior uncertainty accounts perfectly for the uncertainty about how well the new

livers the correct uncertainty interval conditional on the model in the form of posterior predictive inference (Gelman et al. 2004), which averages over the posterior uncertainty on the unknowns  $(\theta, \Sigma_\theta)$ . Formally, the posterior predictive distribution is:

$$(4) \quad f(\theta_{K+1}|\mathcal{Y}) = \int \psi(\theta_{K+1}|\theta) f(\theta|\mathcal{Y}) d\theta$$

The characterization of the full joint posterior distribution on all parameters is not without costs: it can exacerbate the already-formidable tractability issues inherent in hierarchical models. The relationships between  $\theta$ ,  $\Sigma_\theta$  and  $\{\theta_k\}_{k=1}^K$  are highly nonlinear, leading to challenges optimizing and even characterizing the shape of the likelihood function. Common maximum likelihood approaches estimate the upper level first and then condition on the point estimates using the "empirical Bayesian" approach from Efron and Morris (1975). This ignores the uncertainty about the upper level parameters,  $\theta$  and  $\Sigma_\theta$ , when computing uncertainty intervals on the lower level parameters, and thereby systematically underestimates uncertainty (Rubin 1981).<sup>13</sup> To surmount these tractability issues in the Bayesian setting I turn to Markov-Chain Monte Carlo methods; in particular, I use Hamiltonian Monte Carlo (HMC) methods which are well-suited to the peculiar geometry introduced by hierarchical structure (Betancourt and Girolami, 2013).<sup>14</sup>

### B. Tailored Mixture PDF Models for Distributional Effects

#### CHALLENGES OF DISTRIBUTIONAL EFFECT AGGREGATION

Applying the Bayesian hierarchical framework to distributional effects aggregation poses additional challenges which demand new modelling approaches. If one were interested in only a single quantile treatment effect, such as the median, one could directly apply a Rubin-style model to the quantile as long as the data meets the condition for the Mosteller (1946) theorem providing consistency and asymptotic Normality of the sample quantile estimate. However, for the micro-credit data, even the estimation of quantiles themselves is not straightforward because many households outcomes are not continuously distributed (as required by the theorem), but instead have point masses of households at zero. The sampling variance of the quantiles is no longer Gaussian, and may even be estimated

location matches the old location, if the new site is exchangeable with the old sites, and the model structures are correct. If these conditions do not hold, we have modeling uncertainty, which is not accounted for in any meta-analytic methods at present (nor in any of the popular analytic methods used in empirical economics).

<sup>13</sup>While MLE methods that do not condition on point estimates of unknowns are theoretically available, they seem to be largely unused in practice.

<sup>14</sup>HMC uses discretized Hamiltonian dynamics to sample from the posterior, which can be combined with the No-U-Turn sampling method (NUTS) to auto-tune the step sizes in the chain (Hoffman and Gelman, 2014). This algorithm is automated in the software package Stan, a free statistical library which calls C++ to fit Bayesian models from R or Python (Stan Development Team, 2017).

as zero in the finite sample (as in Angelucci et al 2015). Even if researchers are willing to accept zero estimated sampling error, this causes difficulties for a Rubin-style approach, which constructs weighted averages of estimated effects in which the *inverse* standard error enters the weight.<sup>15</sup>

Moreover, aggregating sets of quantiles requires facing the constraint that the true quantiles must be monotonically increasing, because CDFs are monotonically increasing by definition. Violating this constraint leads to the "quantile crossing" problem, which can arise even in simple analyses (see for example Chernozhukov, Fernandez-Val and Galichon 2010). Aggregation exacerbates this problem because most methods both estimate the aggregate effect and revise the site-specific effect estimates by incorporating information on the local variation or sampling uncertainty, and this can be different at different quantiles across the different sites. But differentially-weighted averages of monotonic functions need not themselves be monotonic, so the aggregation procedure could introduce quantile crossing where none was present in the original studies.<sup>16</sup>

#### TAILORED MIXTURE PDF APPROACH

I address both of these problems by directly modelling the distributions of these outcome variables using a flexible set of mixture probability density functions (PDFs). These models specify treatment effects on the parameters of the tailored density function and partially pool information on these effects across studies by placing a hierarchical structure on these treatment effects. The conceptual approach is here is quite general, and even the specific models below are likely to apply beyond the microcredit application. In every case, however, building these parametric models relies on contextual economic knowledge of the variables at hand, and researchers must understand the data generating process well enough to know when it ought to exhibit, say, discreteness, skewness or kurtosis. The incorporation of this specific contextual knowledge is critical to the modelling process, as follows.

Consider the household business variables in the microcredit data, generated by surveying a random sample of households. In most contexts, not all households will choose to own businesses, nor will they necessarily operate businesses they do have at all times, as they often do seasonal agricultural or manual labour for part of the year (Macours and Vakis 2010, Gibson and McKenzie 2014, Fink,

<sup>15</sup>It might be possible to overcome this problem by investigating CDF effects rather than quantiles, but applied researchers are directly interested in quantiles, and the transformation between CDFs and quantiles is not straightforward in the partially discrete data case.

<sup>16</sup>While it is possible to derive an aggregation model for sets of quantiles based on the Mosteller approximation that respects the logical properties of the estimands via variable transformation (see Appendix A), even this model would only be applicable to data for which the asymptotic theorem applies. In some cases of discrete data, such as with count data, it is possible to "dither" or "jitter" the discrete data to produce a new, continuous distribution for which the quantiles have a one-to-one relationship with the discrete distribution (Machado and Santos Silva, 2005). This does not work in the case of a partially discrete "spike and slab" distribution because the dithered data points from the spike generally leapfrog some of the data in the slab, destroying the one-to-one relationship between the quantiles.

Jack and Masiye, 2014). So measured business variables should be mixture of a discrete mass at zero and continuous tails, because these variables are the output of a partially discrete decision process. First, a household has an extensive margin decision to make about whether to start a business, and then they face a secondary extensive margin decision about whether to operate that business in a given season. Only those households who decide to open and operate their businesses go on to make an intensive margin decision, the result of which manifests some continuous expenditures, revenues and profit. In fact, this pattern also arises in consumer durables spending and temptation goods spending, to a lesser extent.

I therefore use a set of mixture models to model this partially-discrete data, each of which has a point mass or "spike" of households at zero and a continuous distribution or "slab" on the rest of the real line. Economic theory and prior research further suggest that the continuous portions of business variables such as revenues and profit will tend to follow power laws or other fat-tailed laws (Roy 1950, Stiglitz 1969, Gabaix 2008, Allen 2014, Piketty 2015, Bazzi 2016). Hence, the outcome PDF can be modeled as a mixture of three distributions: a lower tail, a spike at zero, and an upper tail. Modeling the two tails separately allows for skewness as well as differential kurtosis or variance. As the binary treatment indicator variable  $T_{nk}$  may affect the mass in the components and the shape of the tail components, I specify treatment effects on all aspects of this mixture PDF. The model can then aggregate the effect of the treatment on each of the parameters that govern the distribution, as well as the implied quantile treatment effects. Because these particular mixture models have components with disjoint supports, I can further analytically compute and report the quantile treatment effects employing the method of Castellaci (2012).

Yet the risk in specifying any parametric structure based on contextual and prior information is that our knowledge may be insufficient or incorrect in a manner that leads to poor inference. It is advisable therefore to assess the sensitivity to the choice of functional form, as well as to assess model fit and avoid reliance on models that fail to approximate the data well. In the case of business variables the distribution of the tails could reasonably be modeled by a Pareto distribution, as in Piketty 2015 or Bazzi 2016. However, a lognormal distribution would allow for more mass near the lower bound of the distribution per Roy 1950 and is analogous to log transforming the positive values in the sample, a common practice in applied microeconomics (see for example Banerjee et al 2015b). The lognormal model both fits the data better in this case and is more tractable to use, as it always has all moments (see figure ??model fit and Appendix C for more details ).<sup>17</sup>

<sup>17</sup>While a nonparametric model such as an infinite mixture of Gaussians implemented via a Dirichlet Process prior would be maximally flexible, the parameters of such models are not identified when the outcome data are univariate (Compiani and Kitamura 2016, Kasahara and Shimotsu 2014). Even for a bivariate outcome, on which one can put a lower bound on the number of components, the component distributions themselves are not identifiable (Hall and Zhou 2003). Indeed, the Dirichlet Process mixture

## MIXTURE LOGNORMAL AND PARETO MODELS FOR MICROCREDIT DATA

The above discussion leads to the following tailored hierarchical PDF model to aggregate the quantile effects on household business profit. Denote the probability mass in the  $j$ th mixture component for a household  $n$  with treatment status  $T_{nk}$  to be  $\Lambda_j(T_{nk})$  for  $j = 1, 2, 3$ . This dependence can be modelled using a multinomial logit specification, denoting the intercept in site  $k$  for mixture component  $j$  as  $\alpha_{jk}$  and the treatment effect as  $\pi_{jk}$ . For the spike at zero, the Dirac delta function can be used as a distribution, denoted  $\delta(x)$  for a point mass at  $x$ . If using the lognormal distribution for the tails, then each are governed by a location parameter and a scale parameter. The latter can only be positive valued so I employ the exponential transform to ensure the support constraint is satisfied. I model the location parameter using a linear regression format in which the value for the control group in site  $k$  is  $\mu_k$  and the value for the treatment group is  $\mu_k + \tau_k$ . The scale parameter is modeled with the control group's value being  $\exp(\sigma_k^c)$  and the treatment group's value being  $\exp(\sigma_k^c + \sigma_k^t)$ .

