Erasmus University Rotterdam
Erasmus School of Economics
Bachelor Thesis - BSc^2 Econometrics/Economics

Estimating the shape and evolution of the billionaire wealth distribution using Bayesian statistics

Philipp Höcker
564398

Supervisor:	Prof. dr. Phyllis Wan
Second assessor:	Jens Klooster
Date:	1st of July 2024

Abstract

We present Bayesian modelling approaches to the estimation of the distribution of billionaire wealth using the Forbes list of billionaires from 1997 to 2023. Models with Pareto, Truncated Weibull and Generalised Pareto distributions are considered. First, we show that measurement error has no important effect on the conclusion that the Pareto distribution is a poor fit to the billionaire wealth distribution. We compare Bayesian models that impose regularisation on the shape parameter of the wealth distribution to frequentist estimation techniques. As evaluated by the Kolmogorov-Smirnov goodness of fit test, Bayesian models based on the Generalised Pareto distribution provide the best fit to the empirical billionaire wealth distribution. We show that the Pareto distribution is a good fit for a smaller subset of top wealth, but a poor fit to overall billionaire wealth, as can be seen by the empirical cumulative hazard rate. We further introduce covariates for the variation in the scale of the wealth distribution, and conduct out-of-sample forecasts. We show that over half of the time-series variance in the change in scale of the top wealth distribution can be explained by changes in gross domestic product per capita and returns of the regional stock market. In forecasting, the Weibull distribution models provide the best forecasts for mean wealth. Using comparisons of results across countries, we draw conclusions on the effects of tax policies on top wealth.

1 Introduction

There will be the world's first trillionaire by 2034. At least, this is a prediction of Oxfam in their most recent "Inequality Inc." report (Riddell et al., 2024). Although the purpose of this statement may be to simply draw attention to global inequality, rather than to make an optimal prediction, it does raise the question of how top wealth can be studied from a statistical point of view.

1.1 Literature

Much research has been conducted on the distribution of wealth. Prominently, the 'Pareto principle', also known as the '80-20 rule', states that 20% of the population controls 80% of the wealth. Originally, Pareto (1896) used this principle to describe inequality of wealth and income. More recently, Thomas Piketty's *Capital in the Twenty-First Century* (2014) emphasises the drastic increase in wealth concentration in the top 1% in the recent decades.

For the larger part of history, models for the wealth distribution have both assumed and motivated Paretianity. Stiglitz (1969) proposes a simple capital accumulation model of an economy with inheritance. In this model, the upper tail of the wealth distribution possesses a Pareto shape. Expanding, Atkinson and Harrison (1979) provide an explicit formula for the Pareto index in Stiglitz' model as a function of population growth, taxes and savings, and relate it to empirical data from the United Kingdom. With a stronger focus on the very top end of the wealth distribution, both Drăgulescu and Yakovenko (2001) and Klass et al. (2006) show that the distribution of billionaires in the United States (US) and globally, respectively, is well described by a power law. This suggests a common pattern in the top wealth distribution across countries, with only a difference in the magnitude of parameters. Finally, Vermeulen (2018) utilises rich lists about billionaires to demonstrate Paretianity in the top wealth distribution.

However, the consensus about the Pareto shaped wealth distribution has been opposed. Using data from the Forbes list of billionaires, Teulings and Toussaint (2023) argue that the Pareto distribution is a poor fit for top wealth above one billion dollars. The authors develop a new test for Paretianity that utilises the $\hat{\mathcal{R}}_k$ statistic dependent on the sample moments of observations. Further, they observe that the empirical cumulative hazard rate of log wealth is exponentially increasing, and thus argue that a truncated Weibull distribution with this property provides a better fit. In fact, Teulings and Toussaint are not the first to investigate models using the Weibull distribution. Jacobi and Tzur (2020) study the fit of the Weibull distribution to wealth, but come to the conclusion that a Burr XII distribution is a comparatively better fit to the data.

A natural extension to the Pareto distribution is the Generalised Pareto Distribution (GPD). The Generalised Pareto Distribution finds a strong mathematical foundation in an Extreme Value Theory (EVT) theorem from Pickands III (1975), Balkema and De Haan (1974), guaranteeing that for a large class of distributions, the conditional excess distribution over a sufficiently high threshold converges to a GPD. Based on this result, one may expect the distribution of excess wealth above one billion dollars to be well modelled by a GPD. Correspondingly, Blanchet et al. (2022) use generalised Pareto curves to model the overall wealth distribution. With a stronger focus on top wealth, Charpentier and Flachaire (2022) argue that the GPD fits the data better than a simple Pareto model.

In the context of top wealth, measurement error is an issue that must be addressed. Teulings and Toussaint (2023) argue that their newly defined $\hat{\mathcal{R}}_k$ test statistic preserves important properties under unbiased and homoskedastic measurement error. To adapt any test statistic to measurement error, Capehart (2014) suggests to estimate measurement error empirically by comparing wealth estimates for the same individuals across different sources.

Finally, the growth in the wealth of the world's billionaires has been considerable and persistent over the last two decades, as noted by Bagchi and Svejnar (2015) and Flanigan and Freiman (2022). The causes for this growth are generally assumed to be economic growth and returns from capital markets. For instance, Gandhi and Walton (2012) display via exploratory analysis the seeming correlation between Indian billionaire wealth and Indian gross domestic product (GDP), as well as the Indian stock market. Cross-sectionally, differences in the number of billionaires across countries can largely be explained by differences in GDP per capita too, as shown by Popov (2018).

1.2 Research question

The literature in Section 1.1 exhibits three key gaps. First, while the measurement error relevant for top wealth data is often acknowledged, it is rarely addressed. Second, on the topic of which distribution best fits top wealth, most studies on billionaires from the later 2010s reject the earlier consensus of Paretianity. It is therefore worth studying the top end of the wealth distribution using Forbes' billionaire lists not only in a single country in a given year, but across the world and over time. Third, while there has been work done in describing the evolution of top wealth over time, there have been no attempts to statistically measure the effects of macroeconomic variables on the wealth distribution. Covariate analysis is limited to cross-sectional settings, but has not yet been performed in a panel setting.

Summarising, there is disagreement on the shape of the top wealth distribution, and a lack of studies of its evolution over time. Thus, the following research question arises.

Research Question

What is the shape of the wealth distribution of the world's billionaires, how does it change over time, and what factors underly the changes?

In order to resolve the open questions mentioned above, this research innovates in several ways. First, on the topic of measurement error in top net worth, this research adapts the approach of Capehart (2014) to not only estimate the measurement error on billionaire wealth, but to also study its impact on the $\hat{\mathcal{R}}_k$ statistics from Teulings and Toussaint (2023). The Paretianity tests are improved by using the more powerful Kolmogorov-Smirnov goodness-of-fit test, as suggested by Chu et al. (2019).

To infer which distribution best describes billionaire wealth, this research considers not only a Pareto model, but also a truncated Weibull and a Generalised Pareto model. As the literature suggests common patterns in the shape of the wealth distribution around the world, this research innovates by estimating models for top wealth in a Bayesian setting. As such, we allow for a form of regularisation when characterising the distribution of billionaire wealth.

Finally, the evolution of the top wealth distribution over time has received very limited attention thus far. Hence, we consider Bayesian time series models that incorporate stock market returns and GDP growth into the data generating processes. We let the covariates take influence on the unobserved scale parameter of the wealth distribution. This allows conclusions on the economic importance of the aforementioned covariates. Further, it facilitates distribution forecasts of future wealth, an exercise not touched upon in the literature thus far.

In brief, we find that the global overall increase of wealth, and subsequent increase in the number of billionaires, has resulted in a wealth distribution among billionaires that can no longer be characterised as Pareto. However, when considering a smaller subset of the world's wealthiest, Paretianity seems to still be valid. The best fit to the top wealth distribution is provided by a Bayesian model utilising the Generalised Pareto Distribution, closely followed by the Weibull sister model.

This research finds that over 60% of the variance in the changes of the scale of the top wealth distribution over time can be explained by changes in GDP and returns of the stock market. Billionaire wealth is positively related but inelastic to both the stock market and GDP growth. Out-of-sample forecasts for mean billionaire wealth using a Bayesian time series model based on the truncated Weibull distribution outperform random walk forecasts. This suggests that top wealth is predictable to the extent that the macro-economy is.

1.3 Scientific and economic implications

Understanding the shape of the top wealth distribution is of scientific importance. Besides simple interest in whether Bayesian methods are more adequate to understand billionaire wealth than conventional frequentist techniques, it allows conclusions on how wealth is generated. For example, it indicates whether the upper tail of wealth is mainly the result of economic diffusion processes - combinations of return on capital and exogenous income - that can be modelled by random difference equations as in Gabaix (2009), or whether it originates from different data generating processes. If top wealth is Pareto, it is evidence in favour of diffusion processes, but in case of other, economic theory must be revised.

In direct relation to diffusion processes, the research is also of social relevance. Kesten (1973) and Goldie (1991) show that for random difference equations, only the renewal rate determines the shape of the upper tail. As a direct consequence, one may conclude that the shape of top wealth is only impacted by taxes on capital, such as a wealth tax, but not by taxes on income.

In addition, the value of shape parameters is relevant for optimal taxation rates as well. Saez and Zucman (2019) show that the revenue maximising wealth tax rate depends on the heavy-tailedness of the wealth distribution.

Moreover, estimates of the influence that covariates have on the high tail of wealth do not only provide economic interpretation, but also a practical advantage. For governments using a wealth tax, they can provide tax revenue forecasts for many years ahead by making simplifying assumptions on economic growth. The coefficient estimates also help understanding the risk of changes in tax revenue during economic downturns.

The remainder of this paper is organised as follows. Section 2 presents the data used in this study, together with key exploratory insights. Section 3 outlines the estimation and evaluation methods required for the analysis. Section 4 presents the results in detail, and Section 5 presents economic implications from the results. Section 6 summarises the main findings and concludes.

