Appendix C: Tabular Results and Robustness Checks

Rachael Meager

February 4, 2021

1 Tabular results

Table 1: Lender and Study Attributes by Country

Country	Bosnia & Herzegovina	Ethiopia	India	Mexico	Mongolia	Morocco	The Philippines
Study Citation	Augsburg et al (2015)	Tarozzi et al (2015)	Banerjee et al (2015b)	Angelucci et al (2015)	Attanasio et al (2015)	Crepon et al (2015)	Karlan and Zinman (2011)
Treatment	Lend to marginally rejected borrowers	Open branches	Open branches	Open branches, promote loans	Open branches, target likely borrowers	Open branches	Lend to marginal applicants
Randomization Level	Individual	Community	Community	Community	Community	Community	Individual
Urban or Rural?	Both	Rural	Urban	Both	Rural	Rural	Urban
Target Women?	No	No	Yes	Yes	Yes	No	No
MFI already operates locally?	Yes	No	No	No	No	No	Yes
Microloan Liability Type	Individual	Group	Group	Group	Both	Group	Individual
Collateralized?	Yes	Yes	No	No	Yes	No	No
Any other MFIs competing?	Yes	No	Yes	Yes	Yes	No	Yes
Household Panel?	Yes	No	No	Partial	Yes	Yes	No
Interest Rate (Intended on Average)	22% APR	12% APR	24% APR	100% APR	24% APR	13.5% APR	63% APR
Sampling Frame	Marginal Applicants	Random Sample	Households with at least 1 woman age 18-55 of stable residence	Women ages 18-60 who own businesses or wish to start them	Women who registered interest in loans and met eligibility criteria	Random Sample plus Likely Borrowers	Marginal Applicants
Study Duration	14 months	36 months	40 months	16 months	19 months	24 months	36 months

Note: The construction of the interest rates here is different to the construction of Banerjee et al (2015a); they have taken the maximal interest rate, whereas I have taken the average of the intended range specified by the MFI. In practice the differences in these constructions are numerically small. This table was also printed in Meager (2018) which used the same studies.

Excess Kurtosis in LogNormal distributions is the extent to which tail indices are greater, and thus the extent to which the tails are heavier, than those of the Gaussian. For a LogNormal parameterised as

$$LogNormal(y|\mu, \omega) = \frac{1}{\sqrt{2\pi}\omega y} \exp\left(\frac{-(\log(y) - \mu)^2}{2\omega^2}\right)$$

the excess kurtosis is

$$\exp(4\omega^2) + 2\exp(3\omega^2) + 3\exp(2\omega^2) - 6.$$

I compute the kurtosis for the general control group based on the posterior mean values of μ and σ for this group in the tables below. The μ parameters are the same in the model and in the formula above, but the ω parameter requires some explanation. ω in the formula is the Lognormal scale. The σ parameters in the models are the log versions of this parameter. For example, the posterior mean scale parameter for the LogNormal in the "generalized" control group's positive tail is actually $\omega = \exp(\sigma_2^c)$, and this must be squared further to enter the formula above.The examples in the text are obtained using σ_2^c from profit and from consumption respectively, to give excess kurtosis values of 810.5 and 13.9 respectively (there will be a small rounding error if plugging in the values from the table here).

Table 2: All General-Level Posterior Marginals for the LogNormal Profit Model

	mean	MCMC error	sd	2.5%	25%	50%	75%	97.5%	# effective draws	$\hat{R}$
μ_1	3.200	0.008	0.732	1.722	2.784	3.200	3.615	4.698	9,099	1.000
μ_2	3.843	0.007	0.818	2.225	3.356	3.845	4.324	5.496	15,000	1.000
τ_1	0.094	0.001	0.094	-0.099	0.045	0.095	0.143	0.273	6,719.600	1.001
τ_2	0.077	0.0005	0.042	-0.007	0.054	0.078	0.102	0.157	7,566.232	1.000
σ_{μ_1}	1.659	0.008	0.654	0.867	1.227	1.514	1.923	3.302	7,284.792	1.000
σ_{μ_2}	2.033	0.006	0.677	1.153	1.574	1.889	2.332	3.711	15,000	1.000
σ_{τ_1}	0.117	0.004	0.128	0.005	0.035	0.079	0.154	0.459	1,090.338	1.003
σ_{τ_2}	0.055	0.001	0.052	0.002	0.020	0.043	0.075	0.183	1,323.050	1.004
σ_1^c	0.452	0.002	0.145	0.180	0.374	0.447	0.525	0.761	7,205.404	1.000
σ_2^c	0.225	0.001	0.101	0.022	0.167	0.225	0.284	0.428	10,278.910	1.000
σ_1^t	0.022	0.001	0.094	-0.162	-0.024	0.022	0.067	0.206	6,128.028	1.001
σ_2^t	0.017	0.0003	0.029	-0.043	0.001	0.017	0.032	0.072	9,321.264	1.000
$\sigma_{\sigma_1^c}$	0.302	0.002	0.164	0.122	0.196	0.262	0.357	0.724	5,126.273	1.001
$\sigma_{\sigma_2^c}$	0.242	0.001	0.100	0.125	0.176	0.220	0.280	0.499	9,328.806	1.000
$\sigma_{\sigma_1^t}$	0.163	0.002	0.116	0.034	0.089	0.134	0.201	0.467	2,860.338	1.001
$\sigma_{\sigma_2^t}$	0.046	0.001	0.037	0.002	0.020	0.038	0.062	0.140	2,034.778	1.002
β_{11}	-1.965	0.016	1.273	-4.525	-2.715	-1.958	-1.193	0.527	6,334.358	1.000
β_{12}	0.025	0.001	0.114	-0.187	-0.035	0.019	0.080	0.265	6,957.068	1.001
β_{21}	0.390	0.010	0.906	-1.379	-0.168	0.367	0.918	2.255	7,964.995	1.000
β_{22}	-0.067	0.001	0.104	-0.279	-0.124	-0.066	-0.012	0.143	8,309.348	1.001
$\sigma_{\beta_{11}}$	2.767	0.017	1.277	0.770	1.959	2.560	3.346	5.904	5,636.316	1.000
$\sigma_{\beta_{12}}$	0.128	0.002	0.125	0.005	0.047	0.096	0.168	0.446	5,901.720	1.001
$\sigma_{\beta_{21}}$	1.603	0.014	0.902	0.130	0.990	1.532	2.093	3.672	3,987.814	1.002
$\sigma_{\beta_{22}}$	0.146	0.002	0.114	0.007	0.065	0.124	0.197	0.432	5,234.755	1.001
$\sigma_{\beta_{31}}$	1.450	0.014	0.889	0.091	0.815	1.381	1.942	3.493	3,896.658	1.001
$\sigma_{\beta_{32}}$	0.117	0.002	0.109	0.004	0.041	0.089	0.161	0.390	5,085.964	1.002

Note: The β_3 parameters are normalized to be zero at the general level as required for multinomial logit models. The site-specific effects still have variation around this zero anchor as reported.

Table 3: All General-Level Posterior Marginals for the LogNormal Revenues Model

	mean	MCMC error	sd	2.5%	25%	50%	75%	97.5%	# effective draws	$\hat{R}$
μ_1	4.472	0.007	0.873	2.733	3.959	4.479	4.992	6.193	$15,000$	1.000
τ_1	0.083	0.001	0.068	-0.058	0.045	0.086	0.123	0.211	$10,482.840$	1.000
σ_{μ_1}	2.181	0.007	0.718	1.258	1.693	2.030	2.496	3.982	$10,285.460$	1.000
$\sigma_{\tau_1}]$	0.140	0.001	0.080	0.039	0.089	0.124	0.171	0.329	$5,189.630$	1.001
σ_1^c	0.213	0.001	0.136	-0.063	0.134	0.214	0.292	0.485	$11,190.950$	1.000
σ_1^t	-0.010	0.0003	0.031	-0.071	-0.028	-0.011	0.008	0.052	$9,554.774$	1.000
$\sigma_{\sigma_1^c}$	0.331	0.001	0.135	0.171	0.241	0.301	0.383	0.668	$8,452.406$	1.001
$\sigma_{\sigma_1^t}$	0.062	0.0004	0.033	0.020	0.040	0.055	0.075	0.146	$6,447.524$	1.000
β_{11}	0.011	0.008	0.734	-1.464	-0.424	-0.004	0.443	1.521	$8,107.184$	1.001
β_{12}	-0.063	0.001	0.081	-0.235	-0.101	-0.058	-0.020	0.091	$6,772.048$	1.001
$\sigma_{\beta_{11}}$	1.209	0.010	0.760	0.064	0.637	1.164	1.645	2.912	$5,305.339$	1.001
$\sigma_{\beta_{12}}$	0.095	0.001	0.091	0.003	0.032	0.071	0.129	0.327	$5,418.020$	1.001
$\sigma_{\beta_{21}}$	1.192	0.010	0.762	0.062	0.615	1.147	1.631	2.894	$5,341.343$	1.001
$\sigma_{\beta_{22}}$	0.095	0.001	0.091	0.003	0.033	0.071	0.130	0.328	$5,944.329$	1.000

Note: The β_3 parameters are normalized to be zero at the general level as required for multinomial logit models. The site-specific effects still have variation around this zero anchor as reported. Note also that σ_1^t can be negative as this is the effect specified on the exponential level.