The lower level of the likelihood  $f(\mathcal{Y}_k|\theta_k)$  is specified according to this mixture distribution. Let  $j = 1$  denote the negative tail of the household profit distribution, let  $j = 2$  denote the spike at zero, and let  $j = 3$  denote the positive tail. Then the household's business profit is distributed as follows:

$$\begin{aligned}
 y_{nk}|T_{nk} &\sim \Lambda_{1k}(T_{nk})\text{Lognormal}(-y_{nk}|\mu_{1k} + \tau_{1k}T_{nk}, \exp(\sigma_{1k}^c + \sigma_{1k}^t T_{nk})) \\
 &\quad + \Lambda_{2k}(T_{nk})\delta_{(0)} \\
 &\quad + \Lambda_{3k}(T_{nk})\text{Lognormal}(y_{nk}|\mu_{3k} + \tau_{3k}T_{nk}, \exp(\sigma_{3k}^c + \sigma_{3k}^t T_{nk})) \quad \forall k
 \end{aligned}
 \tag{5}$$

where  $\Lambda_{jk}(T_{nk}) = \frac{\exp(\alpha_{jk} + \pi_{jk}T_{nk})}{\sum_{j=1,2,3} \exp(\alpha_{jk} + \pi_{jk}T_{nk})}$

The upper level  $\psi(\theta_k|\theta)$  is:

$$(\alpha_{1k}, \alpha_{2k}, \alpha_{3k}, \pi_{1k}, \dots)' \equiv \zeta_k \sim N(\zeta, \Upsilon) \quad \forall k
 \tag{6}$$

The Gaussian at the upper level aligns with Rubin (1981) and Meager (2019), and can provide good estimates of the hypermean and hypervariance even if it is misspecified (McCulloch and Neuhaus, 2011). For tractability and simplicity I enforce diagonal  $\Upsilon$  for the microcredit analysis. This prevents the model from using correlations in the distribution of say  $\{\alpha_{1k}\}_{k=1}^K$  and  $\{\alpha_{2k}\}_{k=1}^K$ , a restriction which has practical value because the correlations are hard to estimate with seven sites and this can introduce substantial additional variance into the estimation

model cannot even consistently identify the number of components when it is itself the ground truth model (Miller and Harrison, 2013). The lack of identification does not impede curve-fitting, yet creates challenges for shrinkage on treatment effects when one conceives of these effects as operating on the underlying parameters, since they are not identified in this case. As shrinkage is a nonlinear operation in the hypervariances, the hierarchy must be applied directly to the object of interest (see Appendix B for a short proof).

procedure. This independence restriction can be thought of as a form of particularly strong discrete or dogmatic regularization on these correlations. While strong continuous regularization is generally preferable to discrete or dogmatic regularization, the latter in this case is necessary to ensure tractability given the limitations of computational power I face (and is aligned with econometric tradition, as frequentist models in economics rarely exploit correlations in parameters). The remaining priors  $\mathcal{P}(\theta)$  as follows:

$$\begin{aligned}
 (7) \quad & \zeta \sim N(0, 10) \\
 & \Upsilon \equiv \text{diag}(\nu_{\Upsilon}) \Omega_{\Upsilon} \text{diag}(\nu_{\Upsilon})' \\
 & \nu_{\Upsilon} \sim \text{halfCauchy}(0, 5) \\
 & \Omega_{\Upsilon} = I_{|\zeta|} \\
 & \alpha_{mk} \sim N(0, 5).
 \end{aligned}$$

The model above uses contextual information and balances concerns about flexibility and tractability. It may be desirable to fit a fatter tail model such as the Pareto, or a more flexible tail such as the double-Pareto-lognormal likelihood model of Reed and Jorgensen (2004). I do both in the remainder of this paper; one needs only substitute these densities for the Lognormal in the equations above and then fit the resulting model. However, as I discuss in my results section, the Lognormal fits best. Indeed, the more flexible joint tail model can face substantial convergence issues: Reed and Jorgensen (2004) notes that if there is power law behavior only in one tail, the EM algorithm is "unlikely to converge". Keeping this issue in mind, I report results of a version of the above model using a Pareto-lognormal tail only out of interest.<sup>18</sup> <sup>19</sup> To fully avoid convergence issues while benefitting from the flexibility of two functional forms, I also fit a version of this model which manually separates the tail data at 80th extremal percentile, treating the data before this point as lognormal and the data beyond it as Pareto.<sup>20</sup> See Appendix C, section 2.1, for the full models and technical details of these two alternative approaches as well as several caveats on their use.

#### RECOVERING QUANTILE EFFECTS FROM MIXTURE DISTRIBUTION MODELS

Quantile recovery is nontrivial in this setting because mixture distributions in general do not have analytical quantile functions. However, because the mixture

<sup>18</sup>In practice I found this to be because the simple lognormal case is nested by this model only when certain parameters are infinity, which leads to wandering behavior: even a single Pareto-lognormal tail may have convergence issues if the power law behavior is sufficiently similar to the lognormal's tail behavior. The convergence issues in this case can be somewhat mitigated by strong priors, in particular by placing a strong prior on the  $\alpha$  parameter from Reed and Jorgensen (2004) to force it to take a moderately sized value and mitigate the "wandering" behavior noted in the paper, and by replacing the half Cauchy on  $\nu_{\Upsilon}$  with a half Gaussian.

<sup>19</sup>I thank Dr Michael Betancourt in particular, as well as Dr Ben Goodrich and Professor Aki Vehtari, for their advice and assistance with this problem. A public record of our work can be found here <https://discourse.mc-stan.org/t/double-Pareto-lognormal-distribution-in-stan/10097/20>.

<sup>20</sup>I thank Ulrich Müller and Andriy Norets for this suggestion.

distribution in this particular model has components with disjoint supports, one can apply the method of Castellacci (2012) to compute the quantiles analytically in this case.

Castellacci's method is based on the observation that mixture distribution with disjoint components are decomposable functions, and that thus one can invert such functions by piecing the component inverses together. Thus, while in general the quantiles of mixtures would need to be estimated numerically, when the mixture components have disjoint supports the CDF is decomposable and thus it is possible to analytically derive the quantiles. In fact, in my case the supports are not just disjoint but ordered, making the calculation relatively straightforward.

Thus Castellacci (2012) derives the key result (equation 2.3 of his Proposition 2.1) for the CDF of the mixture distribution  $F_w(x) = w_1 G_1(x) + w_2 G_2(x) + \dots + w_n G_n(x)$  when the component distributions  $G_i(x)$  have disjoint, ordered supports: (8)

$$F_w^{-1}(p) = G_1^{-1}\left(\frac{p}{w_1}\right) \mathbb{1}_{B_1} + G_2^{-1}\left(\frac{p - w_1}{w_2}\right) \mathbb{1}_{B_2} + \dots + G_n^{-1}\left(\frac{p - \sum_{i=1}^{n-1} w_i}{w_n}\right) \mathbb{1}_{B_n}$$

where  $B_i := \left[ \sum_{j=0}^{i-1} w_j, \sum_{j=0}^i w_j \right)$  for  $i = 1, \dots, n-1$

$$B_n := \left[ \sum_{j=0}^{n-1} w_j, 1 \right]$$

Proceeding with this general formula, I make only one adjustment to it: because the negative tail of profits is always modeled using a "reverse" of a distribution with positive support, if one wishes to compute the  $u$ th quantile of the lower tail one needs the negative value of the  $(1 - u)$ th quantile of the positive version of the distribution.