2 Data

2.1 Billionaire lists

Similar to Teulings and Toussaint (2023), we use the *Forbes List of Billionaires* for the years 1997-2023[1]. Among other information, Forbes offers an annual estimate of an individual's net worth, rounded to 100 million $US, as well as the citizenship of individuals. We can classify individuals into regions based on nationality, using the geographical classification from Teulings and Toussaint (2023), shown in Appendix A, Table 4.

Table 1: Summary statistics of Forbes billionaires data grouped by region

Region	Mean N	Mean Net Worth	Mean Log Net Worth	Min N	Max N
Central Eurasia	78.2	3.7	0.9	5	132
China	211.4	3.2	0.8	1	698
East Asia	133.3	2.7	0.8	2	306
Europe	248.8	4.1	1.1	3	509
India	61.9	4.0	0.9	1	169
Middle East	53.2	2.9	0.8	1	90
North America	471.7	4.2	1.0	176	800
South America	55.5	4.0	0.9	1	113
Rest of World	11.1	3.5	1.0	2	23

Note: 'N' refers to the number of billionaires. Mean values are averaged by pooling observations from all years, 1997 up to and including 2023.

Table 1 shows some summary statistics for the Forbes dataset. There are significant differences across regions. There are many more billionaires in North America than in Europe, despite roughly equal populations. Further, billionaires in some regions are on average richer than billionaires in other regions. As besides North America, not all regions have a substantial number of billionaires in the beginning of the sample, all models in this research are only estimated starting in 2005.

More importantly, the same billionaires show up every year. Every year, around 80% of the world's billionaires were also billionaires the year before. Details can be found in Appendix B, Table 6. This implies that observations of the top wealth distribution are dependent across time. From a modelling perspective, this encourages a form of regularisation when estimating distribution parameters to prevent overfitting.

Further, we use a list of the 500 richest people in the year 2021 based on the Bloomberg Billionaires Index[2]. Similarly to Forbes, Bloomberg estimates net worth to a precision of 100 million $US. This list is leveraged to estimate measurement error. Note that since this dataset orders billionaires and only preserves the top 500 observations, the means are much higher than in Forbes' data. Details can be found in Appendix B, Table 7.

[1]Retrieved from: `https://www.kaggle.com/datasets/guillemservera/forbes-billionaires-1997-2023`
[2]Retrieved from: `https://www.kaggle.com/datasets/frtgnn/500-richest-people-2021`

2.2 Covariates

In addition to data on individuals' net worth, the literature prompts us to focus on two key covariates to model changes in the top wealth distribution: GDP per capita and stock indices.

For the former, we use annual data for GDP per capita from the World Bank on a country level, reaching until 2022[3]. As for most use cases of this research, we aggregate countries into larger sub-regions to have a sufficient number of observations in a group, we compute the population weighted GDP per capita in a sub-region. The sub-region classification of countries follows Teulings and Toussaint (2023), and can be found in Appendix A, Table 5. Global population data are available from Worldometers[4].

Furthermore, we use several stock market indices. When possible, we use a market index with a geographical focus on the countries of a respective sub-region. That is, the CAC40 for France, the DAX for Germany, the FTSE100 for the British Islands, the MOEX for Russia, the NIFTY for India, the OMX40 for Scandinavia, the S&P500 (SPX) for the United States, and the SSE for China. For all other sub-regions, we opt to use the MSCI World for simplicity. Historical data for all of the above indices are available on Yahoo Finance's Quote page at a monthly frequency[5].

3 Methodology

This section outlines the different estimation techniques used for the distribution of billionaire wealth. We begin with frequentist techniques, and subsequently move to more appropriate Bayesian estimation techniques, ultimately integrating covariates into the models.

3.1 Frequentist models

3.1.1 Notation

We follow the notation of Teulings and Toussaint (2023). Let absolute wealth be denoted by X. The lower bound on wealth is $\Omega = 10^9$ (one billion \$US). We define $W \equiv \frac{X}{\Omega}$, $x \equiv \ln X$, $\omega \equiv \ln \Omega$, and hence $w \equiv x - \omega = \ln W$. Further, let Y denote the excess wealth of an individual beyond a billion \$US, then we denote $Y \equiv X - \Omega$. To clarify the meaning of parameters when denoting probability distributions, all are shown in Appendix C.

3.1.2 Pareto and truncated Weibull

In Teulings and Toussaint (2023), the authors investigate two distribution fits. Specifically,

$$Y \sim \text{Pareto}(\alpha), \tag{1}$$

and

$$Y \sim \text{TruncatedWeibull}(\gamma, \alpha).$$

[3] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD
[4] https://www.kaggle.com/datasets/whenamancodes/world-population-live-dataset
[5] https://finance.yahoo.com/lookup

Naturally, the thresholds for the Pareto and truncated Weibull distribution are chosen to be 1. The parameters of the truncated Weibull distribution can be estimated by maximum likelihood (ML) and using the parameter relationships shown by Teulings and Toussaint (2023).

Extreme Value Theory (EVT) proposes a natural extension to the model in (1): the Generalised Pareto Distribution (GPD). If $Y \sim \text{GPD}(\gamma, \sigma)$, we have the probability density function (PDF) of Y,

$$f_{(\gamma,\sigma)}(y) = \frac{1}{\sigma}\left(1 + \frac{\gamma}{\sigma} \cdot y\right)^{\left(-\frac{1}{\gamma}-1\right)},$$

for $y \geq 0$ when $\gamma \geq 0$, and $0 \leq y \leq -\frac{\sigma}{\gamma}$ when $\gamma < 0$. It may be noted that the regular Pareto distribution is a special case of the GPD.

The GPD appears as the natural excess distribution function for a large class of probability distributions. In particular, define $F_u(y) = P(Y - u \leq y | Y > u)$ the cumulative density function (CDF) of the excesses. Then an important result of EVT stemming from Balkema and De Haan (1974) and Pickands III (1975) states that this excess distribution function can be approximated by the GPD.

Theorem 1 (Balkema and de Haan, 1974, Pickands, 1975)**.** *For a large class of distributions, a function $\sigma(u)$ can be found such that*

$$\lim_{u \to \bar{y}} \sup_{0 \leq y < \bar{y} - u} \left| F_u(y) - G_{\gamma,\sigma(u)}(y) \right| = 0,$$

where $G_{\gamma,\sigma(u)}$ is the cumulative distribution funtion of the GPD, $\bar{y}$ is the rightmost point of the distribution, u is the threshold and F_u is the excess distribution function.

Parameters of the GPD may generally be estimated by maximum likelihood for adequate values of $\gamma > 0$ (infinite supremum of the distribution). A separate technique for γ only is provided by the Hill estimator (Hill, 1975). In the context of this research, a natural choice of the threshold is one billion \$US.

As outlined multiple times before, one would expect the shape of the wealth distribution across countries and time to be roughly similar. When estimating any of the three frequentist models - Pareto, truncated Weibull, or Generalised Pareto - a form of regularisation is required for the shape parameter. In this case, as suggested in Teulings and Toussaint (2023), we can first jointly estimate shape and scale for all sub-region/year combinations, and then fix the shape parameter to a single value, in this case the median of all estimated shape parameters. In a second step, we then estimate the scale parameters.

3.2 Bayesian estimation

A key feature of the Forbes data is that we can divide billionaires into geographical sub-groups. It is reasonable to assume that the shape parameter of the top wealth distribution will be stable within a country across time, and similar across countries, while the scale may be totally different, for example due to differences in economic development.

Bayesian Hierarchical Models (BHMs) enable the integration of prior distributions at various levels within a hierarchical framework (Lindley & Smith, 1972). This capability is particularly advantageous for analysing group-level effects. A key benefit of BHMs in this scenario is their ability to leverage information across time and different geographical groups, particularly when dealing with limited data. In our case, the number of billionaires in a single year and region may be somewhat small, but the total number of billionaires is quite large.

Let θ be a vector of parameters of a model, for example the parameters of a distribution, $\mathbf{y}$ the observed dependent variable and X a set of observations from covariates. Then we are interested in finding $\theta \mid \mathbf{y}, X$, so the distribution of θ given our data. For some simple BHMs, it is possible to analytically derive the posterior distribution $\theta \mid \mathbf{y}, X$ by using conjugate priors. However, for more complex models, deriving the posterior distribution analytically becomes challenging or even infeasible. The key idea for numerically estimating this distribution relies on

$$\mathbb{P}(\theta \mid \mathbf{y}, X) = \frac{\mathbb{P}(\mathbf{y}, X \mid \theta) \cdot \mathbb{P}(\theta)}{\mathbb{P}(\mathbf{y}, X)} \propto \mathbb{P}(\mathbf{y}, X \mid \theta) \cdot \mathbb{P}(\theta).$$

While the numerator is easily computed, the denominator is not. Fortunately, when using the Metropolis-Hastings method to sample from the posterior distribution, the denominator cancels out in the ratio of likelihoods. Metropolis-Hastings is a specific type of Markov Chain Monte Carlo (MCMC) methods, a broader class of techniques that can numerically estimate the posterior distribution by generating a Markov chain of samples.

We present Bayesian hierarchical approaches for all three considered distributions: Pareto, Weibull and Generalised Pareto. We further present those in two settings: a cross-sectional estimation, where we estimate distributions year by year, ignoring information across years, and a panel approach.

In the coming sections, we always use the following notation. $Y_{i,j,t}$ is the excess wealth as defined in Section 3.1.1 of individual $i \in \mathcal{I}$, in sub-region $j \in \mathcal{J}$ at time (i.e. year) $t \in \mathcal{T}$.

3.2.1 Un-regularised cross-sectional estimation

Naturally, we can estimate Bayesian models for single combinations of a sub-region and year, allowing for different shape parameters across groups and across years. We estimate the cross-sectional Pareto BHM without regularisation with the following specification,

$$\begin{aligned}
Y_{i,j,t} \mid \alpha_{j,t} &\overset{\text{i.i.d.}}{\sim} \text{Pareto}(\alpha_{j,t}), \\
\alpha_{j,t} \mid \alpha, \beta &\sim \text{InverseGamma}(\alpha, \beta).
\end{aligned}$$

The choice of the Inverse Gamma distribution prior for α_j is natural as it is a conjugate prior. That is, it can be shown that the posterior distribution of α_j will also be an Inverse Gamma distribution. This is one of the rare cases where the posterior parameter distribution can be derived analytically, see the proof in Appendix D.