Table 4: All General-Level Posterior Marginals for the LogNormal Expenditures Model

	mean	MCMC error	sd	2.5%	25%	50%	75%	97.5%	# effective draws	$\hat{R}$
μ_1	4.042	0.006	0.733	2.563	3.593	4.047	4.483	5.528	$15,000$	1.000
τ_1	0.103	0.001	0.048	0.005	0.076	0.104	0.132	0.198	$8,840.624$	1.000
σ_{μ_1}	1.867	0.005	0.624	1.061	1.449	1.735	2.135	3.449	$15,000$	1.001
σ_{τ_1}	0.078	0.001	0.060	0.004	0.035	0.067	0.106	0.226	$1,919.668$	1.002
σ_1^c	0.303	0.002	0.171	-0.037	0.204	0.304	0.401	0.649	$8,974.738$	1.001
σ_1^t	-0.008	0.001	0.045	-0.092	-0.033	-0.009	0.016	0.082	$5,069.866$	1.000
$\sigma_{\sigma_1^c}$	0.421	0.002	0.171	0.218	0.309	0.382	0.489	0.845	$8,374.404$	1.001
$\sigma_{\sigma_1^t}$	0.094	0.001	0.051	0.035	0.062	0.082	0.111	0.217	$3,164.881$	1.001
β_{11}	0.234	0.009	0.694	-1.177	-0.180	0.233	0.653	1.645	$6,027.909$	1.000
β_{12}	-0.116	0.001	0.117	-0.349	-0.177	-0.114	-0.053	0.112	$7,262.210$	1.000
$\sigma_{\beta_{11}}$	1.148	0.011	0.712	0.062	0.613	1.102	1.565	2.729	$4,414.652$	1.001
$\sigma_{\beta_{12}}$	0.157	0.002	0.125	0.007	0.071	0.132	0.209	0.465	$5,601.528$	1.000
$\sigma_{\beta_{21}}$	1.119	0.011	0.707	0.056	0.580	1.075	1.535	2.714	$4,076.193$	1.001
$\sigma_{\beta_{22}}$	0.159	0.002	0.124	0.007	0.074	0.136	0.212	0.463	$5,427.373$	1.001

Note: The β_3 parameters are normalized to be zero at the general level as required for multinomial logit models. The site-specific effects still have variation around this zero anchor as reported.Note also that σ_1^t can be negative as this is the effect specified on the exponential level.

For visual ease, the figures below graph the treatment effects and posterior predicted effects for each of the dimensions of change permitted in the model.

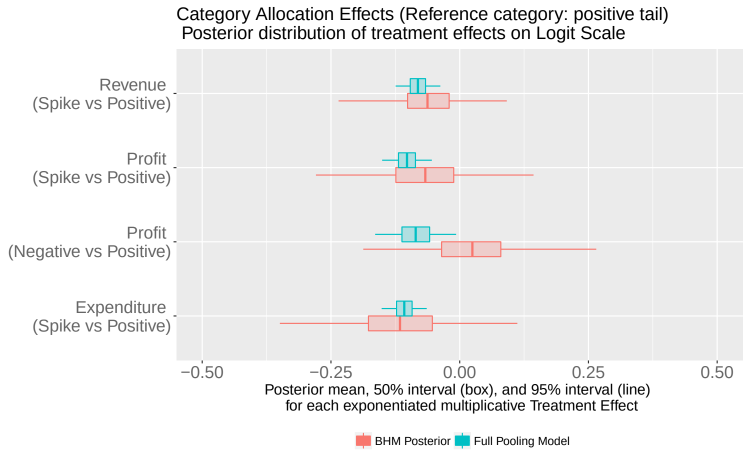

Figure 1: Posterior distributions for the logit treatment effects (π_j) on category assignment. These treatment effects are specified as an exponentiated multiplicative factor on the control group proportion of households in the category: if $\tilde{\pi_j} = 0$ the effect is zero, if $\tilde{\pi_j} < 0$ the treatment increases the proportion of households in the positive tail relative to other categories.

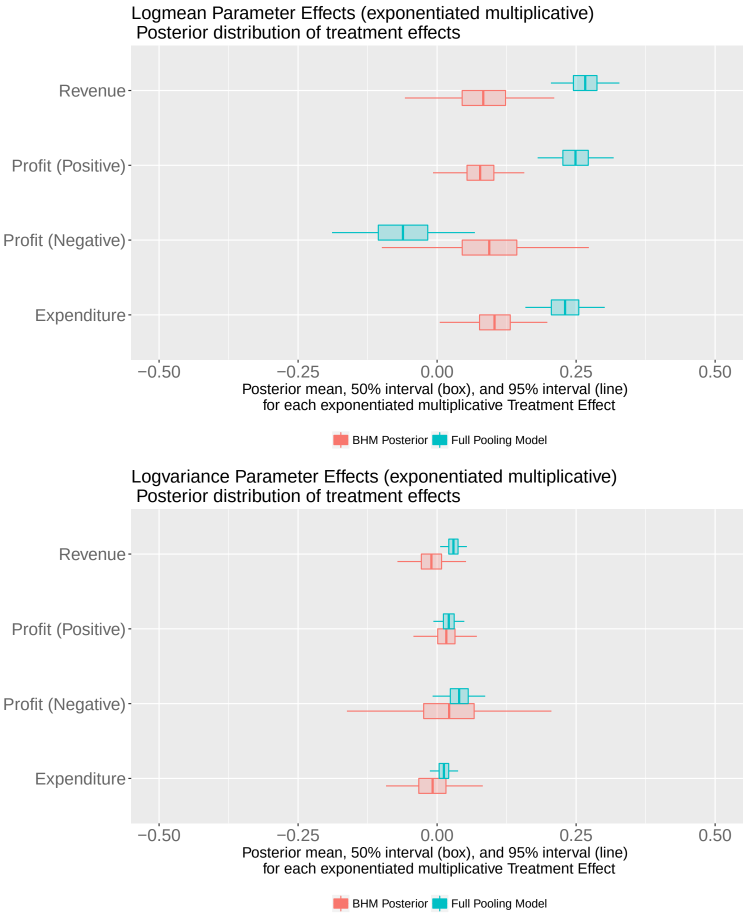

Figure 2: Posterior distributions for the location treatment effects (τ_j) and the scale treatment effects (σ_j^t).

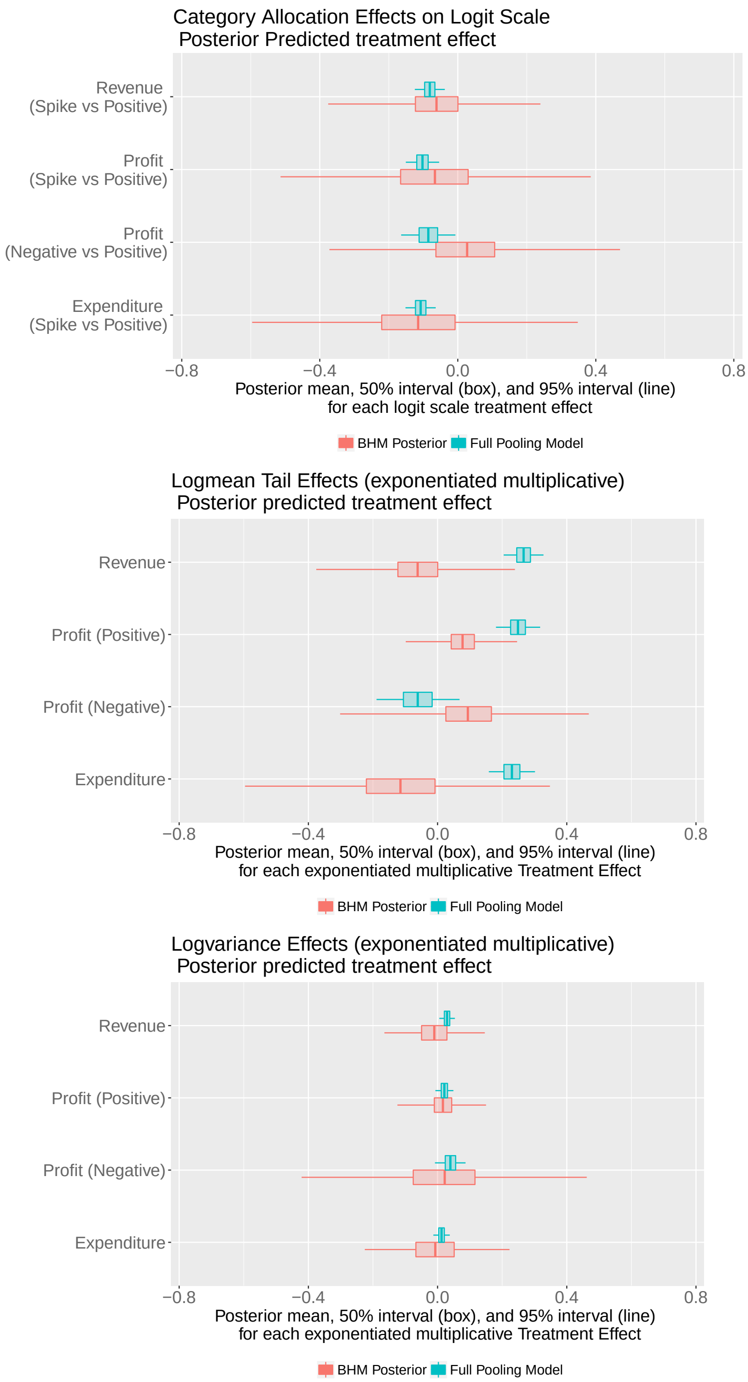

Figure 3: Posterior predicted distributions for the logit treatment effects on category assignment and tail shape effects.