Given the lognormal profit model above I derive the following parametric quantile function using the method above, noting that the only difference from the general Castellacci (2012) formula is the use of the reverse quantile for the first component due to it being a reverse distribution:

$$(9) \quad Q(u) = -\text{Lognormal}^{-1}\left(1 - \frac{u}{\Lambda_1(T_n)} \mid \mu_{1k} + \tau_{1k}T_{nk}, \exp(\sigma_{1k}^c + \sigma_{1k}^t T_{nk})\right) * \mathbb{1}\{u < \Lambda_1(T_n)\} \\ + 0 * \mathbb{1}\{\Lambda_1(T_n) < u < (\Lambda_1(T_n) + \Lambda_2(T_n))\} \\ + \text{Lognormal}^{-1}\left(\frac{u - (1 - \Lambda_3(T_n))}{\Lambda_3(T_n)} \mid \mu_{3k} + \tau_{3k}T_{nk}, \exp(\sigma_{3k}^c + \sigma_{3k}^t T_{nk})\right) * \mathbb{1}\{u > (1 - \Lambda_3(T_n))\}$$

Any disjoint mixture model's quantiles can be computed analogously. The posterior distribution of the entire set of quantiles and thus the implied quantile treatment effects is easily computed from the posterior distribution of the

unknown parameters by applying the computation above to every MCMC draw from the joint posterior distribution. This method ensures that the uncertainty on the quantiles implied by the uncertainty on the parameters that govern the model is translated exactly.

### III. Results

#### A. Main Results

The expected effects of access to microcredit on the quantiles of each of the six outcomes produced by fitting the tailored hierarchical mixture model with lognormal tails are shown in figure ??posterior general quantiles. I focus on profit and consumption as the key outcomes of interest, recalling that while profit is recorded in all studies, consumption is recorded in only five. The full details of the site-specific results for household consumption and profits are shown in tables ??big consumption table and ??big profit table respectively. The pattern across all variables is similar: there is a reasonably precise zero impact from the 5th quantile to the 75th quantile, after which point there is a large and highly uncertain effect in a "desirable" direction (business variables and consumption increase, temptation spending decreases). The posterior probability that the quantile treatment effects are equal across the quantiles is extremely small; in the case of profit, for example, the formal posterior probability that the quantile effects are within 50 US cents (PPP) of each other is less than 1%. There is no evidence of harm on average at any quantile, contrary to the fears that microcredit systematically causes harm to communities which receive it.

To understand why this pattern emerges, it is useful to consult the detailed results in tables ??big consumption table and ??big profit table respectively. As well as the country-specific results from the partial pooling Bayesian model, I also show the results of running a "no pooling" analysis country by country, and of performing a full pooling analysis (as in Banerjee et al 2015c). In the case of consumption, the no pooling and full pooling models can be computed simply using the canonical Koenker and Basset 1978 estimator, while the partial pooling results are generated by the tailored parametric model; they are nevertheless remarkably similar and tell a coherent story of zero effects along most of the distribution with noisy positive effects on the right tail. For profit, all results are generated via the lognormal model, but in this case full pooling makes quite a difference – it reports a much larger and more significant effect than the partial pooling models. This is driven largely by Mexico and the Philippines with their large samples; as they nevertheless provide quite imprecise estimates, the partial pooling model puts less weight on them than the full pooling model.

To some extent, one should expect the results of the full pooling model to differ from the partial pooling model because in general they are concerned with different estimands. The partial pooling model estimates the average effects across the different settings, with uncertainty intervals referring to extrapolation to an



unobserved setting. The full pooling model estimates the average effect across all the individuals in the existing settings, with uncertainty intervals referring to an extrapolation to an unobserved individual within these given settings. These estimands will differ except in the case that the full pooling model is literally correct.<sup>21</sup>

Conceptually, to gain a general understanding of what a particular intervention will do across settings rather than across individuals within settings, the relevant exercise is not to quantify the average effect but to predict the differences in the distributions in a new setting and report the associated uncertainty. To answer this question, I now turn to the posterior predicted quantile results, which provide inference on the effects in the next comparable study location. The results are shown in figure ??posterior predicted quantiles. The predicted effect for the majority of the distribution remains a precise zero, but above the 75th percentile there is even greater uncertainty about the exact impact microcredit will be likely to have on the right tail of the next distribution to which it is applied. The estimates in the upper tails are so imprecise that the estimates look as if they are zero when they are graphed at this scale, even though they are generally quite large (as tables tables ??big consumption table and ??big profit table show). Essentially, the models decline to make any generalizable prediction for these effects across settings.

This analysis further reveals that the business variables in this sample exhibit extreme kurtosis; the tails of these distributions are very heavy. The positive tail of profit, which is less heavy than that of revenues and expenditures, has a log scale parameter of 1.25, indicating an excess kurtosis of approximately 811 in the lognormal model (see calculations in Appendix C). Figure ??model fit shows that the model's predicted tails are actually somewhat thinner than the real data, so if anything this is an underestimate of the true kurtosis. As shown in figure ??complex quantiles this is further supported by the results of the more flexible models, incorporating either the Pareto-lognormal likelihood in the tails or a composite tail model in which the lognormal is fit below the 80th percentile and the Pareto beyond this point, which show even more uncertainty in the tails. Consumption, by comparison, has a log scale parameter of 0.65, and thus an excess kurtosis of 14 in the model. For reference, the standard Laplace distribution has an excess kurtosis of 3, yet even in that case the sample median is more efficient than the sample mean as an estimator of the location parameter (Koenker and Bassett 1978).<sup>22</sup> This suggests that the average treatment effects estimated via OLS regression in the original studies and thus the analysis in Meager (2019) may

<sup>21</sup>If there is no heterogeneity across settings, or "external validity problem", the full pooling model also estimates the average effect across settings and the correct uncertainty, because an unseen individual in an unseen setting is no different from an unseen individual in a setting we do see.

<sup>22</sup>The Pareto models fit to the business data find scale parameters close to zero, indicating that the kurtosis is undefined or infinite. It seems likely that the tails of these business variables are heavy enough to impede the functioning of the central limit theorem and even the law of large numbers (see Appendix C).

be unreliable for these variables, both because they invoke Gaussian asymptotics which do not hold, and because in this case the sample mean is not a reliable indicator of the underlying distribution's location parameter (Koenker and Basset 1978, Koenker and Hallock 2001).

### *B. The role of business experience*

Previous studies including Banerjee et al (2015b) suggested that a household's previous business experience may explain some of the evident variation in treatment effects. Meager (2019) investigated the role of this covariate across all the studies, and found that there were notable differences in impacts between this group and the other households without business experience, although there was still no strong evidence of a generalizably positive average impact for any group. However, as the main results above indicate that these means are composed of precise zeroes and noisy positive results in the tails, a quantile analysis may be able to shed further light on this question.

To assess the role of previous business experience in determining the distributional treatment effects, I split the entire sample by a binary indicator of prior business ownership, denoted  $PB = 1$  if the household does have such experience and 0 otherwise, and separately analyze the two subsamples. Fitting the Bayesian hierarchical quantile aggregation models to each group shows that the impact of microcredit differs somewhat across the two types of households, but to a lesser extent than was suggested in Meager (2019) and Banerjee et al (2015b). Figures ??consumption posterior general quantiles pb split and ??posterior mean quantiles profit pb split show the general distributional impact of microcredit on the six household outcomes of interest for each of the household types.

There does seem to be a general pattern of larger impacts in the tails of the group which does have previous business experience, as well as greater uncertainty surrounding these estimates. The posterior probability that the impact for the two groups is the same at each quantile is reasonably high for the middle quantiles, but fall below 2% at the 95th percentile even at high tolerances, as shown in table 3. The probability that the two groups' quantile effects are within 5 USD PPP of each other are only 10% at the 95th quantile. However, as there is substantial overlap in the posterior inference for the two groups and particularly in the posterior predictive effects, these results do not necessarily provide a rationale for targeting those with prior business experience in future settings.