We estimate the cross-sectional Weibull BHM without regularisation with the following specification, for a given $t \in \mathcal{T}$,

$$\begin{aligned}
&Y_{i,j,t} \mid \gamma_{j,t}, \alpha_{j,t} \overset{\text{i.i.d.}}{\sim} \text{TruncatedWeibull}(\gamma_{j,t}, \alpha_j), \\
&\gamma_{j,t} \sim \text{InverseGamma}(\alpha_\gamma, \beta_\gamma), \\
&\alpha_{j,t} \sim \Gamma(\alpha_\alpha, \beta_\alpha).
\end{aligned}$$

Note that from here onwards, $\Gamma(\cdot)$ denotes the Gamma function, and $\Gamma(\cdot,\cdot)$ denotes the Gamma distribution. The choice for the priors in this case can be motivated as if one parameter is known (shape or scale), then the prior distribution of the other would again be a conjugate prior.

Finally, we estimate the cross-sectional GPD BHM without regularisation with the following specification, for a given $t \in \mathcal{T}$,

$$\begin{aligned}
&Y_{i,j,t} \mid \gamma_{j,t}, \sigma_{j,t} \overset{\text{i.i.d.}}{\sim} \text{GPD}(\gamma, \sigma_j), \\
&\gamma_{j,t} = \gamma^f_{j,t} - \frac{1}{2}, \\
&\gamma^f_{j,t} \sim \Gamma(\alpha_\gamma, \beta_\gamma), \\
&\sigma_{j,t} \sim \Gamma(\alpha_\sigma, \beta_\sigma).
\end{aligned}$$

As noted by Diebolt et al. (2005), there exists no Bayesian conjugate class for the GPD. However, the choice of priors is not arbitrary. In particular, Dombry et al. (2023) show that if the priors on the parameters fulfil some regularity conditions, then the posterior distribution is asymptotically normal with asymptotically nominal coverage. Two (shifted) Gamma distributions fulfil these conditions. These properties are important, as they ensure that one will obtain non-degenerate posterior distributions when sampling using Markov-Chain Monte Carlo.

The hyperparameters of the above models can be chosen to be informative or uninformative. In our case, it makes sense to impose informative priors on the shape parameters, as those can be assumed to be similar across countries, and uninformative priors on scale parameters. A possible way of doing this is to fit an Inverse Gamma and shifted Gamma distribution to all the maximum likelihood estimates across years and sub-regions of the Weibull and Generalised Pareto distributions respectively. Whilst an unconventional approach for setting hyperparameters in Bayesian statistics, it is adequate in this context to ensure that the prior predictive distribution of each model is plausible. For the Generalised Pareto Distribution, we know that the mean does not exist for shape values larger than one, and that the distribution is bounded for negative shapes, and hence a shape prior with the coverage focused on the interval from zero to one is desirable in the context of wealth data.

3.2.2 Regularised cross-sectional estimation

Note that the model specifications in Section 3.2.1 are even more flexible than the frequentist approaches, as they allow the shape parameters to vary not only across groups, but also across time within a group. However, this can produce strongly varying shape estimates within a group, which is an unrealistic model, and prone to overfitting to single years. A more parsimonious approach is to impose further regularisation by restricting the shape parameters of the Weibull and Generalised Pareto models to be constant across time. That is, for all j, we have

$$\gamma_{j,1997} = \gamma_{j,2000} = \ldots = \gamma_{j,2023}. \tag{2}$$

This allows for variation in scale (expansion or contraction of wealth), but estimates an individual γ for each sub-region, utilising all observations for that region across time.

3.2.3 Time series estimation

We already discussed the imposition of regularisation on parameters in the previous subsection. One may however criticise that even a regularised BHM with the restriction from Relation 2 ignores a key feature of the data. That is, that the wealth observations are serially correlated. Therefore, one would expect the scale parameter of the distribution of a particular sub-region not to change drastically from one year to the next.

Again, for the Pareto, truncated Weibull and Generalised Pareto Distributions, we model this additional aspect as follows. For the Pareto distribution, we specify

$$
\begin{aligned}
&Y_{i,j,t} \mid \alpha_{j,t} \overset{\text{i.i.d.}}{\sim} \text{Pareto}(\alpha_{j,t}), && \beta_j \sim \mathcal{N}(0, \sigma_j^{\beta}),\\
&\alpha_{j,t=0} \mid \alpha_\alpha, \beta_\beta \sim \text{InverseGamma}(\alpha_\alpha, \beta_\alpha), && \epsilon_{j,t} \sim \mathcal{N}(0, \sigma_j^{\epsilon}),\\
&\ln \alpha_{j,t+1} = \ln \alpha_{j,t} + x'_{j,t+1}\beta_j + \epsilon_{j,t+1}, && \sigma_j^{\beta}, \sigma_j^{\epsilon} \overset{\text{i.i.d.}}{\sim} \Gamma(1,1).
\end{aligned}
$$

For the Weibull distribution, we specify

$$
\begin{aligned}
&Y_{i,j,t} \mid \gamma_j, \alpha_{j,t} \overset{\text{i.i.d.}}{\sim} \text{TruncatedWeibull}(\gamma_j, \alpha_j), && \beta_j \sim \mathcal{N}(0, \sigma_j^{\beta}),\\
&\gamma_j \sim \text{InverseGamma}(\alpha_\gamma, \beta_\gamma), && \epsilon_{j,t} \sim \mathcal{N}(0, \sigma_j^{\epsilon}),\\
&\alpha_{j,t=0} \mid \alpha_\alpha, \beta_\beta \sim \Gamma(\alpha_\alpha, \beta_\alpha), && \sigma_j^{\beta} \sim \Gamma(1,1),\\
&\ln \alpha_{j,t+1} = \ln \alpha_{j,t} + x'_{j,t+1}\beta_j + \epsilon_{j,t+1}, && \sigma_j^{\epsilon} \sim \Gamma(1,1).
\end{aligned}
$$

For the Generalised Pareto distribution, we specify

$$
\begin{aligned}
&Y_{i,j,t} \mid \gamma_{j,t}, \sigma_{j,t} \overset{\text{i.i.d.}}{\sim} \text{GPD}(\gamma, \sigma_j), && \gamma_{j,t}^{f} \sim \Gamma(\alpha_\gamma, \beta_\gamma),\\
&\gamma_{j,t} = \gamma_{j,t}^{f} - \frac{1}{2}, && \beta_j \sim \mathcal{N}(0, \sigma_j^{\beta}),\\
&\sigma_{j,t=0} \mid \alpha_\sigma, \beta_\sigma \sim \Gamma(\alpha_\sigma, \beta_\sigma), && \epsilon_{j,t} \sim \mathcal{N}(0, \sigma_j^{\epsilon}),\\
&\ln \sigma_{j,t+1} = \ln \sigma_{j,t} + x'_{j,t+1}\beta_j + \epsilon_{j,t+1}, && \sigma_j^{\beta}, \sigma_j^{\epsilon} \overset{\text{i.i.d.}}{\sim} \Gamma(1,1).
\end{aligned}
$$

In other words, we assume that the log-shape parameter in one period equals, in expectation, the log-shape parameter from the previous period, plus the effect of some covariates. In the context of top wealth, the log structure can be well motivated. To begin, this parameterisation ensures that the shape parameters are always strictly greater than zero, thus ensuring a finite likelihood when estimating the models by Markov Chain Monte Carlo. In addition, the log model is intuitive from an economic interpretation point of view. For the sake of top wealth, we will consider three covariates: a constant (i.e. trend component), the log GDP growth, and the log return of the corresponding regional stock market index. Now consider the population mean of the Generalised Pareto Distribution $\mu_{GPD} = \frac{\sigma}{1-\gamma} \propto \sigma$. As the mean is clearly proportional to the scale parameter, the coefficients of a log-log model have a simple economic interpretation - they indirectly show the elasticity of mean billionaire wealth to the stock market and economic output. In other words, a 1% change in the stock market has a β% impact on the mean wealth of billionaires.

Again, the above models have to be estimated numerically - recall that unlike in conventional time series models, the scale parameters are now unobserved. Further, the high level of dependence of the model parameters makes sampling from the posterior distribution computationally expensive. In practice, we thus limit ourselves to an estimation window of at most 15 years. This hardly results in a loss of data, as besides the United States nearly no countries have a sufficient number of billionaires for reliable estimation prior to 2005 anyways. The exact time frames can be found in Appendix E.

Finally, the time series models offer a forecasting capacity. Of course, it would be unreasonable to assume that stock market returns can be forecasted accurately for one, if not more, years ahead. However, GDP growth numbers are usually forecasted to a reasonable degree of accuracy for a year in advance (in the absence of extreme shocks like financial crisis, that is). This yields a pseudo-forecasting exercise: fit a model until a given year with only a constant and the GDP per capita growth covariate, and assume future GDP growth is known. Then use the estimated parameters and the model specification to perform h-step ahead distribution forecasts.

3.3 Evaluating fit

This subsection focuses on different ways to evaluate the fit of the aforementioned models.

3.3.1 $\hat{\mathcal{R}}_k$ statistics

Teulings and Toussaint (2023) use the $\hat{\mathcal{R}}_k$ statistics defined in Equation 3 to test for Paretianity,

$$\mathcal{R}_k := \frac{\mathbb{E}\left[w^k\right]}{k! \cdot \mathbb{E}\left[w\right]^k}, \quad \hat{\mathcal{R}}_k := \frac{\overline{w^k}}{k! \cdot \overline{w}^k}. \tag{3}$$

The authors show that if w follows an exponential distribution, that is, if W follows a Pareto distribution, then all $\mathcal{R}_k$ statistics are one.