2 Robustness Checks

2.1 Pareto Tail Models

If using the Pareto distribution for the continuous component, the tails are governed by a location parameter which controls the lower bound of the support and a scale parameter which controls the thickness of the tail. The location parameter ι_{jk} is exactly known because I have already defined the domain of each of the components by manually splitting the data. However the shape parameter is unknown and may be affected by treatment, which I model using a multiplicative exponential regression specification to impose a non-negativity constraint on the parameter. The shape parameter in mixture component j for household n in site k is therefore $\exp(\rho_{jk} + \kappa_{jk} T_{nk})$.

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:

$$\begin{aligned} y_{nk}|T_{nk} \sim\ & \Lambda_{1k}(T_{nk}) Pareto(-y_{nk}|\iota_{1k}, \exp(\rho_{1k} + \kappa_{1k} T_{nk}) \\ & + \Lambda_{2k}(T_n)\delta_{(0)} \\ & + \Lambda_{3k}(T_n) Pareto(y_{nk}|\iota_{3k}, \exp(\rho_{3k} + \kappa_{3k} T_{nk})\ \forall\ k \\ \text{where}\ \ \Lambda_{jk}(T_{nk}) =\ & \frac{\exp(\alpha_{jk} + \pi_{jk} T_{nk})}{\sum_{j=1,2,3} \exp(\alpha_{jk} + \pi_{jk} T_{nk})} \end{aligned} \tag{2.1}$$

The quantiles are recovered thus using the Castellaci method:

$$\begin{aligned} Q(u) = & -\text{Pareto}^{-1}\left(1 - \frac{u}{\Lambda_1(T_n)} \mid \iota_{1k}, \rho_{1k}(\exp(\kappa_{1k} T_n))\right) * \mathbb{1}\{u < \Lambda_1(T_n)\} \\ & + 0 * \mathbb{1}\{\Lambda_1(T_n) < u < (\Lambda_1(T_n) + \Lambda_2(T_n)\} \\ & + \text{Pareto}^{-1}\left(\frac{u - (1 - \Lambda_3(T_n))}{\Lambda_3(T_n)} \mid \iota_{3k}, \rho_{3k}(\exp(\kappa_{3k} T_n)\right) * \mathbb{1}\{u > (1 - \Lambda_3(T_n))\} \end{aligned} \tag{2.2}$$

The fit of this model to the microcredit data is not good. The table below shows the posterior predictive fit of this model and the LogNormal model.

Table 5: Posterior Predictive Comparison of LogNormal and Pareto Models

Control Group Quantiles	5%	15%	25%	35%	45%	55%	65%	75%	85%	95%
Revenues Data	0	0	0	0	0	0	4	41	154	622
Lognormal Prediction	0	0	0	0	0	12	37	77	154	408
Pareto Prediction	0	0	0	0	0	0	0	5	337	2, 793, 933
Expenditures Data	0	0	0	0	0	0	0	17	85	411
Lognormal Prediction	0	0	0	0	0	0	15	40	93	283
Pareto Prediction	0	0	0	0	0	0	0	1	94	1, 172, 324
Profit Data	-29	0	0	0	0	0	0	4	49	226
Lognormal Prediction	-2	0	0	0	0	0	4	21	56	173
Pareto Prediction	0	0	0	0	0	0	0	0	21	70, 170

Notes: The posterior predictive distributions are generated by drawing samples of data from the likelihood averaged over the posterior probability of the unknown parameters. Because this data is itself fat tailed, I have compared the actual sample quantiles from the fully pooled control group against the posterior predicted median value of each quantile from each model. [Back to main]

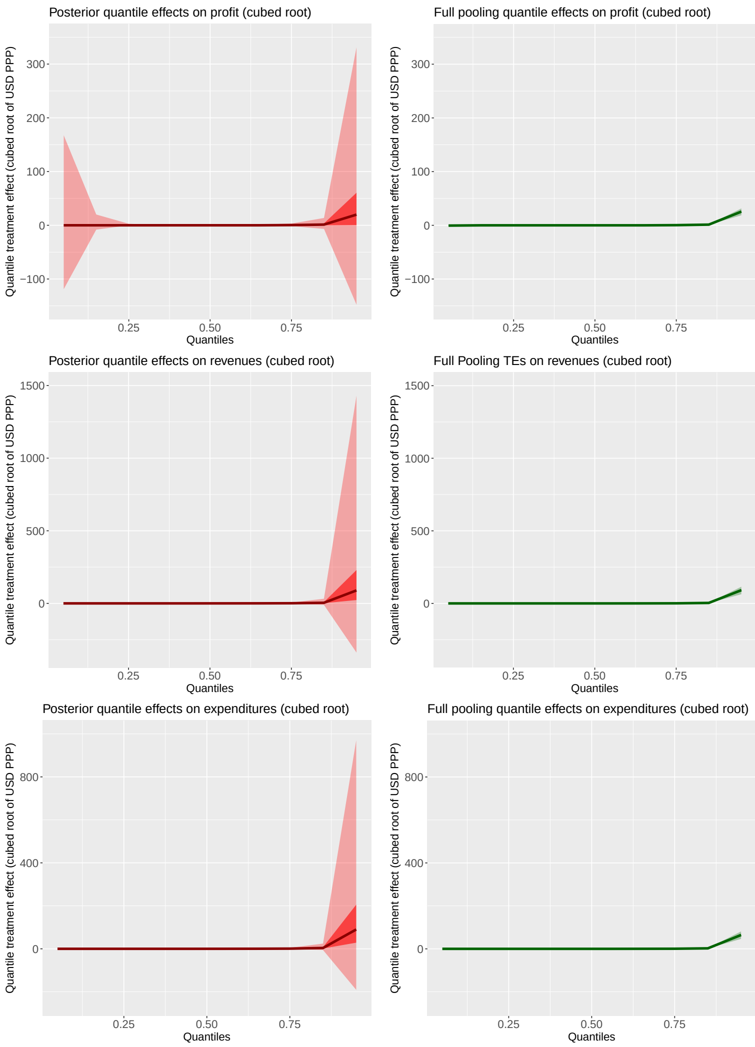

Figure 4: General Quantile Treatment Effect Curves (β_1) for business variables. The dark line is the median, the opaque bars are the central 50% interval, the translucent bands are the central 95% interval. Display is in cubed root of USD PPP due to the scale differences in the uncertainty at the right tail versus the rest of the distribution.

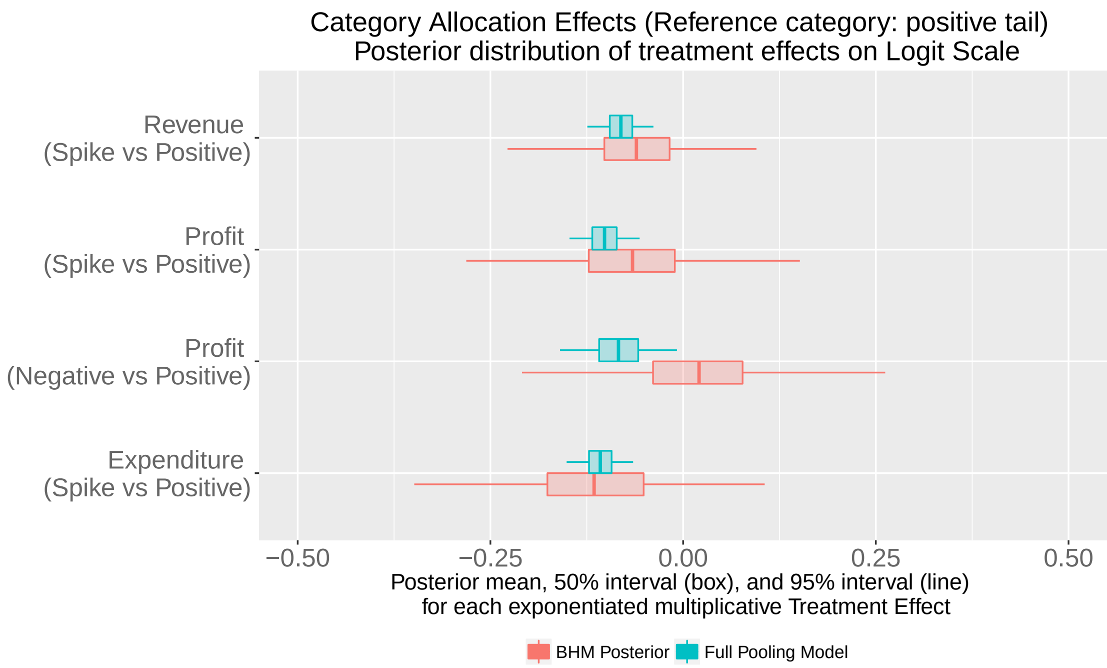

Figure 5: Posterior distributions for the logit treatment effects (π_j) on category assignment. These treatment effects are specified as an exponentiated multiplicative factor on the control group proportion of households in the category: if $\tilde{\pi}_j = 0$ the effect is zero, if $\tilde{\pi}_j < 0$ the treatment increases the proportion of households in the positive tail relative to other categories.