## **IV. Discussion**

The aggregated distributional effects show no evidence that access to microcredit causes any negative shifts in the distribution of household outcomes. While moderately negative impacts are within the 95% posterior interval of the effects on the upper tails of most of the distributions, the point estimate and vast majority of the posterior mass is positive in those cases. The only variable with

larger uncertainty at the lower tail is profit, but the point estimate is zero and the uncertainty is symmetric around that point. This provides evidence against the notion articulated by some critics that microcredit causes substantially worse outcomes for some group of households than they would have experienced in its absence (Schicks 2013). While a lack of negative quantile effects does not imply that no household experiences any harm from microcredit, it does imply that any households who do experience harm are approximately canceled out by other types of households who experience benefits, such that these two types of households are swapping ranks in the outcome distribution rather than contributing to any change in the shape of that distribution. Each population in each of the studies taken as a whole however does not experience any systematic worsening of its economic outcomes.

The precise zero effect from the 5th to 75th percentile of most of the household outcomes appears to be a true zero and not a mechanical artifact of the spike at zero nor an economic consequence of the low takeup. Consumption does not exhibit a spike of households who record an outcome equal to zero yet microcredit still has a precisely estimated zero effect for most of the distribution. Even for profit, revenues and consumption the spike only accounts for at most 50% of the outcome distribution, yet the zero effect applies to 75% of the distribution. Similarly, regarding concerns about takeup: Bosnia and the Philippines had over 90% takeup and yet still exhibited zero effects from the 5th to 75th percentile in consumption and profit (see tables ??big consumption table and ??big profit table respectively). Concerns that the small scale of profits in Mongolia, or lack of negative profits in Bosnia, may be affecting the results are addressed in Appendix C by re-running the models without these sites, which largely confirms the main results. The bounding exercise in Appendix D aggregates all the data on the question of takeup and shows that even the effects on the outcome distribution for those who take up microloans are likely to be zero along most of the distribution.

What are the economic consequences of potentially increasing the right tail of consumption and business outcomes while leaving the rest of the distribution unchanged? This pattern means that expanding access to microcredit is likely to cause an ex-post increase in economic inequality across households, which may be important if inequality leads to capture of local political institutions or other adverse social consequences (Acemoglu and Robinson 2008). A rightward expansion of the upper tail does not mean that the richer households are getting richer, because quantile effects cannot be localized to any particular households without invoking a rank invariance assumption or some comparable structure (which is unrealistic for credit market interventions). More detailed baseline data may have permitted an exploration of this question, although such households may well look identical to others along all the covariates we can measure (as suggested in Kaboski and Townsend 2011).

This pattern of probable expansion in the right tail, combined with the inability to localize the effects to particular households in these data sets, highlights

the potential value of locating and studying these highly productive individuals. Studies such as Hussam, Rigol and Roth 2017, which leverages local knowledge to lend to borrowers with high marginal returns to capital, are valuable both because these individuals seem to be the only households positively benefiting and because the benefits are large. My analysis suggests that we cannot necessarily expect the results observed in such papers to replicate elsewhere, and there may well be contexts in which these positive tail effects will not materialize. However, my analysis also demonstrates the challenges of inference on these highly productive households because their returns follow heavy-tailed distributions. Under such circumstances, studies that appear to be well-powered may be underpowered to detect these effects, giving further reason to emphasize aggregated results rather than individual studies.

The presence of heavy tails also has econometric implications: in these populations, certain individual households account for large percentages of the total business activity. It may be challenging to understand the economies of developing countries if we trim or winsorize the most productive households who make up a large percentage of total economic activity. It might be more useful to study mechanisms that can produce fat-tailed outcomes, such as multiplicative production functions, experimentation or investments with a relatively high risk exposure and long maturation horizons.<sup>23</sup>

## V. Conclusion

The microcredit results demonstrate the value of using appropriate methodology for the question of interest, rather than a "default" approach restricting oneself to conditional means. The models developed in this paper could be used to study the distributional effects of other financial interventions, trade and innovation policies, educational subsidies, and local migration incentives, all of which have social welfare implications (Borusyak and Jaravel 2018, Duflo, Dupas and Kremer 2017, Chetty, Hendren, and Katz 2016, Bryan, Chowdhury and Mo-barak 2014). There are many settings in which quantiles are implicated in policy directly, often because taxation and welfare policies are explicitly made with reference to quantiles of the income distribution; in such settings, the conditional mean often contains little information about the policy's impacts. The models provided here are especially relevant when the data contains discrete spikes of individuals who record certain outcome values due to their extensive margin decisions. In these cases, aggregation of distributional effects within the Bayesian hierarchical framework may be both more informative and more reliable than individual analyses of average treatment effects alone.

<sup>23</sup>One might ask whether if microcredit interventions were studied over a 10 or 20 year horizon the imprecise tail effects we observe after two years could either become precise or could lead to benefits across the entire distribution, but this data set does not contain the answer to that question.

TABLE 1—CONSUMPTION: COMPARISON OF NO POOLING, PARTIAL POOLING AND FULL POOLING RESULTS

Quantile:	5th	15th	25th	35th	45th	55th	65th	75th	85th	95th
<b>No Pooling</b>										
Bosnia	-5.2 (-11.1,0.8)	-7.1 (-16.9,2.7)	-4.7 (-17.5,8.1)	-7.7 (-22.8,7.3)	4.1 (-13.1,21.4)	0.9 (-20,21.7)	-16.3 (-46.2,13.6)	-34.4 (-74.9,6.1)	-64.5 (-131.1,2.2)	104 (-77.4,285.5)
India	0.2 (-5.9,6.3)	-1 (-6.3,4.3)	-2.3 (-8.3,3.8)	-1.2 (-7.4,4.9)	-1.4 (-8.3,5.6)	-2.6 (-10.1,4.8)	2.2 (-6.3,10.7)	4.6 (-6.3,15.6)	8.2 (-7.5,24)	40.1 (-4.5,84.7)
Mexico	-9.3 (-13.7,-4.9)	-1.5 (-5.9,2.9)	-2 (-6.2,2.2)	-2 (-6.5,2.5)	-0.5 (-6.5)	4.1 (-1.6,9.7)	5.5 (0.11)	11 (2.7,19.2)	13.2 (1.8,24.7)	16.6 (-6.7,39.9)
Mongolia	12.3 (-7.31,5)	6.5 (-9.4,22.4)	5.1 (-13.6,23.7)	8 (-13.7,29.8)	-8.2 (-38.7,22.2)	0.8 (-18.5,20.1)	-0.9 (-32.7,30.9)	-2.5 (-42.1,37.1)	-12.8 (-70.3,44.8)	87.4 (-40.6,215.4)
Morocco	1.6 (-6.2,9.3)	5.3 (-0.9,11.5)	4.1 (-2.8,10.9)	-1.1 (-7.3,5.1)	-2.6 (-10.1,4.9)	3.6 (-4.3,11.5)	3.7 (-7.3,14.7)	0.4 (-12.4,13.2)	-6.4 (-23.8,11)	-54 (-104,-4)
<b>Partial Pooling</b>										
Bosnia	-5.2 (-9.6,2.6)	-3.7 (-8.3,0.5)	-3.8 (-9.0,7)	-3.9 (-9.9,1)	-3.6 (-10.9,1.7)	-2.8 (-12.4,3.6)	-1.1 (-14.7,8.6)	2.6 (-19.4,20.9)	11.8 (-30.4,52.1)	52.4 (-75.8,188.3)
India	-2 (-5.3,1.4)	-1.7 (-5.2,2)	-1.2 (-4.9,2.7)	-0.6 (-4.7,3.6)	0.1 (-4.4,4.9)	1.1 (-4.2,6.6)	2.4 (-4.9)	4.3 (-4.1,12.8)	7.5 (-4.2,19.6)	16 (-5.6,37.9)
Mexico	-4.7 (-7.3,-2.1)	-4.2 (-7.1,-1.3)	-3.4 (-6.5,-0.3)	-2.2 (-5.8,1.2)	-0.8 (-4.7,3)	1.2 (-3.5,5.6)	3.9 (-1.7,9.3)	8 (0.7,15)	15.1 (4.8,25.1)	34.1 (15.5,52.7)
Mongolia	-3 (-11.4,5.3)	-2.5 (-10.7,7.6)	-1.7 (-9.2,9.8)	-0.6 (-7.5,12.2)	0.7 (-6.15,8)	2.7 (-5.3,20.2)	5.8 (-5.7,26.5)	10.3 (-7.3,36.2)	18 (-10.6,55)	38.4 (-22.4,108)
Morocco	4.3 (-0.5,9)	3.7 (-1.2,8.5)	2.9 (-2.1,7.8)	2 (-3.1,7.1)	0.9 (-4.6,6.4)	-0.4 (-6.7,5.6)	-2.2 (-9.6,5)	-4.6 (-14.4,4)	-8.7 (-21.7,4)	-18.8 (-41.5,3.3)
<b>Average</b>	-1.3 (-12.9,10.7)	-1.3 (-12.3,8.4)	-1 (-11.8,8.5)	-0.6 (-10.9,9.2)	0 (-10.3,10.5)	1 (-10.5,13.6)	2.3 (-11.9,20.8)	4.3 (-15.5,35.8)	7.7 (-23.6,63.8)	16.9 (-48.9,163.9)
<b>Full Pooling</b>										
<b>Average</b>	-3.9 (-6.8,-0.9)	0.2 (-2.4,2.9)	-0.9 (-3.7,1.9)	-1.8 (-5.1,1.4)	-1.3 (-4.8,2.2)	2.5 (-1.4,6.3)	3.6 (-0.8,7.9)	6.1 (0.2,11.9)	6.4 (-1.8,14.6)	13.9 (-6.1,33.9)