Theorem 2 (Teulings and Toussaint, 2023)**.** *Assume w is exponentially distributed. Then, for any integer $k \geq 1$, $\mathcal{R}_k$ defined in Equation 3 satisfies $\mathcal{R}_k = 1$, while $\mathbb{E}\left[\hat{\mathcal{R}}_k\right] = 1 + N^{-1}\left(\frac{(2k)!}{k!^2} - \frac{3k^2-k+2}{2}\right) + O\left(N^{-2}\right)$ and $\mathrm{Var}\left[\hat{\mathcal{R}}_k\right] = N^{-1}\left(\frac{(2k)!}{k!^2} - k^2 - 1\right) + O\left(N^{-2}\right)$, where N is the sample size.*

Further, Teulings and Toussaint (2023) give an explicit formula for the population mean of the truncated Weibull distribution.

Theorem 3 (Teulings and Toussaint, 2023)**.** *Assume W is TruncatedWeibull(γ, α) distributed. Then, the moments for $k \in \mathbb{N}$, $k > 0$ of W read $\mathbb{E}\left[W^k\right] = (\alpha\gamma)^k/\gamma\, e^{(\alpha\gamma)^{-1}}\Gamma\left(1 + k/\gamma, (\alpha\gamma)^{-1}\right)$, where $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function.*

3.3.2 Kolmogorov-Smirnov test

Using $\hat{\mathcal{R}}_k$ statistics to test for the GPD is unfortunately problematic, as the k-th non-central moment only exists if $\gamma < \frac{1}{k}$ (Hosking & Wallis, 1987). To test for generalised Paretianity, we must therefore resort to other tests, such as the Kolmogorov-Smirnov (KS) test as described in Chu et al. (2019). The Kolmogorov-Smirnov statistic for a sample of size n with empirical CDF F_n and null hypothesis CDF F is $D_n = \sup_x |F_n(x) - F(x)|$. In simple terms, the KS-statistic penalises the maximum difference between the CDFs. Note also that the KS-statistic can easily be modified to a two-sample test by replacing the analytical CDF by a second empirical CDF, then testing whether two samples came from the same distribution. This proves especially useful for evaluating the fit of Bayesian models, where no analytical posterior distribution is tractable.

3.3.3 Bayesian R-squared

For the time series models presented in Section 3.2.3, it is of interest to know what portion of the variance in the scale parameters is explained by our chosen covariates. As the year-on-year change in variance is estimated but unobserved, we must use a different metric. We resort to the proposal of Gelman et al. (2019) to use an alternative version of the R-squared, namely

$$\text{Alternative } R^2 = \frac{\text{Explained variance}}{\text{Explained variance} + \text{Residual Variance}} = \frac{V_{t=1}^{T-1} \ln \frac{\sigma_{t+1}^{\text{pred}}}{\sigma_t^{\text{pred}}}}{V_{t=1}^{T-1} \ln \frac{\sigma_{t+1}^{\text{pred}}}{\sigma_t^{\text{pred}}} + V_{t=1}^{T-1} \epsilon_{t+1}^{\text{pred}}}.$$

That is, we use the posterior predictive for the scale and residuals to compute a measure of explained variance, for which we obtain a distribution of R^2 values by doing so across all MCMC samples.

3.4 Measurement error

Data on billionaire wealth from Forbes suffer from two types of measurement error. First, the values are rounded to the nearest $US 100 million. Second, there is uncertainty in the estimates themselves. This can be seen when comparing Forbes' data from April 2021 to the data from Bloomberg in May 2021. Whilst Teulings and Toussaint (2023) show that symmetric and independent measurement errors of w have no impact on the $\hat{\mathcal{R}}_k$ statistics, they ignore rounding errors.

We consider the Bloomberg dataset on the world's 500 richest people. These data are meant to be a snapshot corresponding to the state of affairs in May 2021. The Forbes data from that year are from April, so the measurements of the same individuals should be quite close. Measurement errors cannot be observed, but we can obtain an idea of their magnitude by comparing measurements of the same person's net worth across both lists.

Capehart (2014) suggests a method to adjust statistical tests for this measurement error: fit a kernel density to the log-ratio between Forbes and Bloomberg estimates for individual billionaires, and draw measurement errors from this kernel density to compute the distribution of the test statistic when sampled from a given distribution contaminated by these measurement errors. Capehart shows that the significance of KS-tests shrinks under this procedure. We may apply the same procedure to the $\hat{\mathcal{R}}_k$ statistics.

4 Results

4.1 Assessing Paretianity under measurement error

We begin by presenting the results of testing for Paretianity using the methods from Teulings and Toussaint (2023), but adjusted for measurement error.

As discussed before, we proxy measurement error by comparing the lists from Bloomberg and Forbes in the same year, 2021. Figure 1 shows a histogram of the differences in measurements between Forbes and Bloomberg.

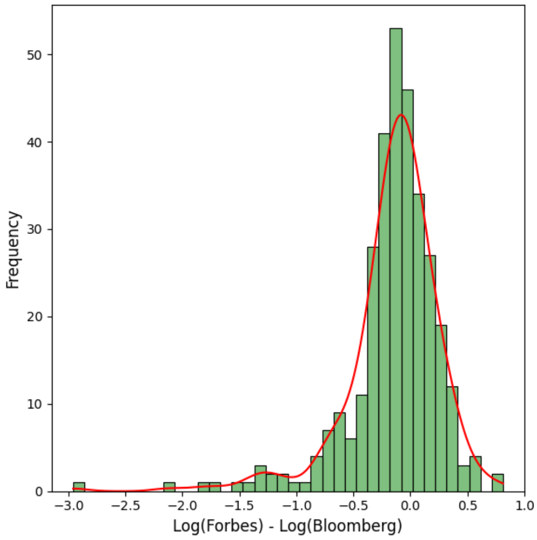


Figure 1: Difference between measurements of the same individuals between Forbes and Bloomberg with fitted kernel density, 2021

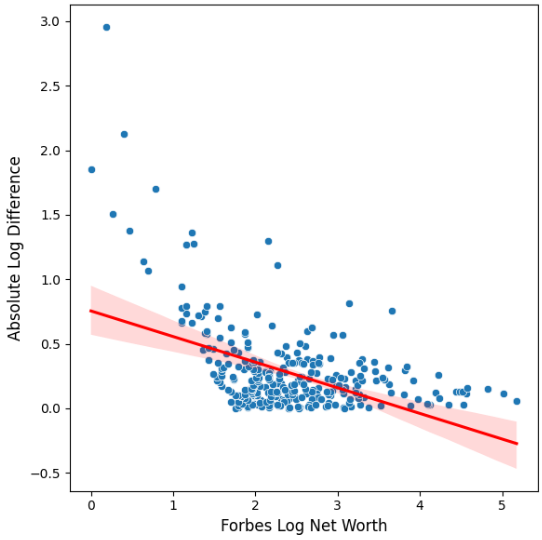


Figure 2: Absolute log differences between Forbes and Bloomberg against estimated log wealth, with OLS regression line

We may notice two issues. First, the log-error distribution is not symmetric: the left tail is much longer than the right tail. Second, its mean is far from zero: On average, Forbes estimates wealth to be around 15% higher than Bloomberg. This violates one of Teulings and Toussaint's (2023) assumptions of mean zero errors. The mean of the log error series is -0.15, which corresponds to a t-statistic of -6.54 and a P-value of 0.000 against the null hypothesis that the mean error is zero.

One may be curious whether some outlier errors can be explained, and should thus be removed from the data. However, upon examination of the top five most 'misestimated' individuals, both on the side where Forbes estimates higher than Bloomberg and vice-versa, no simple explanation for the drastic difference in wealth could be found, either by news exploration or exploratory analysis of the stock price evolution of companies owned by the respective people. The only individual for whom a clear explanation was evident was Ugur Sahin, whose shares in Covid vaccine manufacturer BioNTech soared 110% between April and May 2021. In general, there is no clear evidence that the perceived measurement error is driven by actual changes in wealth instead of disagreements between Forbes and Bloomberg.

Next, we examine whether the differences are homoskedastic. Figure 2 shows a scatterplot of absolute differences between log wealth measurements from the two sources. Clearly, the higher an individual's net worth is, the lower the disagreement on their log wealth. Indeed, the slope coefficient of the red regression line is significant with a P-value of 0.000. That is, Bloomberg and Forbes have stronger agreement on the net worth of richer billionaires. Therefore, we may conclude that the assumptions needed for the results from Teulings and Toussaint (2023) on measurement errors are not satisfied in billionaire data.

We can demonstrate how drastically the distribution of $\hat{\mathcal{R}}_k$ statistics changes when measurement error is added. Figure 3 shows, on the left, a plot of the distribution under the null hypothesis of Paretianity with the test statistic for the U.S. in 2021, under the assumption of no measurement errors. As we can see, the distribution under the null hypothesis is well centred at 1, and the test statistic of around 0.78 results in left-sided P-value of 0.000, clearly rejecting Paretianity. Now observe the right plot. Here, we estimate the measurement error distribution by first trimming the observed log measurement errors' top and bottom deciles, then shifting the location of the remaining errors such that their mean is zero, and then fitting a Gaussian kernel density using Fast Fourier Transform. The distribution of the test statistics under the null hypothesis is now bootstrapped by sampling from a Pareto distribution, and adding measurement errors from the kernel density to the simulated observations. As the effect of rounding to one decimal is negligible, it is omitted. The result is a test statistic distribution that is shifted far to the left, with a mean of 0.73. The null hypothesis of Paretianity can no longer be rejected from the left, but from the right with a P-value of 0.011. Therefore, even a very conservative measurement error has drastic consequences for the properties of the $\hat{\mathcal{R}}_k$ statistics.