Table 6: Pooling Factors for Categorical Logit Parameters (Reference Category: Positive)

Outcome	Treatment Effects			Control Group Means		
	$\omega(\kappa_j)$	$\breve{\omega}(\kappa_j)$	$\lambda(\kappa_j)$	$\omega(\rho_j)$	$\breve{\omega}(\rho_j)$	$\lambda(\rho_j)$
Profit (Negative vs Positive)	0.378	0.721	0.907	0.144	0.421	0.240
Profit (Zero vs Positive)	0.137	0.476	0.688	0.013	0.379	0.487
Expenditures (Zero vs Positive)	0.084	0.612	0.783	0.010	0.498	0.570
Revenues (Zero vs Positive)	0.131	0.694	0.881	0.010	0.509	0.562

Notes: All pooling factors have support on [0,1], with 0 indicating no pooling and 1 indicating full pooling. The $\omega(\cdot)$ refers to the conventional pooling metric that scores signal strength at the general level against average signal strength at the local level. The $\breve{\omega}(\cdot)$ refers to the proximity-based "brute force" pooling metric that measures the geometric proximity of the partial pooling estimate to the no-pooling and full-pooling estimates. The $\lambda(\cdot)$ refers to the Gelman and Pardoe (2006) pooling metric that scores the posterior variation at the general level against the average posterior variation at the local level.

Table 7: Pooling Factors for Tail Shape Parameters

Outcome	Treatment Effects			Control Group Means		
	$\omega(\pi_j)$	$\breve{\omega}(\pi_j)$	$\lambda(\pi_j)$	$\omega(\alpha_j)$	$\breve{\omega}(\alpha_j)$	$\lambda(\alpha_j)$
Profit (Negative Tail)	0.389	0.855	0.991	0.284	0.346	0.494
Profit (Positive Tail)	0.219	0.785	0.988	0.036	0.074	0.089
Expenditures	0.175	0.756	0.987	0.019	0.061	0.050
Revenues	0.169	0.692	0.977	0.014	0.036	0.029

Notes: All pooling factors have support on [0,1], with 0 indicating no pooling and 1 indicating full pooling. The $\omega(\cdot)$ refers to the conventional pooling metric that scores signal strength at the general level against average signal strength at the local level. The $\breve{\omega}(\cdot)$ refers to the proximity-based "brute force" pooling metric that measures the geometric proximity of the partial pooling estimate to the no-pooling and full-pooling estimates. The $\lambda(\cdot)$ refers to the Gelman and Pardoe (2006) pooling metric that scores the posterior variation at the general level against the average posterior variation at the local level.

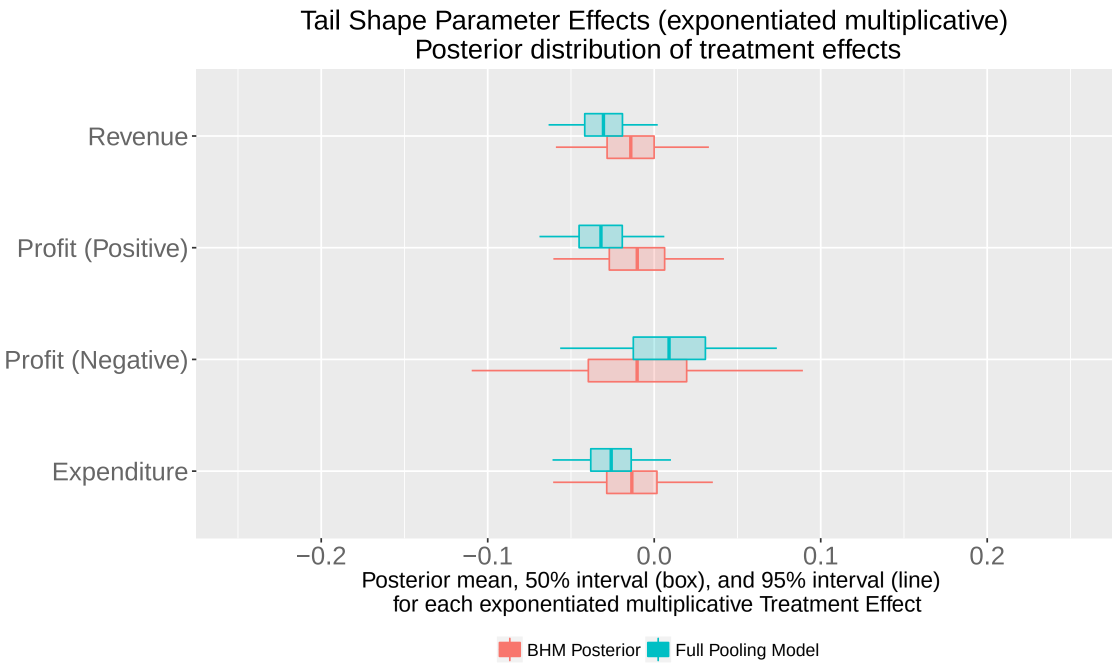

Figure 6: Posterior distributions for the Pareto shape treatment effects (κ_j) in each site. These treatment effects are specified as an exponentiated multiplicative factor on the control group scale parameter: if $\tilde{\kappa_j} = 0$ the effect is zero, if $\tilde{\kappa_j} = 0.7$ the effect is a 100% increase in the scale parameter.

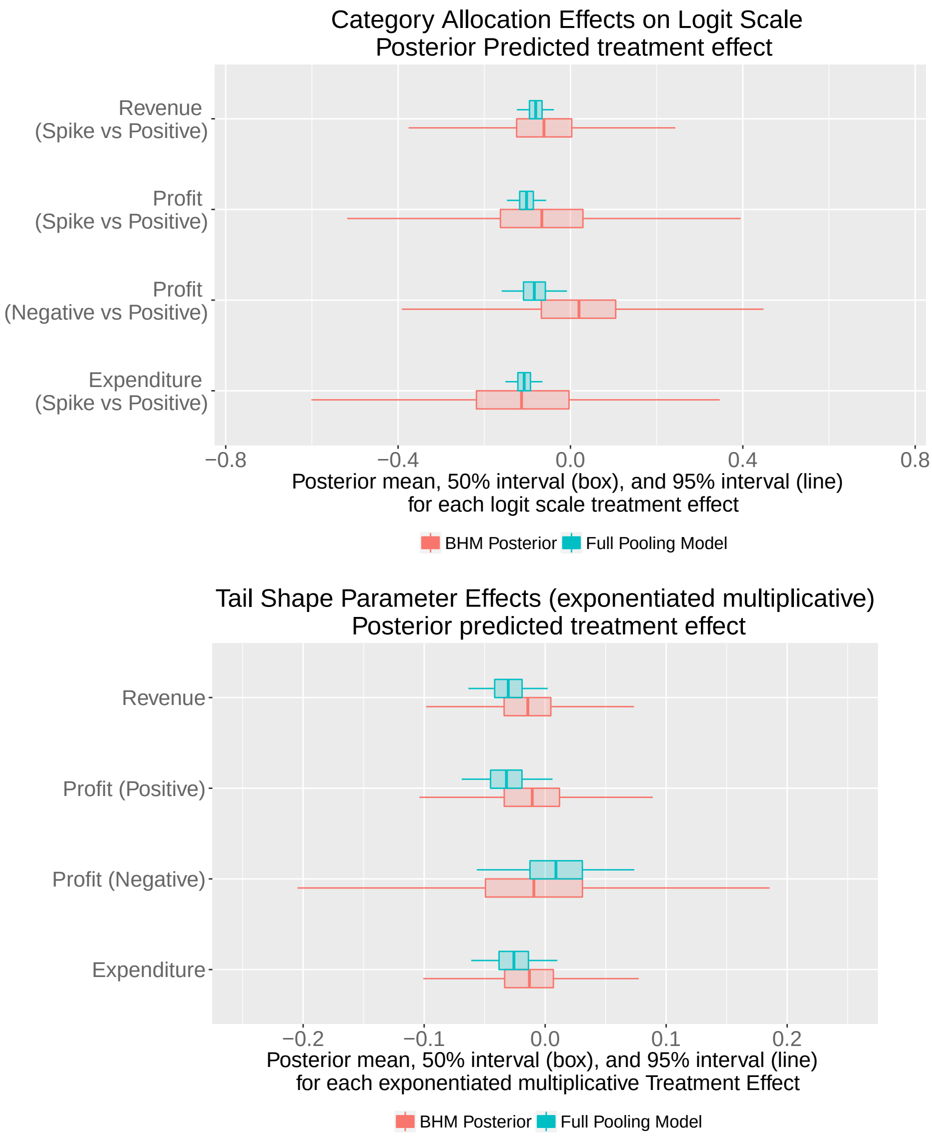

Figure 7: Posterior predicted distributions for the logit treatment effects on category assignment and tail shape effects. In each case this is the predicted treatment effect in a future exchangeable study site, with uncertainty intervals that account for the estimated generalizability (or lack of it).