Notes: All units are USD PPP per two weeks. Estimates are shown with their 95% uncertainty intervals below them in brackets. In this case the full pooling and no pooling models are frequentist, estimated using the quantreg package in R, per Koencker and Basset 1978 with the nonparametric bootstrap providing the standard errors.

TABLE 2—PROFIT: COMPARISON OF NO POOLING, PARTIAL POOLING AND FULL POOLING RESULTS

Quantile:	5th	15th	25th	35th	45th	55th	65th	75th	85th	95th
<b>No Pooling</b>										
Bosnia	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	122.7 (57.5,164.1)	117.1 (30.1,189.1)	148.9 (-53.7,322.2)
Ethiopia	-5.8 (-11.7,0.7)	-1.8 (-3.2,-0.4)	-0.5 (-1.1,0.1)	0 (-0.7,0.7)	0.2 (-0.5,1)	0.5 (-0.4,1.5)	0.9 (-0.3,2.4)	1.7 (-0.3,3.9)	3.4 (-0.5,7.4)	10.1 (-2.6,21.5)
India	0 (-5.2,1.7)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	-6.7 (-17.6,6.1)	7.9 (-36.5,59.3)
Mexico	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	10.2 (-4.3,26.3)
Mongolia	-1.8 (-5.1,2)	-0.8 (-2.2,0.6)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	-0.2 (-1.3,0.6)
Morocco	-30.3 (-70.4,12.2)	1.5 (-6.4,8.7)	2.2 (-1.4,5)	0 (0,0)	0 (0,2.7)	3.8 (0.2,8.1)	7.8 (1.9,14.7)	14.5 (4.26,7)	28.6 (4.7,54.8)	78.5 (-20,183.5)
Philippines	0 (0,0)	-27.8 (-49.7,6.2)	-13.6 (-39.5,6.5)	-10.3 (-42.1,13.8)	-6.3 (-46.1,24.3)	0.2 (-51.1,40.3)	10.8 (-58.8,66.5)	30.6 (-72.7,113.3)	72.8 (-99.2,216.7)	222 (-206.1,614.1)
<b>Partial Pooling</b>										
Bosnia	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	89.5 (15.3,144.7)	79.5 (22.7,146.3)	129.2 (15.3,249.8)
Ethiopia	-3.7 (-9.3,1.5)	-1.2 (-2.6,0)	-0.4 (-0.9,0.1)	0 (-0.7,0.7)	0.2 (-0.5,0.9)	0.4 (-0.4,1.3)	0.8 (-0.3,2)	1.5 (-0.2,3.3)	2.9 (-0.2,6.3)	8.4 (-1.5,18.6)
India	0 (-3.9,1.3)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	-0.1 (-11.4,10.4)	25.3 (-19.64,5)
Mexico	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	12.2 (-1.8,25.7)
Mongolia	-1 (-3.6,1.7)	-0.6 (-1.8,0.6)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (-0.7,0.6)
Morocco	-35.8 (-69.4,-3.2)	-0.6 (-7.6,1)	1.2 (-1.6,4.1)	0 (0,0)	0 (-0.3,2.6)	3 (-0.7,6.9)	5.9 (0.2,12.2)	11.3 (1.7,22.6)	23.6 (2.5,48)	74.6 (-17.5,168.4)
Philippines	0 (0,0)	-4.9 (-33.6,18.7)	-0.7 (-19.1,14.6)	2.6 (-18.2,20.1)	7.1 (-17.8,28.4)	13.7 (-18.1,40.9)	24.1 (-18.5,61.5)	41.5 (-20.9,98.1)	77 (-31.4,180.5)	201 (-80.2,491.6)
<b>Average</b>	0 (-49.6,4.4)	0 (-9.2,0.2)	0 (-2.9,0.5)	0 (-1.2,3.1)	0 (-1.1,7)	0 (-2,12.5)	0 (-3,20)	2.8 (-4.5,31.5)	6.9 (-4.6,57.2)	20.8 (-5.5,165.1)
<b>Full Pooling</b>										
<b>Average</b>	2.4 (-2.6,7.4)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	5.6 (3.5,7.7)	21 (15.8,26.3)	106 (80.1,132.6)

Notes: All units are USD PPP per two weeks. Estimates are shown with their 95% uncertainty intervals below them in brackets. Formal joint posterior probabilities computed by integrating over regions of the joint parameter space show that the probability that the average effects are equal across quantiles is 0.007 at a 0.5 USD PPP "tolerance" level, and 0.011 at a 1 USD PPP tolerance. All intervals are produced by the tailored parametric likelihood model, as due to the discreteness of the data there is no reliable way to compute the standard errors of the classical frequentist quantile effect estimator (see Mosteller 1946 and Koencker and Basset 1978). Because these posteriors are fat tailed, the headline estimate provided is the posterior median value.

TABLE 3—POSTERIOR PROBABILITY THAT EFFECTS ARE EQUAL FOR GROUPS PB = 0 AND 1

Quantile	Tolerance = 0.5 USD PPP	1 USD PPP	3 USD PPP	5 USD PPP
0.050	0.219	0.275	0.426	0.507
0.150	0.608	0.634	0.697	0.734
0.250	0.688	0.718	0.793	0.830
0.350	0.588	0.633	0.761	0.826
0.450	0.425	0.482	0.654	0.756
0.550	0.276	0.341	0.534	0.660
0.650	0.165	0.220	0.401	0.536
0.750	0.082	0.121	0.268	0.399
0.850	0.035	0.060	0.157	0.251
0.950	0.011	0.020	0.062	0.101

*Note:*

Notes: PB stands for "Previous Business", and PB = 1 indicates a household has previous business experience before accessing microcredit, else PB = 0. Posterior probabilities are generated by integrating over areas of the full joint posterior distribution, which automatically adjusts for dependencies in the uncertainty on these parameters. Tolerances must be used in probability statements (or tests) on equality conditions to avoid measure-theoretic paradoxes caused by testing a single point in Real space, which has measure zero under a continuous prior (and indeed, under the Lebesgue measure). In practice, this amounts to testing whether the two values are in a small interval near one another with interval length equal to the tolerance.