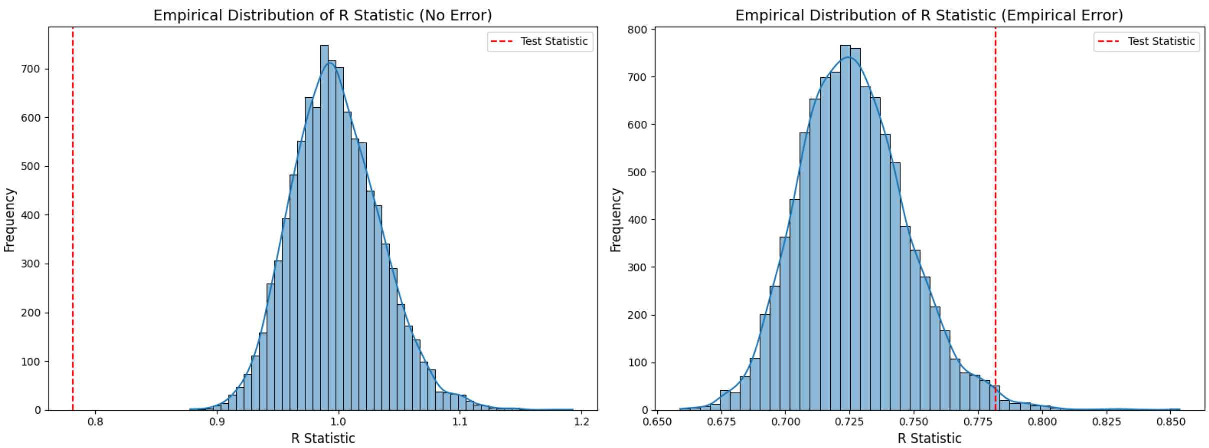
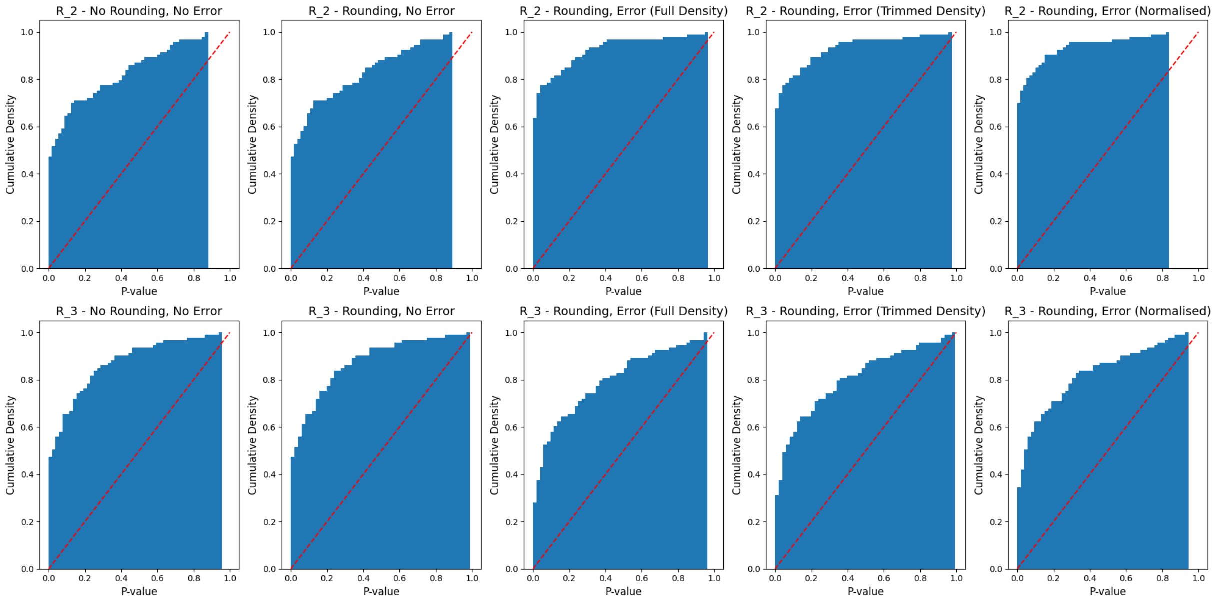


Figure 3: Comparison of distribution of the $\hat{\mathcal{R}}_2$ statistic with and without trimmed and centred measurement error for the 724 billionaires in the U.S. in 2021

Now of course the question arises: do the findings on measurement errors re-validate the Paretianity that Teulings and Toussaint (2023) rejected? For that, we perform individual tests for Paretianity using the $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ statistics with different set-ups: with and without rounding, with measurement errors from the kernel density of the full error sample, a 10% trimmed error sample, and a normalised error sample, centred to have mean zero. We then perform tests for individual sub-region/year pairs, and plot the empirical CDF of the two-sided p-values. If the top wealth samples stemmed from Pareto distributions, the CDFs would resemble the CDF of a uniform distribution, plotted in dotted-red. The results are in Figure 4.

As one can see, all CDFs exhibit a bulge above the diagonal, indicating that the data are not Paretian. One does notice that for small significance levels, the null hypothesis of Paretianity is rejected less often for the $\hat{\mathcal{R}}_3$ statistic than for the $\hat{\mathcal{R}}_2$ statistic. Further, one can notice that for the $\hat{\mathcal{R}}_2$ statistics, the null hypothesis can be rejected even more often when measurement error is accounted for, over 50% of times even at the 1% significance level.

Summarising, one can see that whether measurement error is accounted for or not, Paretianity is clearly rejected by the $\hat{\mathcal{R}}_k$ statistics, thus providing clear motivation for alternative models.



Note: 'Full Density' means measurement errors are drawn from the kernel density of all observations. 'Trimmed density' means measurement errors are drawn from the kernel density of observations that are not in the top and bottom decile. 'Normalised' means measurement errors are drawn from the full kernel density shifted to satisfy a mean of zero.

Figure 4: Empirical Cumulative Density Functions of all P-values from sub-region/year pair tests for Paretianity using different measurement error distributions, using $\hat{\mathcal{R}}_k$ statistics

4.2 Frequentist model evaluation

Let us now make a first comparison of a Pareto model against alternative models. Similarly to Teulings and Toussaint (2023), we begin by estimating, for each sub-region/year combination of more than 64 billionaires, the shape and scale parameters of the Pareto, Weibull and Generalised Pareto model using maximum likelihood estimation. For the Generalised Pareto model, we also estimate γ using Hill's (1975) estimator[6]. Subsequently, we fix the shape parameter for the Weibull and Generalised Pareto model and solely estimate the scale. The corresponding histograms of the estimates can be found in Figures 5 and 6.

From Figures 5 and 6, we obtain mean and median estimates for the shape parameters of the Weibull and Generalised Pareto distributions. The mean and median Weibull γ estimates are both 0.29. For the Generalised Pareto distribution, we notice the Hill estimates are slightly higher than the ML estimates. This is likely caused by the imperfect selection of the index for the estimator, and the resulting bias in estimates. We hence opt to use the ML mean and median estimates, which are 0.60 and 0.57 respectively. For subsequent estimation of the scale, we fix the shape to the median.

From the histograms of the estimated parameters, we may also directly compute the hyperparameters for the informative distribution priors used in later Bayesian models. Simple maximum likelihood estimates suggest a $\Gamma(41, 44)$ distribution for the Pareto α, an InverseGamma$(12, 14)$ distribution for the Weibull γ, and a $\Gamma(37, 33)$ distribution for the GPD γ, shifted by $\frac{1}{2}$ as described in Dombry et al. (2023).

[6]To pick the index for the estimator, as visual inspection for over 200 estimates is impractical, we choose the middle of the region of range 20 with the smallest standard deviation in estimates (stable region).

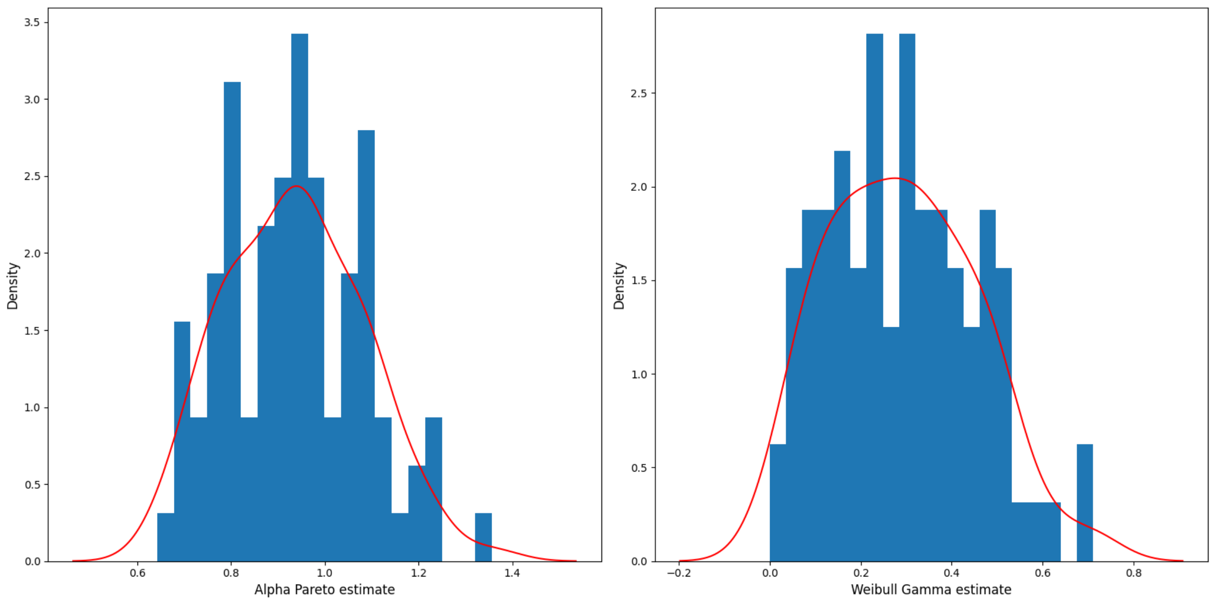


Figure 5: Density histograms of the Pareto α and Weibull γ estimates, with fitted kernel densities