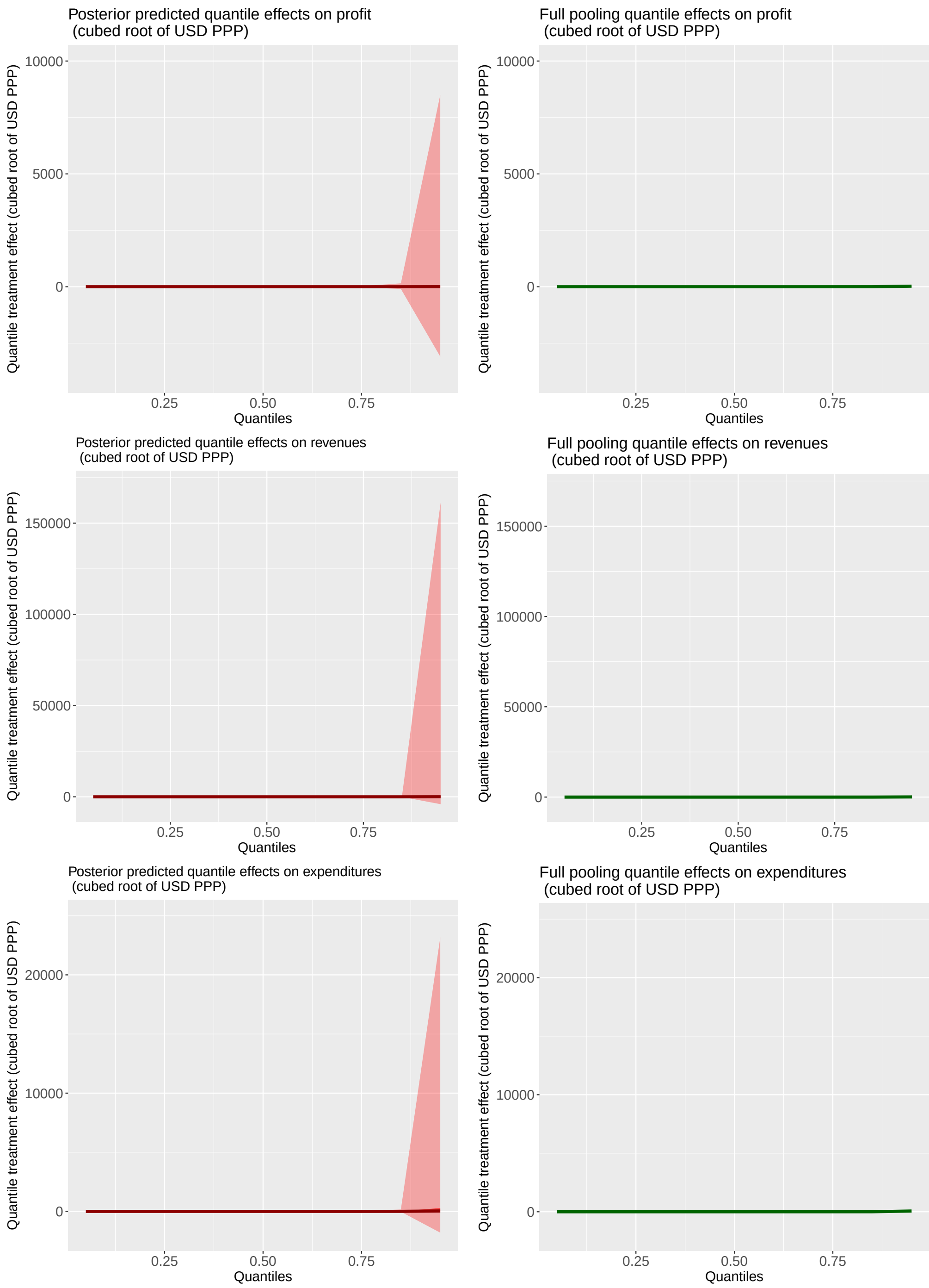

Figure 8: Posterior predicted quantile treatment effect Curves for Business Variables. The dark line is the median, the opaque bars are the central 50% interval, the translucent bands are the central 95% interval. Display is in cubed root of USD PPP due to the scale differences in the uncertainty at the right tail versus the rest of the distribution.

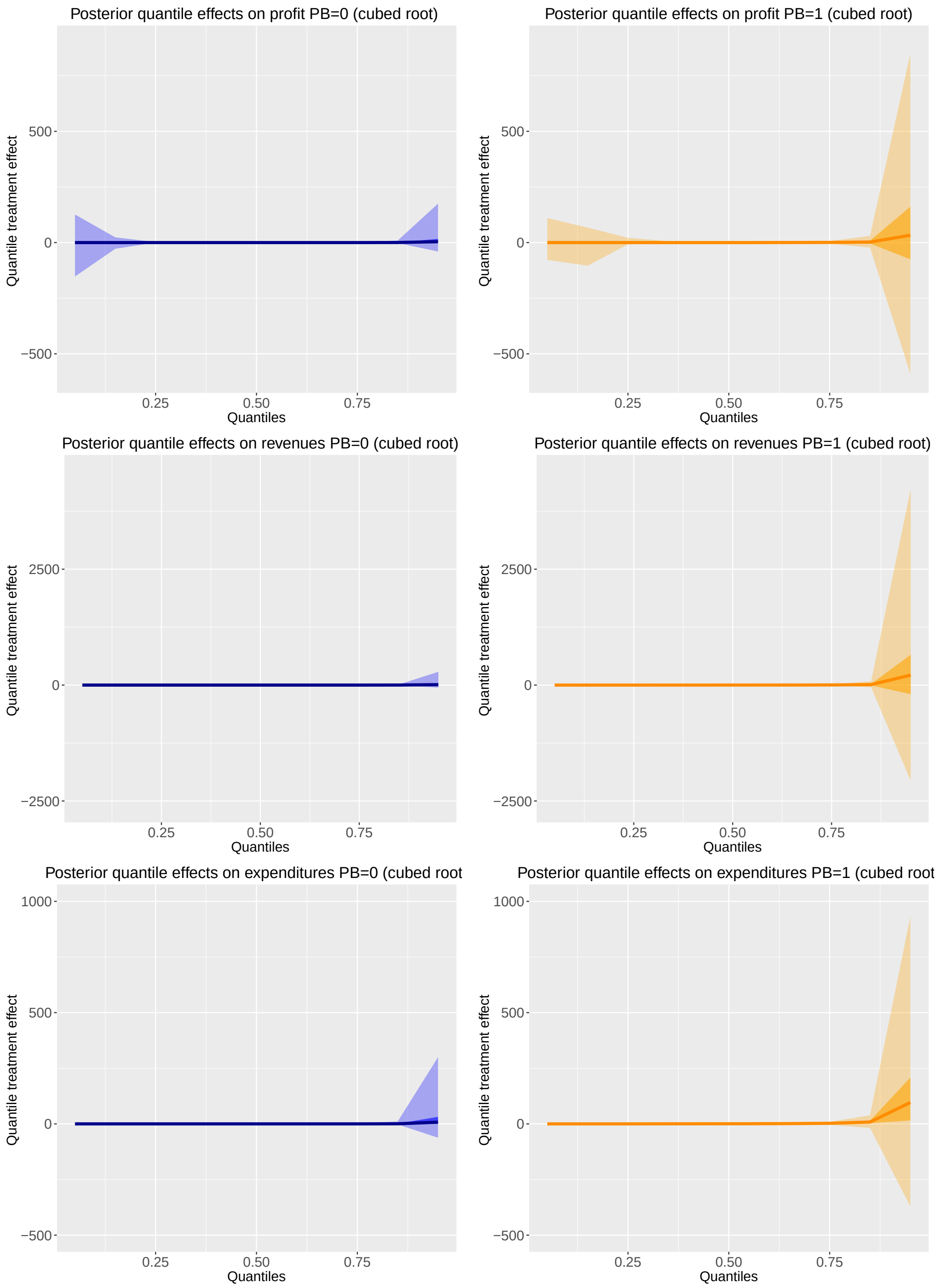

Figure 9: General Quantile Treatment Effect Curves (β_1) for business variables split by prior business ownership. The dark line is the median, the opaque bars are the central 50% interval, the translucent bands are the central 95% interval. Display is in cubed root of USD PPP due to the scale differences in the uncertainty at the right tail versus the rest of the distribution.