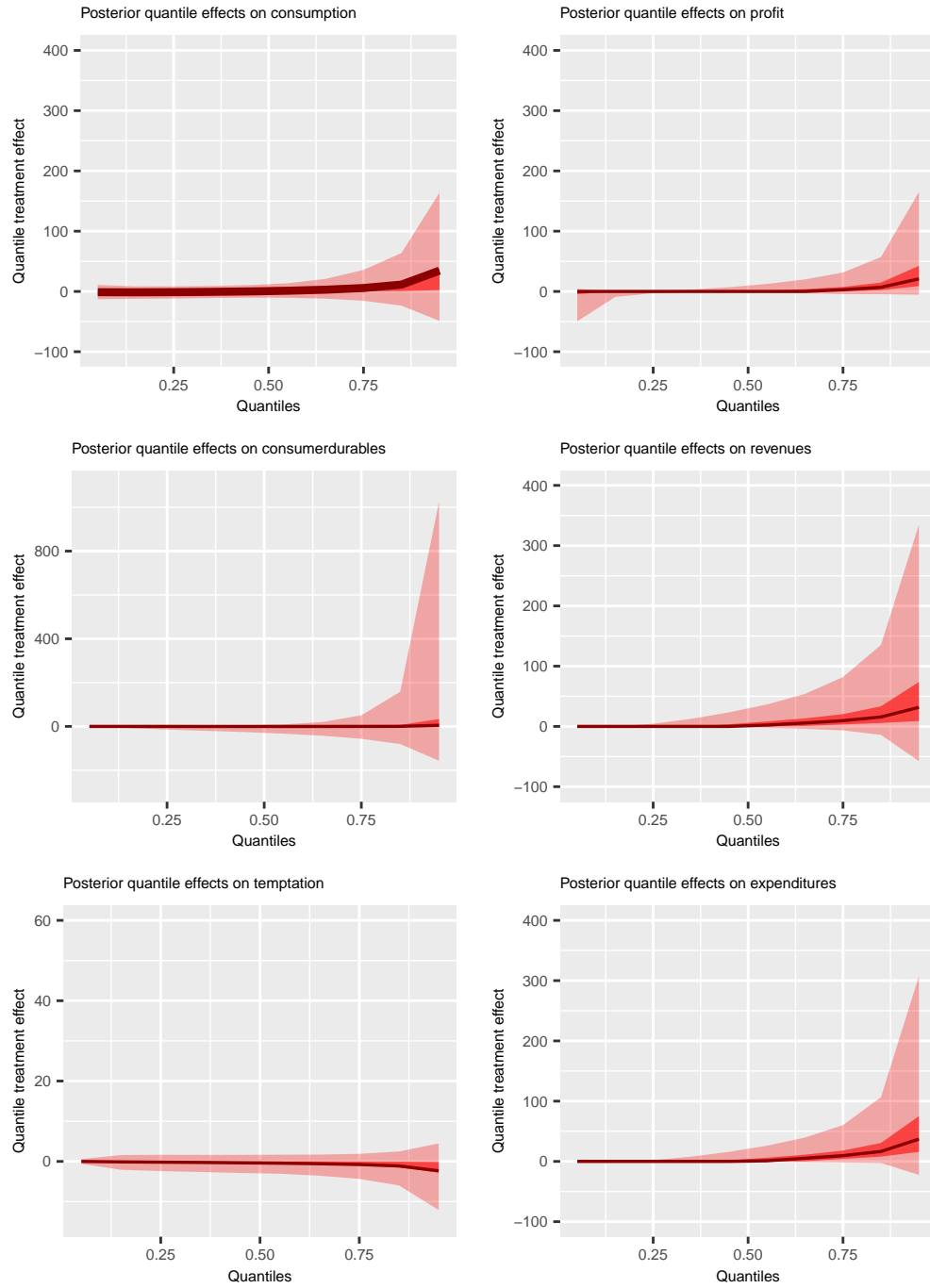


FIGURE 1. AVERAGE QUANTILE TREATMENT EFFECTS IN USD PPP PER 2 WEEK INTERVAL ACROSS ALL SETTINGS FOR ALL VARIABLES. THE DARK LINE IS THE POSTERIOR MEAN, THE OPAQUE COLOR BANDS ARE THE CENTRAL 50% POSTERIOR UNCERTAINTY INTERVAL, THE TRANSLUCENT COLOR BANDS ARE THE CENTRAL 95% POSTERIOR UNCERTAINTY INTERVAL.



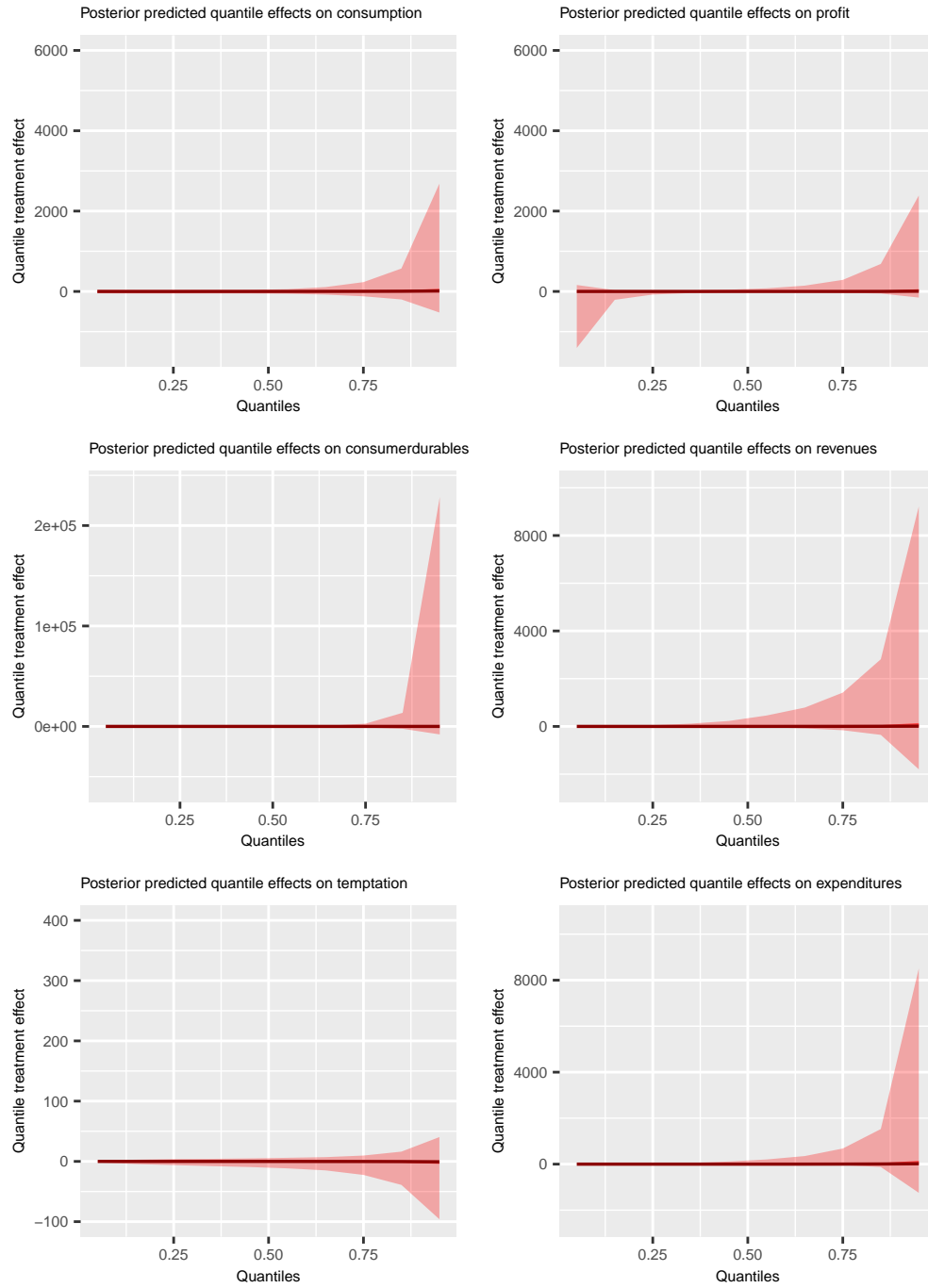


FIGURE 2. POSTERIOR PREDICTIVE QUANTILE TREATMENT EFFECTS IN USD PPP PER 2 WEEK INTERVAL FOR THE NEXT SETTING FOR ALL VARIABLES. THE DARK LINE IS THE POSTERIOR MEAN, THE OPAQUE COLOR BANDS ARE THE CENTRAL 50% POSTERIOR PREDICTIVE UNCERTAINTY INTERVAL, THE TRANSLUCENT COLOR BANDS ARE THE CENTRAL 95% POSTERIOR PREDICTIVE UNCERTAINTY INTERVAL.