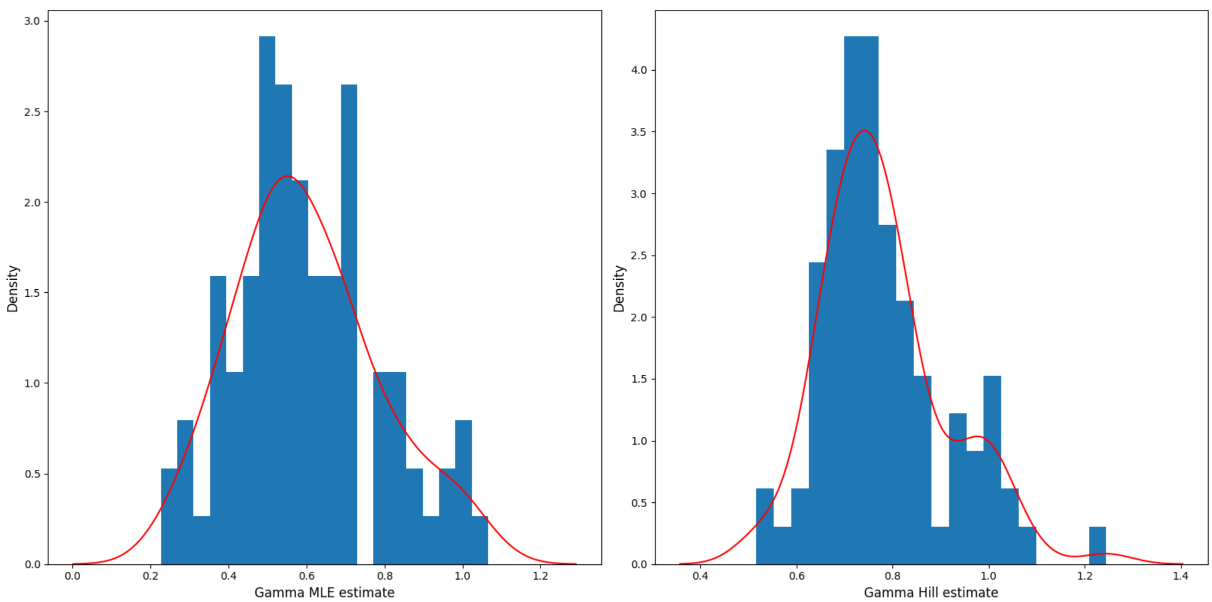


Figure 6: Density histograms of the Generalised Pareto MLE and Hill γ estimates

We now estimate the scale parameters for all observation with more than 20 billionaires using the fixed median shape parameters. Detailed summary statistics can be found in Appendix F, Table 9.

For each sub-region/year pair, we test the empirical distribution against our three distributions, and plot the cumulative P-value distribution. The GPD's overall better fit can best be seen in the cumulative P-value plot of the KS-tests shown in Figure 7, where the Generalised Pareto distribution has the mildest bulge above the main diagonal. The Pareto distribution offers the worst fit.

An important argument must be made in favour of the Pareto model that has been standard in literature for long. Whether top wealth is Pareto shaped is inherently dependent on what one

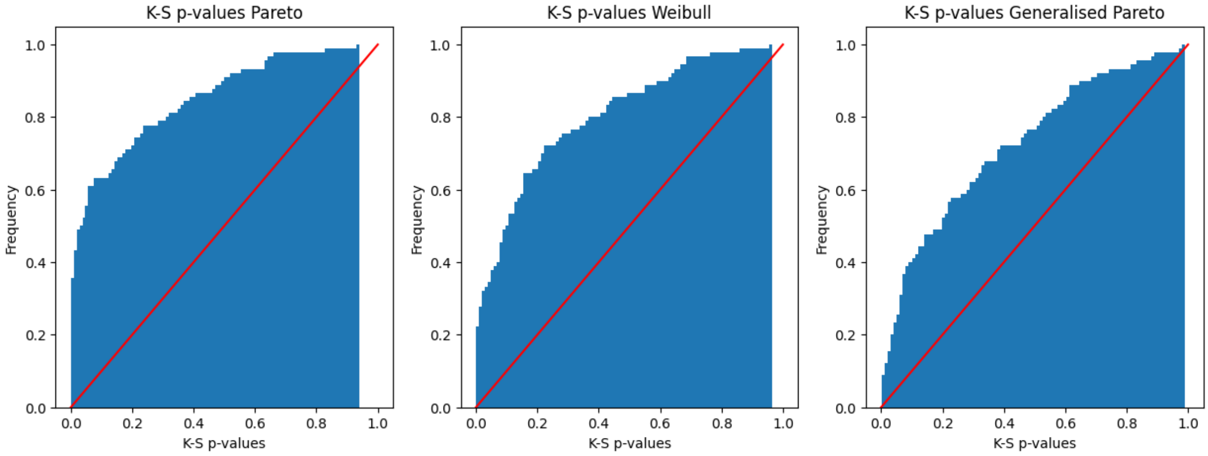


Figure 7: Empirical CDFs of KS-test P-values for different distributions estimated fitted by maximum likelihood

defines to be 'top'. In 2002, less than one in a million Americans was a billionaire. By 2022, the share quadrupled. Being a billionaire has become less special. Essentially, the threshold for being 'top' has shifted. This can be visualised well with plots of the empirical cumulative hazard rate of the log transformed wealth data, computed from the Kaplan-Meier estimator (Kaplan & Meier, 1958), alongside the fitted cumulative hazard rates of the fitted log-transform distributions, as shown in Figure 8.

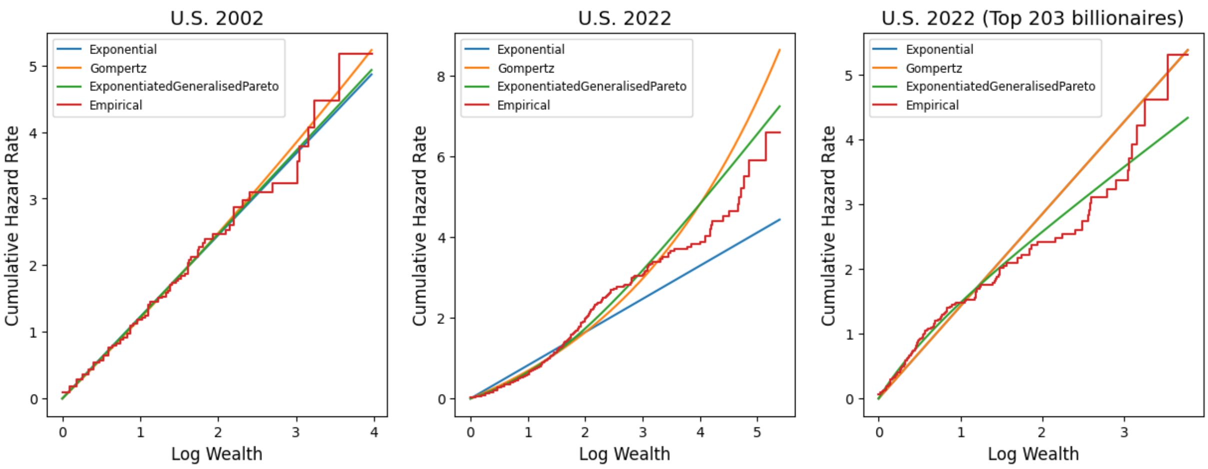


Figure 8: Cumulative Hazard for the U.S. across time and sub-samples using the Kaplan-Meier estimator and fitted frequentist models

We clearly see that in 2002, the Pareto distribution (log-transform is exponential) was an outstanding fit to billionaire net worth, whereas it is not in 2022. However, if we only take the top 203 observations in 2022 that correspond to a 0.6 in a million share of the population, we again obtain a good fit with a straight line (Exponential overlaps with Gompertz in the figure). Simply said, 'top' wealth as it was understood in 2002 may still be Pareto today, but billionaire wealth is not.

4.3 Bayesian cross sectional model evaluation

We now move to the evaluation of Bayesian cross-sectional models. Whenever we refer to a 'regularised' model, we mean a model with a common shape parameter for a sub-region across time, as explained in Section 3.2.1. Non-regularised models are such where the shape parameter is allowed to differ across years.

For each year/sub-group combination, we now perform a two sample KS-test of the true data against a large sample from the estimated posterior wealth distribution. Again, we can plot the KS-statistics as well as the cumulative densities of the P-values to get an impression of overall fit. The results are in Figure 9. We observe that the empirical P-value distribution of the GPD is almost uniform, thus indicating a very good fit. Further, the results for the Pareto distribution have barely improved compared to the frequentist alternative.

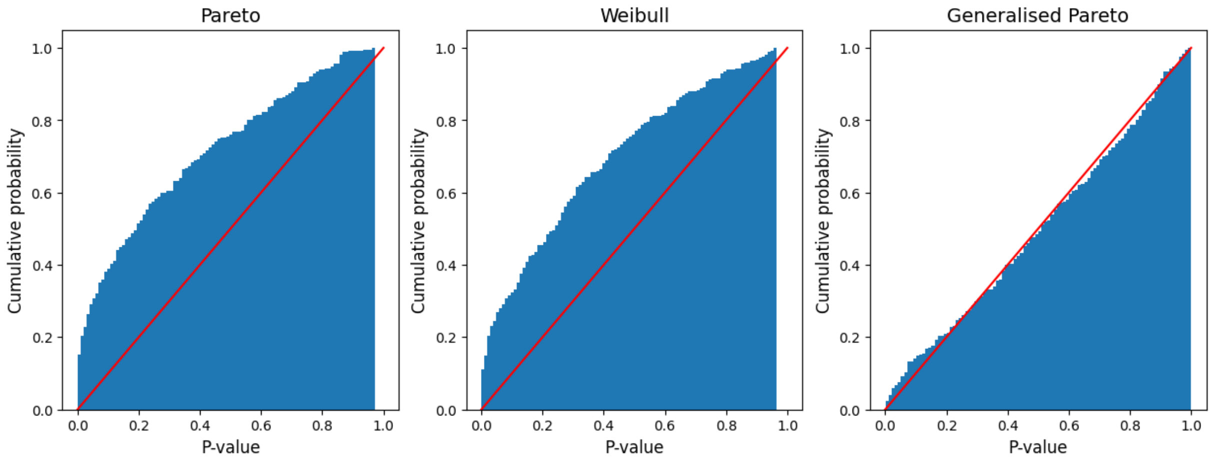


Figure 9: Cumulative P-values for non-regularised cross-sectional Bayes models

One thing to notice is that the Weibull shape estimates of the non-regularised models are quite stable over time, as can be seen in Figure 10. This suggests that allowing for flexibility in the shape estimates mainly adds noise, especially for the Weibull distribution, thus further motivating the regularised cross-sectional approach.