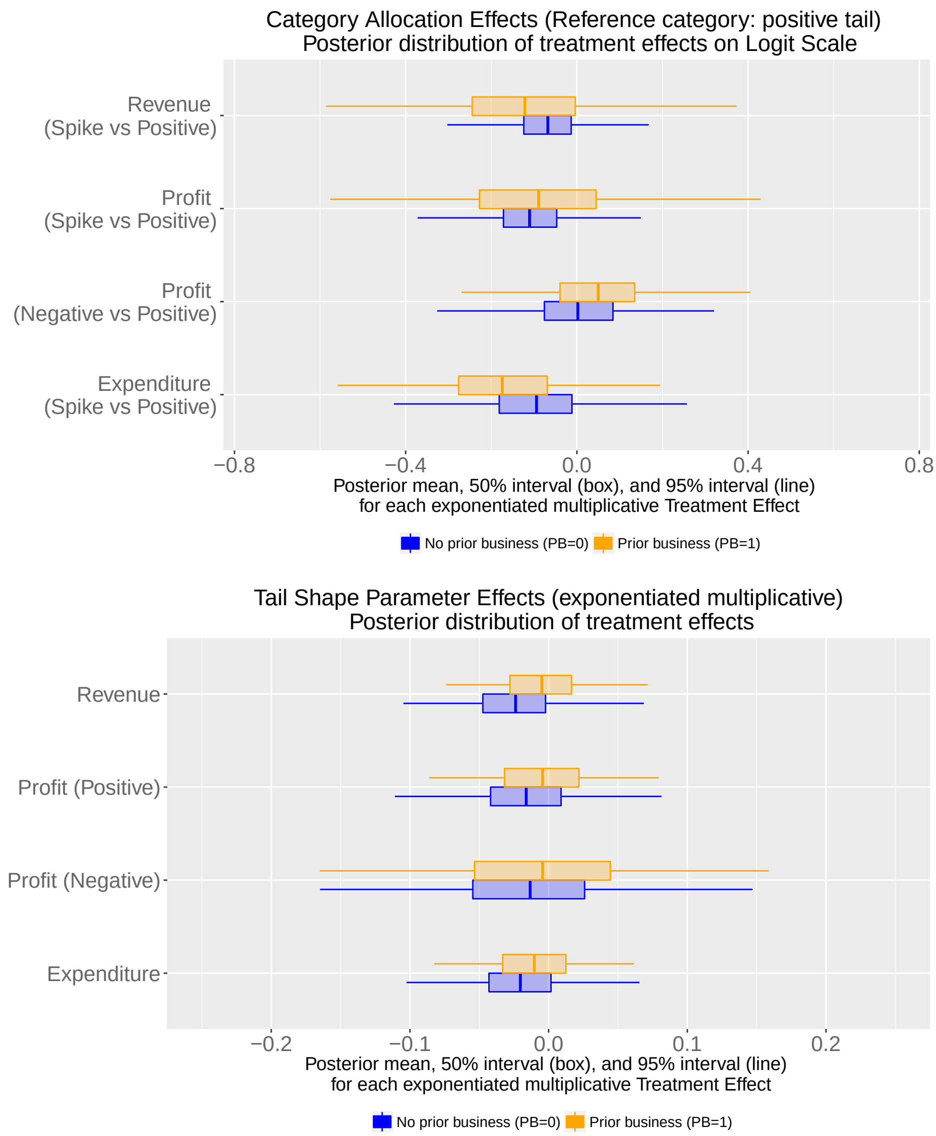

Figure 10: Upper panel: Posterior distributions for the logit treatment effects (π_j) on category assignment split by prior business ownership. These treatment effects are specified as an exponentiated multiplicative factor on the control group proportion of households in the category: if $\tilde{\pi_j} = 0$ the effect is zero, if $\tilde{\pi_j} < 0$ the treatment increases the proportion of households in the positive tail relative to other categories. Lower panel: Posterior distributions for the Pareto shape treatment effects (κ_j) in each site. These treatment effects are specified as an exponentiated multiplicative factor on the control group scale parameter: if $\tilde{\kappa_j} = 0$ the effect is zero, if $\tilde{\kappa_j} = 0.7$ the effect is a 100% increase in the scale parameter.

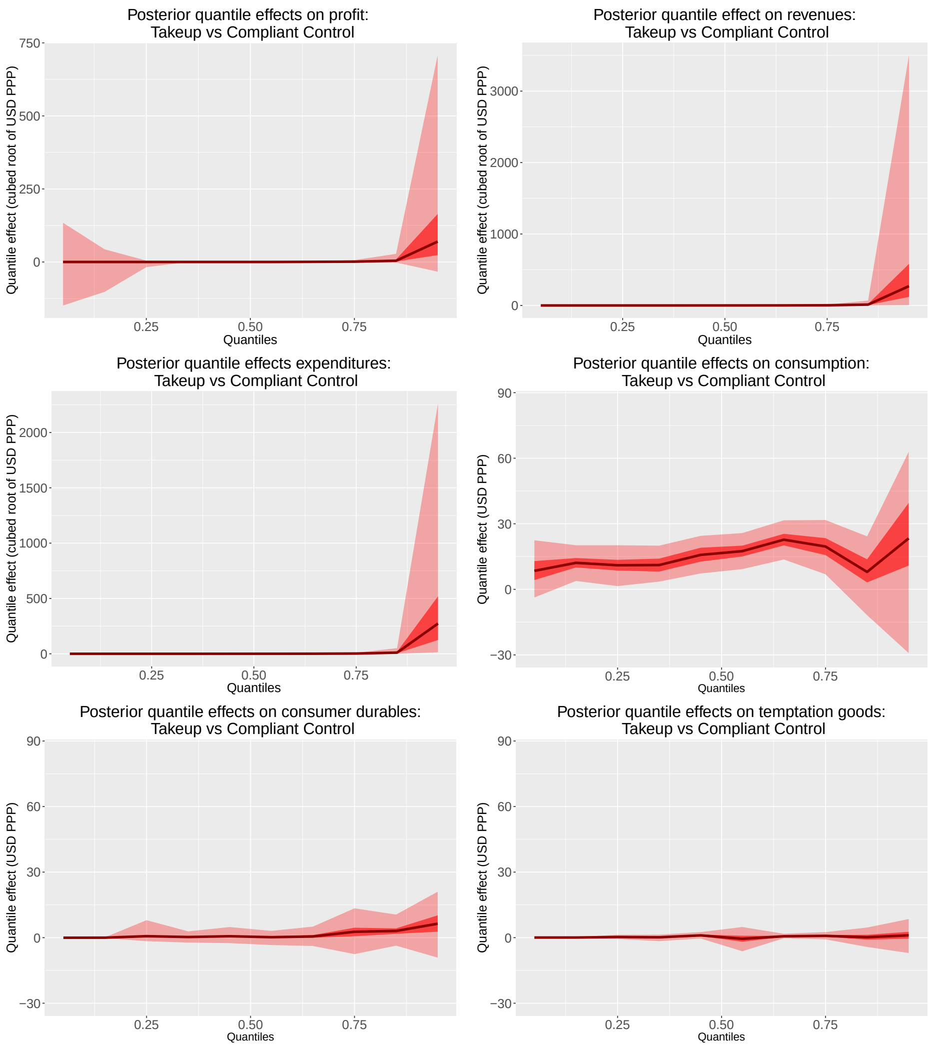

Figure 11: General Quantile Treatment Effect Curves for Business Outcomes: Treated households who took up vs Compliant control households who did not take up. This effect should overestimate the true impact of microcredit on those who take it up in a simple selection framework. Consumption variables are in USD PPP per two weeks, business variables are in cubed root of USD PPP per two weeks due to the scale differences in their uncertainty intervals. The dark line is the median, the opaque bars are the central 50% interval, the translucent bands are the central 95% interval.