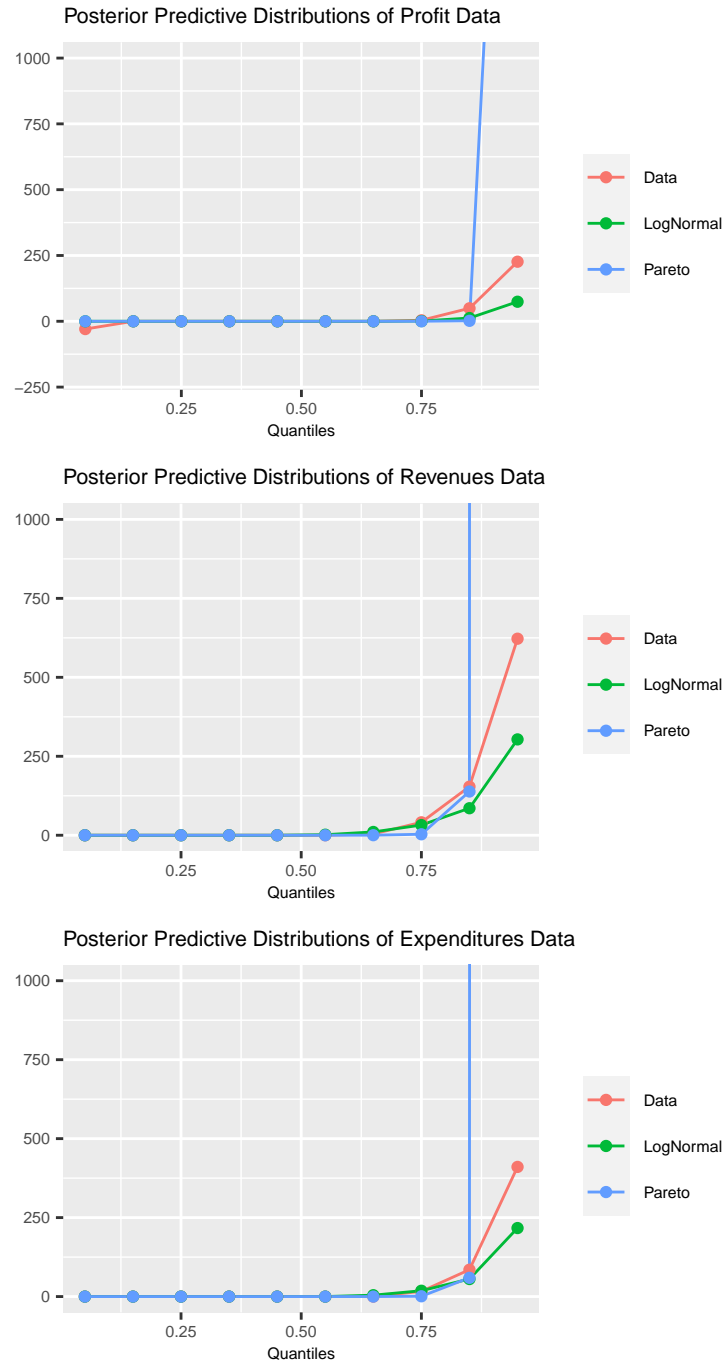


FIGURE 3. MODEL FIT ANALYSIS: POSTERIOR PREDICTIVE QUANTILES IN USD PPP PER 2 WEEK INTERVAL IN THE CONTROL GROUP FROM THE LOGNORMAL AND PARETO MODELS COMPARED TO THE REAL DATA. FURTHER DETAILS AND CAVEATS CAN BE FOUND IN APPENDIX C.

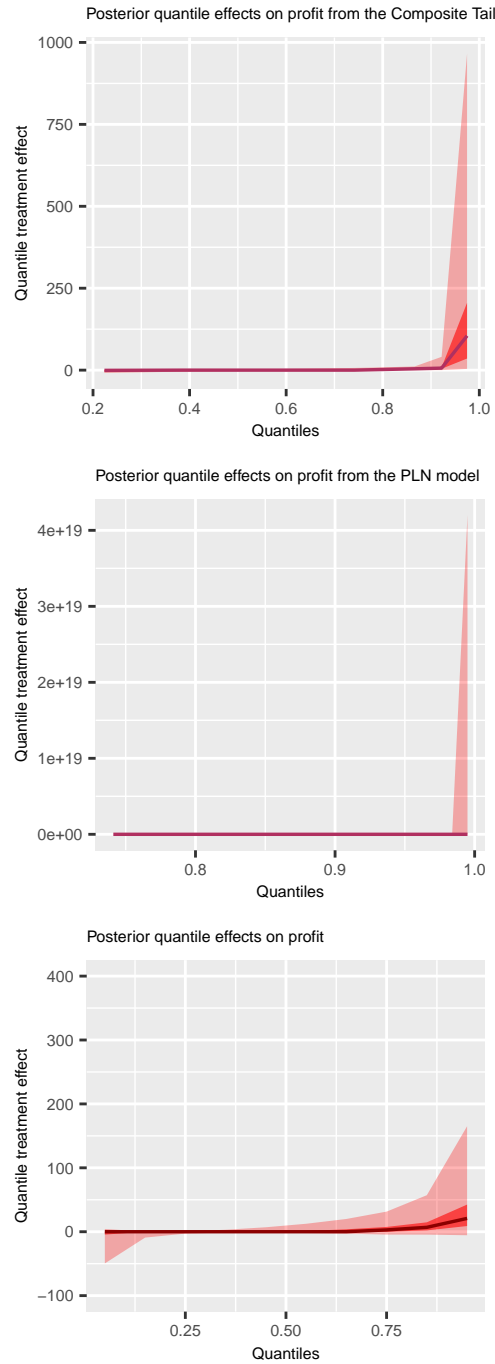
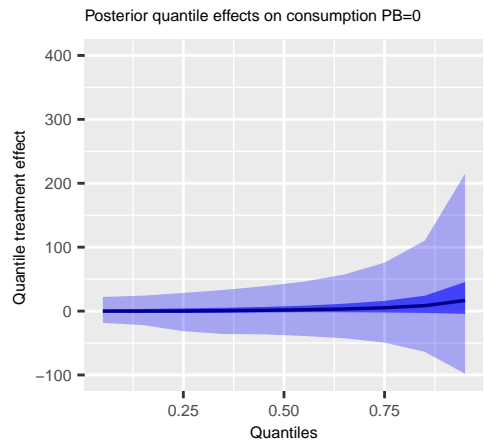
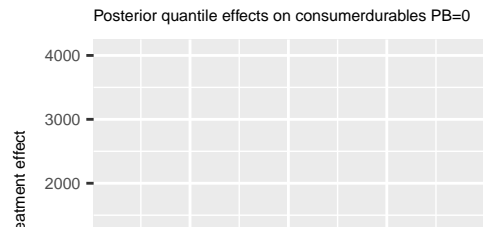
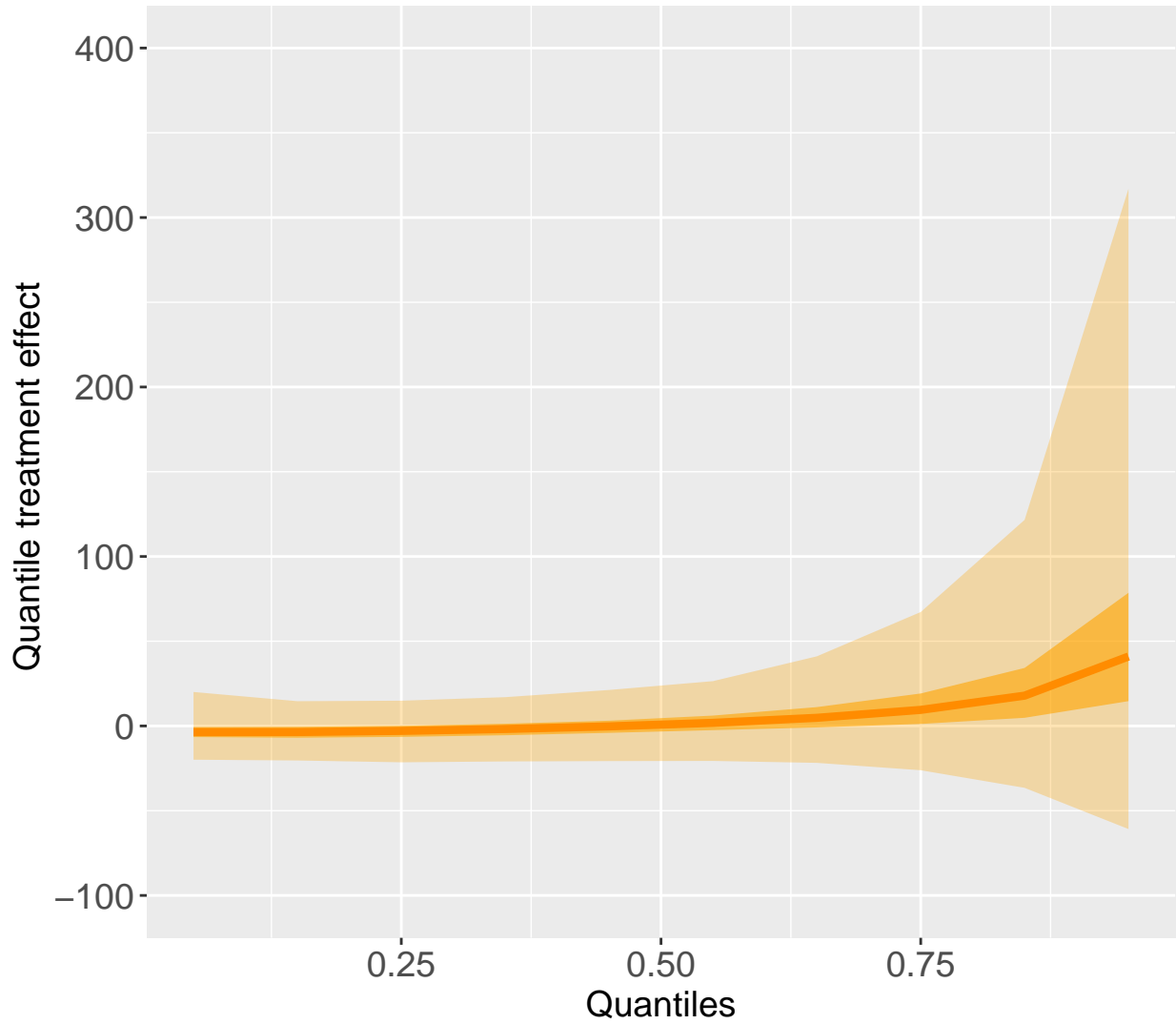
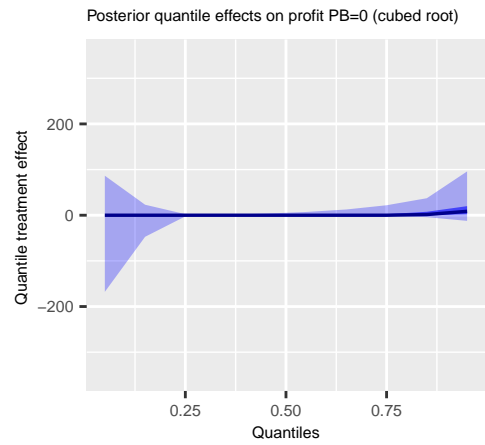