We now turn to the evaluation of the regularised Bayesian models to see if a more parsimonious approach improves the results. Note that since there is no parameter to regularise across time, the results for the regularised Pareto distribution are identical to its non-regularised counterpart (in fact, the models are identical).

In general, the results are much better than for the non-regularised models. Notably, the fit of the regularised Weibull model is much better than that of its non-regularised counterpart. Similarly to the non-regularised models, we present the results in Figure 11.

We can further briefly evaluate in-sample mean wealth predictions. 'In-sample' forecast in this context means that we use the posterior predictive distribution of a given year and simply take its mean. Teulings and Toussaint (2023) have already shown that Pareto performs terribly in this metric. The replication of this result can be found in Appendix G.2, Table 10. In Table 2, we can observe summary statistics of the absolute forecast errors of Weibull and the GPD, aggregating all sub-regions and years. The GPD performs well, much better than Pareto, but still worse than Weibull. However, the Weibull model's performance worsens under regularisation,

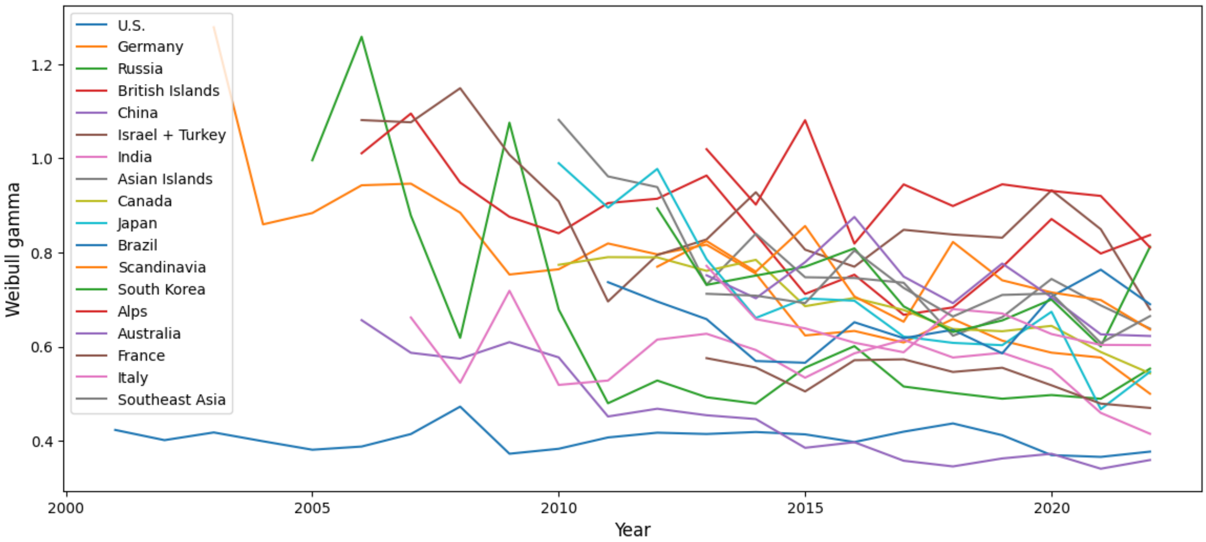


Figure 10: Non-regularised Weibull γ estimates over time for different sub-regions

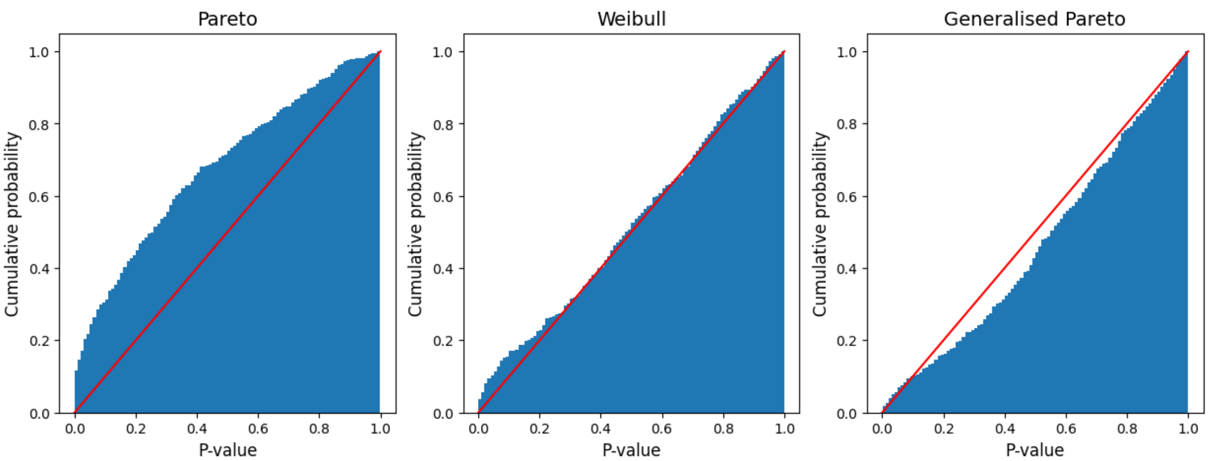


Figure 11: Cumulative P-values for regularised cross-sectional Bayes models

while that of the GPD model improves. Worth noting are also the outliers. They are not only much larger for the GPD models, but also occur mostly in France and Russia, the countries with the highest tail indices. This may be expected, because as the GPD shape parameter tends to 1, and exceeds it some samples in the posterior, the first moment of the GPD cedes to exist, or tends to infinity in sampling practice. Across time, Figure 12 shows that in-sample forecast errors are temporarily much larger in 2008 than in other years, which can be expected as the wealth distribution is perturbed by the financial crisis.

Finally, let us look at the estimated shape coefficients. This is necessary for two reasons. First, we can understand if the informative priors we imposed on the shape coefficients were still flexible enough to allow for differences across groups. If the distributions overlapped too much, it would be an indication that the priors were too strict. Second, we can get an impression of the coverage of the distributions, and thus of the confidence of the coefficient estimates, which is not directly possible for frequentist models. The densities of the GPD γ estimates are shown in Figure 13. The interested reader can find the Weibull estimates in Appendix H, Table 18.

Table 2: Summary statistics of the mean absolute errors of in-sample mean wealth predictions from different models

	Unregularised		Regularised	
	GPD	Weibull	GPD	Weibull
count	250	250	406	406
mean	1.68	0.13	1.08	0.37
std	3.37	0.23	2.04	0.54
min	0.04	0.00	0.00	0.00
25%	0.69	0.02	0.21	0.07
median	0.96	0.06	0.47	0.16
75%	1.43	0.16	1.11	0.41
max	30.10	2.16	25.06	3.45

Note: Un-regularised model predictions are limited to sub-region/year pairs with at least 20 observations.

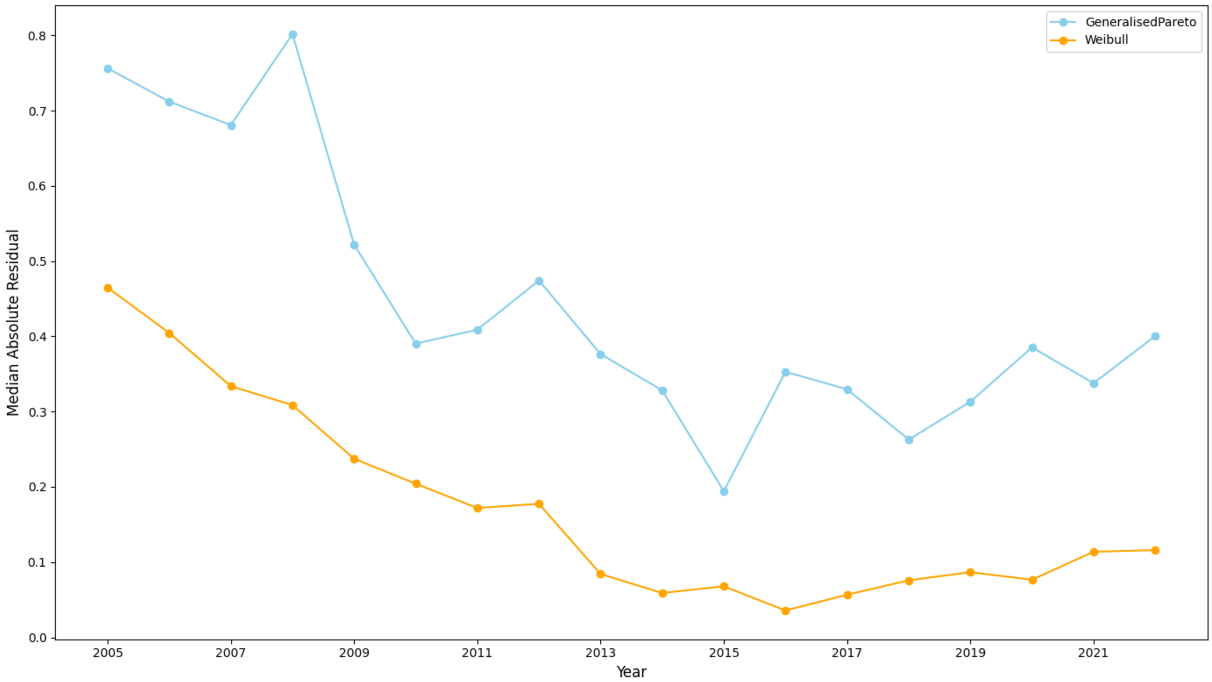


Figure 12: Mean absolute errors for in-sample wealth forecasts of regularised models over time

There is significant variation and limited overlap between the posterior distributions. As expected, we also notice differences in the coverage of the posteriors. For countries with more observations, such as the United States, we notice that the distribution becomes much tighter. This is expected: as one collects more data, one should gain confidence and give less importance to the prior. It is further important that the posteriors are well restricted between 0 and 1, as for values smaller than 0, the tail of the GPD becomes bounded, and for values above 1, the distribution has no moments. Neither of the two is the case.