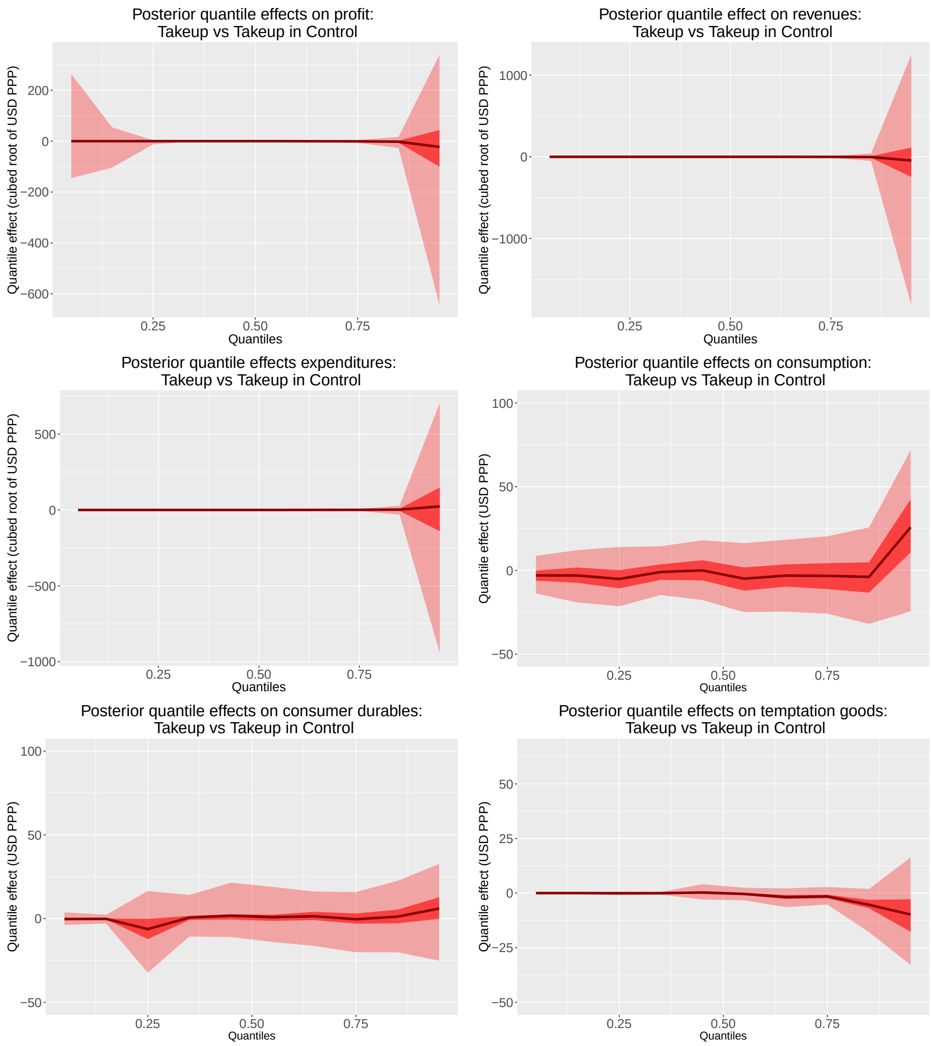

Figure 12: General Quantile Treatment Effect Curves for Business Outcomes: Treated households who took up vs Control households who took up. This effect should underestimate the true impact of microcredit on those who take it up in a simple selection framework. Consumption variables are in USD PPP per two weeks, business variables are in cubed root of USD PPP per two weeks due to the scale differences in their uncertainty intervals. The dark line is the median, the opaque bars are the central 50% interval, the translucent bands are the central 95% interval.

2.2 Flexible Tail Mixture Models

In this section I provide details of the two models fit with more flexible tail specifications than the simple Lognormal. First consider the Pareto-Lognormal model of Reed and Jorgensen (2004). Denote the Mills ratio of the standard Gaussian as $R(z) = (1 - \Phi(z))/\phi(z)$. The log of the Pareto-Lognormal is much more tractable computationally. By taking the log of equation 13 from Reed and Jorgensen (2004), I get the following likelihood function:

$$\ell(y|\alpha, \nu, \tau) := \log(\alpha) + log(\phi((y-\nu)/\tau)) + log(R(\alpha\tau - (y-\nu)/\tau) \tag{2.3}$$

Substituting this tail into the mixture model, for notational clarity given the main paper's notation I now denote these parameters by (A, N, t). Allowing microcredit to affect any of these distributional parameters in any way, I get the following:

$$\begin{aligned} y_{nk}|T_{nk} \sim\ & \Lambda_{1k}(T_{nk})\ell(-\log(y_{nk})|A_{1k} + TE_{A1k}T_{nk}, N_{1k} + TE_{N1k}T_{nk}, t_{1k} + TE_{t1k}T_{nk})) \\ & + \Lambda_{2k}(T_n)\delta_{(0)} \\ & + \Lambda_{3k}(T_n)\ell(\log(y_{nk})|A_{3k} + TE_{A3k}T_{nk}, N_{3k} + TE_{N3k}T_{nk}, t_{3k} + TE_{t3k}T_{nk})) \ \forall\ k \\ \text{where}\ \ \Lambda_{jk}(T_{nk}) = & \frac{\exp(\alpha_{jk} + \pi_{jk}T_{nk})}{\sum_{j=1,2,3}\exp(\alpha_{jk} + \pi_{jk}T_{nk}))} \end{aligned} \tag{2.4}$$

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

$$(\alpha_{1k}, \alpha_{2k}, \alpha_{3k}, \pi_{1k}, ...)' \equiv \zeta_k \sim N(\zeta, \Upsilon)\ \forall\ k \tag{2.5}$$

The priors for this model need to be strong to overcome the convergence issues noted in Reed and Jorgensen (2004), which on this data was particularly problematic on the parameter A. Following extensive testing and discussion with computational experts, I chose the following priors with a view to computational performance.[1]

$$
\begin{aligned}
A &\sim N(3,2) \\
\zeta \backslash A &\sim N(0,3) \\
\Upsilon &\equiv diag(\nu_\Upsilon)\Omega_\Upsilon diag(\nu_\Upsilon)' \\
\nu_\Upsilon &\sim \text{halfNormal}(0,3) \\
\Omega_\Upsilon &= I_{|\zeta|} \\
\alpha_{mk} &\sim N(0,5).
\end{aligned}
\tag{2.6}
$$

Even with these quite specific priors, this model still represents a strict relaxation of the LogNormal model fit in the paper; while the priors are stronger conditional on the given parameters, the parameters themselves construct a weaker structure on the data. However, while the convergence issues are mitigated, they are not eliminated: the "Rhat" criteria statistics from this model are indeed further from 1 than those of the LogNormal tail model, indicating poorer convergence and less reliable posterior inference despite these priors (Gelman and Rubin 1992). For this reason I minimize focus on this model in the main paper.

To avoid the convergence issues without having to employ such strong and specific priors, it is possible to employ the original insight from the methods section again and split up the tail into two components with disjoint supports: a Lognormal for the component with support adjacent to zero, and a Pareto for the extremal tail component.[2] This leads to the following "composite tail" likelihood with the Pareto location parameter ι naturally taking the form of the breakpoint or cutoff location:

$$
Composite(y|\iota,\rho,\mu,\sigma) := \mathbb{1}\{y \leq \iota\} Lognormal(y|\mu,\sigma) + \mathbb{1}\{y > \iota\} Pareto(y|\iota,\rho)
\tag{2.7}
$$

The challenge in practice is how to define the cutoff location ι in a hierarchical context, as one can no longer rely on the convenient scale-invariance of the cutoff location being zero. For tractability, in light of potential convergence issues, I do not allow the proportion of data in the two tail components to change at all in this model and I estimate it before the rest of the model; this two-step procedure is not ideal but it is computationally advantageous. I have defined the cutoff for the microcredit data as at the 80th quantile of the positive continuous tail and the 20th quantile

[1]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

[2]Once again I think Ulrich Müller and Andriy Norets for this insight.

of the negative continuous tail within each site, which corresponds to a model in which 80% of the data in every tail takes a LogNormal form, and the most extremal 20% of draws take a Pareto shape. In practice this is quite easy to implement: one uses any well-behaved quantile estimator, frequentist or Bayesian, within each tail to generate $\hat{\iota}$, with no inferential problems as this data is continuous. Then, one fits the mixture model with the tails taking the form of the composite model above, with $\hat{\iota}$ treated as data. I use the original priors from the main model on all the hyperparameters. To recover the quantiles, one uses the same Castellaci (2012) method, noting that one must rescale the cutoff quantile by 0.2 in the negative tail and 0.8 in the positive tail to determine the "average" cutoff ι at the superpopulation level. Given the suboptimal two-step nature of this procedure I do not focus on this model in the main results.

2.3 Running the Rubin Model Quantile by Quantile

Table 8: Profit: Results of running the Rubin (1981) model quantile by quantile

Quantile:	94th	95th	96th
Partial Pooling			
Bosnia	280.5	255.6	251.7
	(39.1,524.6)	(65.7,445.3)	(-25.1,535.9)
India	-16.3	-16.9	-19.4
	(-53.6,21.5)	(-63.6,29.4)	(-58.4,20.2)
Mexico	-0.1	19.7	20.5
	(-15.6,15.2)	(-0.5,40)	(1.7,39)
Mongolia	0	-0.1	-0.5
	(-0.7,0.7)	(-1.4,1.1)	(-2.6,1.6)
Morocco	95.6	87.3	157.9
	(4.4,188)	(-43.6,217.3)	(4.2,311.6)
Ethiopia	5.5	3.6	4.3
	(-1.8,12.8)	(-7.2,14.4)	(-12.7,21.3)
Philippines	339.8	454.1	681.7
	(-24.9,705.2)	(-16.6,916.4)	(168.3,1208.1)
Average	2.3	7.9	9.5
	(-11.2,59.2)	(-13.3,99.2)	(-19.6,141.8)

Notes: All units are USD PPP per two weeks. Estimates are shown with their 95% uncertainty intervals below them in brackets. These models had difficulty converging and likely do not represent a good fit to the data. This may be because the Gaussian approximation to the sampling error is unlikely to hold for this data given its extreme kurtosis.