FIGURE 4. AVERAGE QUANTILE TREATMENT EFFECTS IN USD PPP PER 2 WEEK INTERVAL FOR THE NEXT SETTING FOR ALL VARIABLES, GRAPHED IN SQUARE ROOT TERMS DUE TO THE VERY LARGE SCALE OF THE PARETO TAIL. THE DARK LINE IS THE POSTERIOR MEDIAN, THE OPAQUE COLOR BANDS ARE THE CENTRAL 50% POSTERIOR PREDICTIVE UNCERTAINTY INTERVAL, THE TRANSLUCENT COLOR BANDS ARE THE CENTRAL 95% POSTERIOR PREDICTIVE UNCERTAINTY INTERVAL. THE PARETO-LOGNORMAL MODEL HAS CONVERGENCE ISSUES AS NOTED IN REED AND JORGENSEN 2004, HERE SOMEWHAT MITIGATED BY STRONG PRIORS, AND SHOULD BE INTERPRETED WITH CAUTION. THE COMPOSITE TAIL MODEL, GENERATED BY MANUALLY CUTTING UP THE SUPPORT OF EACH TAIL AND FITTING A LOGNORMAL BETWEEN ZERO AND THE 80TH PERCENTILE AND A PARETO BEYOND IT, IS COMPUTED IN A TWO-STEP PROCEDURE. FURTHER DETAILS AND CAVEATS CAN BE FOUND IN APPENDIX C.

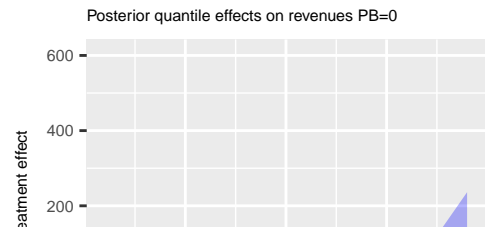
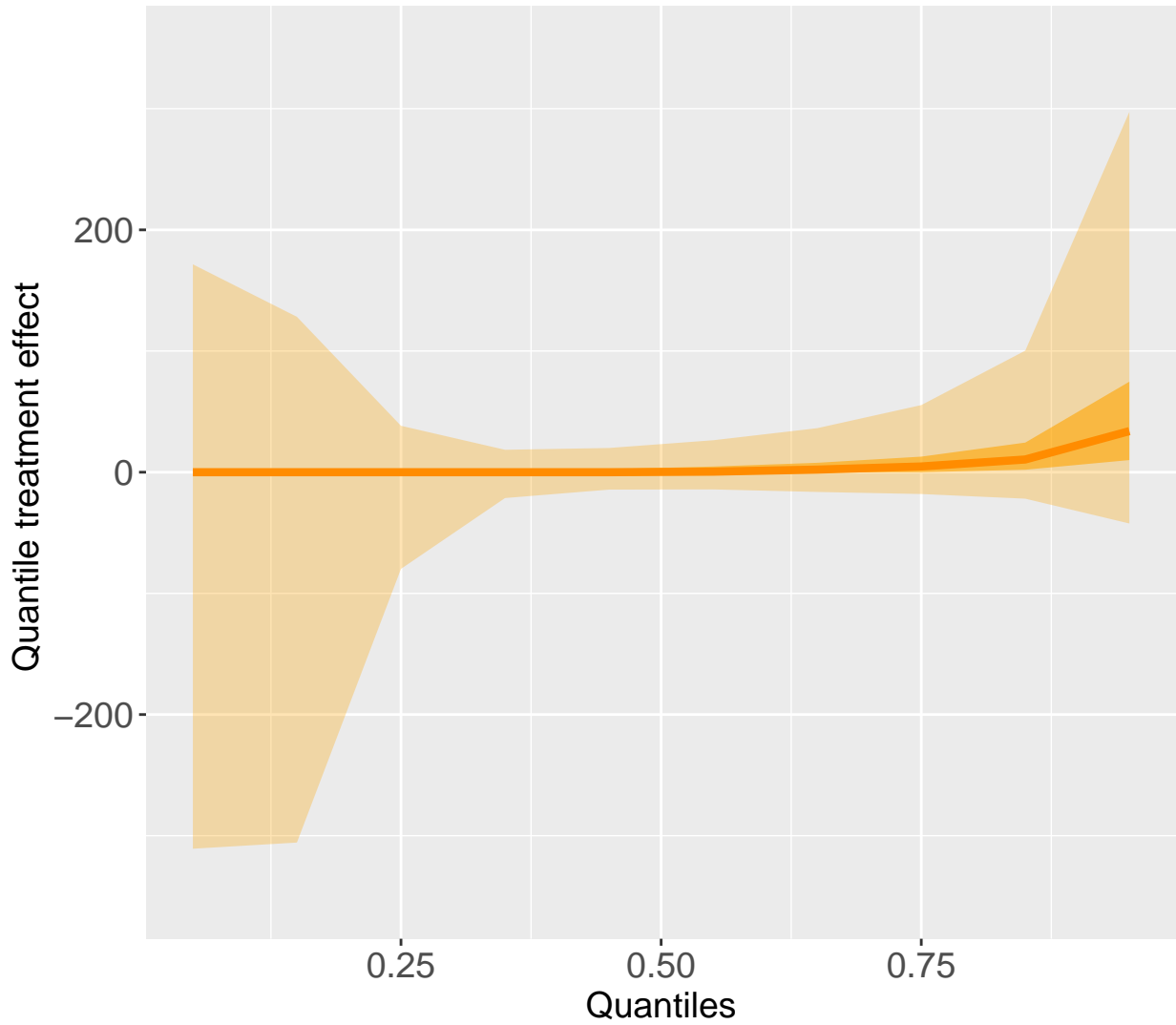


Posterior quantile effects on consumption PB=1





Posterior quantile effects on profit PB=1



Sample figure:

Figure here.

FIGURE 7. CAPTION FOR FIGURE BELOW.

*Note:* Figure notes without optional leadin.  
*Source:* Figure notes with optional leadin (Source, in this case).

Sample table:

TABLE 4—CAPTION FOR TABLE ABOVE.

	Heading 1	Heading 2
Row 1	1	2
Row 2	3	4

*Note:* Table notes environment without optional leadin.  
*Source:* Table notes environment with optional leadin (Source, in this case).

References here (manual or bibTeX). If you are using bibTeX, add your bib file name in place of BibFile in the bibliography command.