Table 9: Consumption: Results of running the Rubin (1981) model quantile by quantile

Quantile:	5th	15th	25th	35th	45th	55th	65th	75th	85th	95th
Partial Pooling										
Bosnia	-4.6	-2.2	-1.4	-2.2	-0.7	1.7	2.1	-1.6	-9.4	35.5
	(-10,0.8)	(-11.7,3.4)	(-10.7,5.3)	(-13.2,4.3)	(-8.9,9.9)	(-8.8,11)	(-19.1,11.2)	(-41.2,13.7)	(-77.1,18.5)	(-57,177)
India	-0.6	-0.6	-1.3	-1.5	-1.3	0.1	3.1	4.6	6.7	34.2
	(-6,5.3)	(-5.4,3.8)	(-6.8,3.5)	(-6.1,3.3)	(-6.3,4.1)	(-7.6,5.6)	(-4.4,8.9)	(-5.1,13.7)	(-7.3,20.8)	(-4.9,78.2)
Mexico	-8.4	-0.9	-1.4	-1.8	-1	2.9	4.5	8.9	11	16.3
	(-12.9,-3.9)	(-5,2.7)	(-5.5,2.3)	(-5.5,2)	(-5.4,3.7)	(-1.8,8.2)	(-0.1,9.5)	(1.5,17.5)	(0.3,22.7)	(-6.9,38.6)
Mongolia	1.8	0.9	-0.2	-1.1	-1.6	1.7	3.2	3.4	-0.4	40.3
	(-8.9,19.2)	(-6.4,12.8)	(-8.1,10.6)	(-8.1,11.7)	(-15.4,7.2)	(-8.7,10.7)	(-12.2,15.1)	(-22.5,22)	(-45.2,28.7)	(-40.5,157.8)
Morocco	0	2.5	1	-1.4	-1.7	2.4	3.5	2.4	-2.9	-37.4
	(-6.3,7.4)	(-2.4,9.5)	(-3.8,8)	(-5.9,3.2)	(-8,3.6)	(-3.4,8.8)	(-4.7,11.3)	(-9.5,12.4)	(-20.2,11.8)	(-89.8,14.4)
Average	-2.6	-0.2	-0.6	-1.6	-1.3	1.6	3.1	4	0.8	17.3
	(-11.2,7)	(-6.1,6.4)	(-6.8,5.3)	(-6.6,3.6)	(-9.2,4.5)	(-5.8,7.6)	(-7.5,11.2)	(-10,25.5)	(-27.9,19.7)	(-56.2,95.7)

Notes: All units are USD PPP per two weeks. Estimates are shown with their 95% uncertainty intervals below them in brackets.

2.4 Leaving Out Certain Studies

Across the different studies, both an eyeball test and the results of the main analysis show that the underlying data are on very different scales, and that the control groups look quite different. This makes the main conclusion that the quantile effects are quite similar for most of the distribution even more striking. However, one might be concerned that a single study with a particularly large or small scale is driving or unduly affecting the results. In this case Mongolia, with its much smaller scale than all other studies, or Bosnia with its unique lack of negative profit observations, are the main concerns.

I have re-run the analysis leaving out Bosnia and Mongolia respectively. The results are shown in the graph below with the main results for comparison. Leaving out Bosnia changes virtually nothing; leaving out Mongolia makes the results much more uncertain and somewhat more positive, but still displays the same fundamental pattern and substantive conclusions of the main analysis.

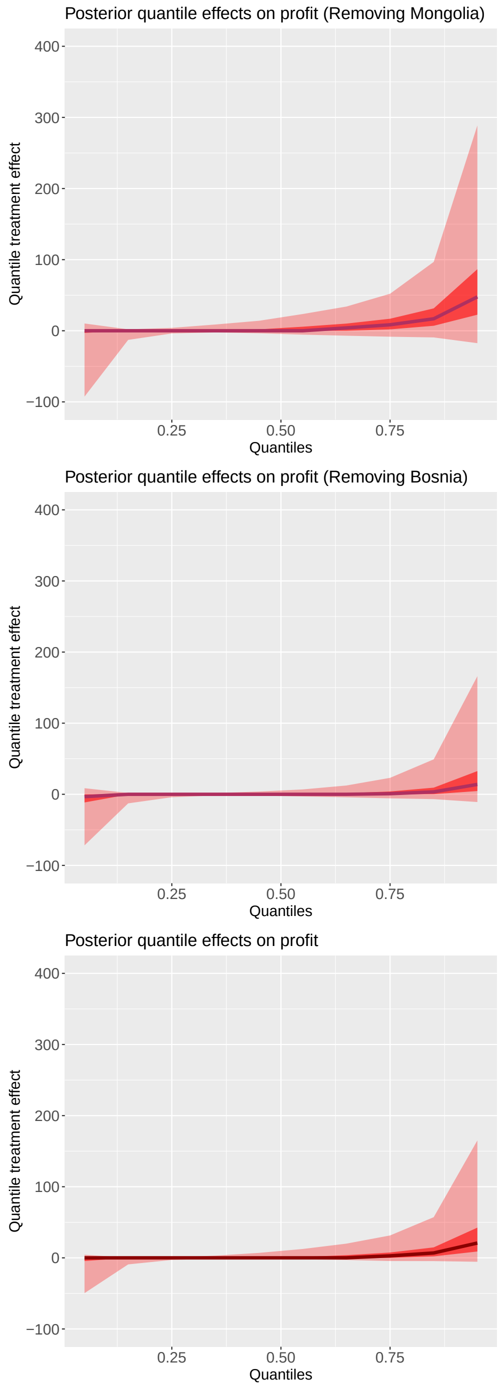

Figure 13: Average quantile treatment effects across all settings for profit (USD PPP per two weeks) without Mongolia (graph 1), without Bosnia (graph 2), and main results with all sites for comparison.

2.5 Trimming the data

With such extreme kurtosis values, it would be of interest to understand whether removing the largest 0.5% of values from the data set as a whole substantially impacts the inference. I examine the positive tail as this is the location of both the greatest uncertainty and greatest potential for positive effects. The table below shows the inference on the lognormal tail parameters for the profit data with these top positive values trimmed out. The posterior mean intervals on these parameters are reasonably stable across the original and trimmed data sets. While the lognormal scale parameters are slightly smaller, and τ_2 has most notably been reduced from approximately 0.077 to 0.057 indicating the important role of the extremal upper tail in generating even these results, this is within half a standard deviation of the original estimate.

Table 10: Profit Tail Inference from Trimmed Data (top 0.5% removed)

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
μ_1	3.228	0.025	0.810	1.733	2.747	3.214	3.679	4.828	1,038.916	0.999
μ_2	3.795	0.034	0.846	2.119	3.264	3.809	4.317	5.449	603.721	1.003
τ_1	0.096	0.003	0.092	-0.078	0.044	0.095	0.146	0.277	1,265.345	1.000
τ_2	0.057	0.001	0.047	-0.037	0.030	0.057	0.083	0.148	1,946.863	1.000
σ_{μ_1}	1.835	0.025	0.784	0.912	1.303	1.640	2.150	3.844	1,002.582	1.000
σ_{μ_2}	2.178	0.024	0.799	1.189	1.639	1.996	2.499	4.092	1,094.257	1.000
σ_{τ_1}	0.111	0.004	0.123	0.003	0.033	0.074	0.146	0.435	1,082.284	1.001
σ_{τ_2}	0.073	0.002	0.059	0.003	0.029	0.060	0.099	0.219	957.807	1.001
σ_1^c	0.453	0.005	0.144	0.181	0.373	0.446	0.526	0.759	837.234	1.004
σ_2^c	0.179	0.004	0.106	-0.049	0.116	0.180	0.242	0.388	868.749	1.003
σ_1^t	0.023	0.002	0.089	-0.154	-0.025	0.025	0.069	0.215	1,370.381	1.001
σ_2^t	0.001	0.001	0.026	-0.051	-0.014	0.001	0.015	0.054	1,624.716	1.001
$\sigma_{\sigma_1^c}$	0.307	0.006	0.169	0.126	0.199	0.264	0.364	0.737	912.078	1.004
$\sigma_{\sigma_2^c}$	0.268	0.004	0.112	0.142	0.195	0.242	0.309	0.530	992.346	1.003
$\sigma_{\sigma_1^t}$	0.164	0.004	0.122	0.037	0.091	0.136	0.202	0.448	1,211.250	1.000
$\sigma_{\sigma_2^t}$	0.038	0.001	0.032	0.002	0.016	0.031	0.051	0.118	1,026.529	1.002