We can also test the GPD posterior distributions for the property of asymptotic normality established by Dombry et al. (2023). Indeed, for the country with the largest sample, the United States, a Jarque-Bera test cannot reject the null hypothesis of normality even at a 10% level (JB-statistic of 1.30, P-value of 0.52). The null hypothesis of normality is rejected for all other

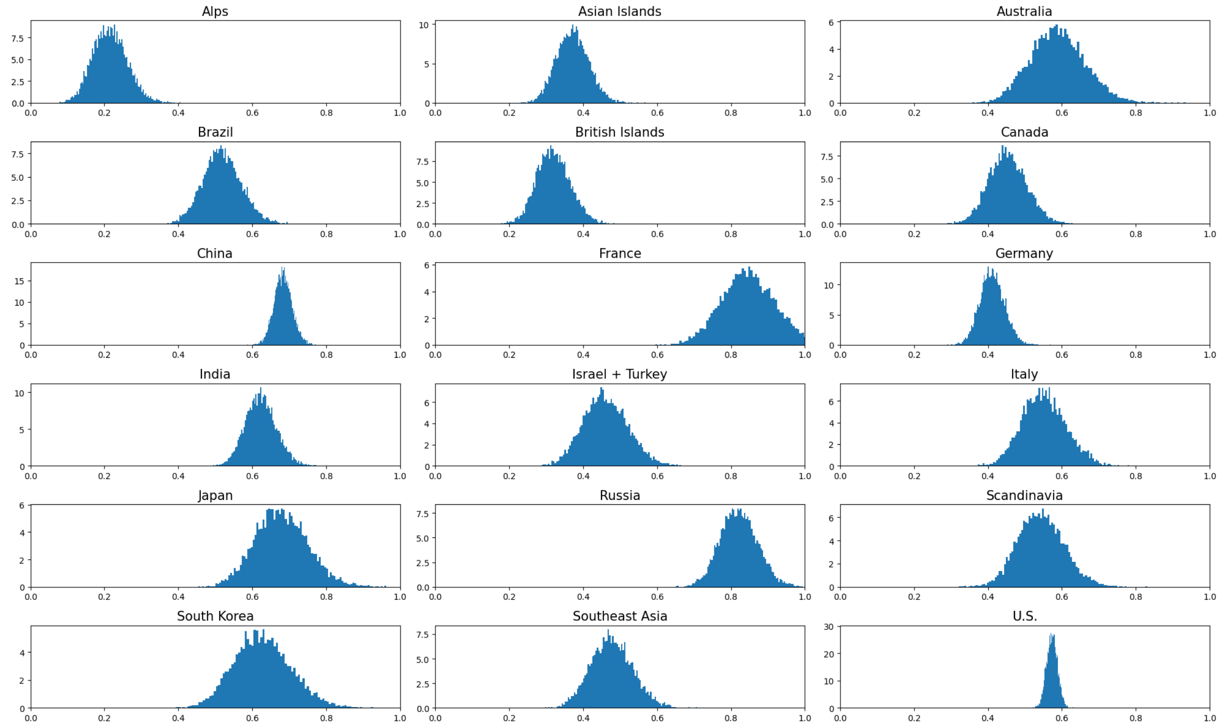


Figure 13: GPD γ posterior distributions across sub-regions

countries at a 1% significance level, indicating that in other cases the sample size is too small for the posterior to sufficiently converge to the normal distribution.

Summarising, regularisation imposed in the Bayesian models seems to clearly improve their fit to the data, on top of avoiding risks of overfitting. The results favour the Weibull and Generalised Pareto distribution, and clearly disapprove of the Pareto models.

4.4 Bayesian time series model evaluation

Last but not least, we evaluate the Bayesian time series models, as specified in Section 3.2.3. Let us first get an impression of the overall in and out of sample fit resulting from the further restriction of the Bayesian models. That is illustrated in the usual cumulative P-value plots of Figure 14.

Three observations can be made from the plots. First, as in all cases before, we notice that the Pareto distribution provides a poor fit to the data, both in and especially out of sample. Second, we see that the in-sample fit of the Weibull and GPD models is compromised by the regularisation via covariates, but not drastically. Of course, a certain loss in fit is expected, as we cannot claim that a trend component, and the growth in GDP per capita and the stock market can explain the entire variation in scale parameters. However, the fact that the overall fit is still this good indicates that the chosen covariates explain a significant part of the variation. Third, the out-of-sample fit is worse than the in-sample fit, as should be expected.

Next, we consider whether the covariate models have out of sample predictive power for mean wealth forecasts. For this, we compare the means of the forecasted posterior distributions, both for Weibull and GPD, to the naive 'random walk' forecast, equal to the mean wealth of the previous year. For the sake of realism, we use only the covariate for GDP growth, as stock

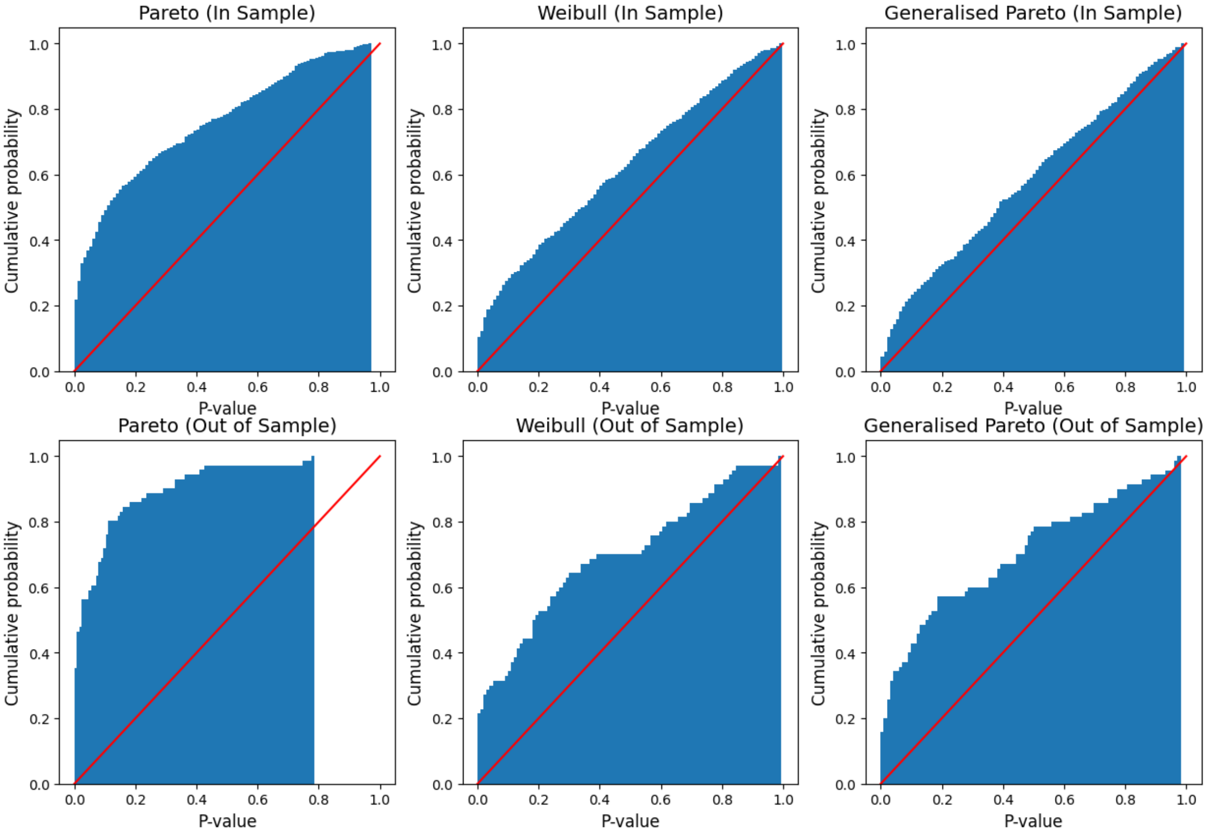


Figure 14: In and out-of-sample (one step ahead) KS-test P-values for Bayesian time series models

market returns are not predictable. Comparative statistics of the predictions for Weibull and GPD are available in Table 3. The exact estimation and prediction windows can be found in Appendix E.

Table 3: Summary statistics of the absolute residuals for the Weibull and Generalised Pareto time series model mean wealth predictions, one year ahead

	Weibull Model		Generalised Pareto Model	
	Model Mean Absolute Residual	Random Walk Mean Absolute Residual	Model Mean Absolute Residual	Random Walk Mean Absolute Residual
count	35	35	34	34
mean	0.43	0.46	0.71	0.45
std	0.40	0.36	0.90	0.36
min	0.01	0.01	0.00	0.01
25%	0.12	0.23	0.22	0.22
median	0.31	0.38	0.34	0.37
75%	0.56	0.57	0.76	0.53
max	1.54	1.28	3.64	1.28

In brief, we see that the mean wealth predictions of the Generalised Pareto model with covariates cannot outperform simple random walk forecasts. However, the Weibull model does outperform the random walk, decreasing the mean forecast error by about 7% compared to the random walk. By the standard deviation, we do notice that the variance in the Weibull residuals is greater than in the random walk residuals. For practitioners, this shows that while the most accurate predictions are obtained from the Weibull model, random walk estimates are more stable.

We can further interpret the model coefficients. We omit presenting results on the constants, as their posteriors are tight around zero for all models. Further, we omit presenting results for the Pareto and Weibull models' slope coefficients, as they are similar to the results for the GPD model. The interested reader can find plots of all posterior distributions in Appendix I. Figures 15 and 16 show the slope coefficient posterior distributions for the British Islands, Germany and the United States.

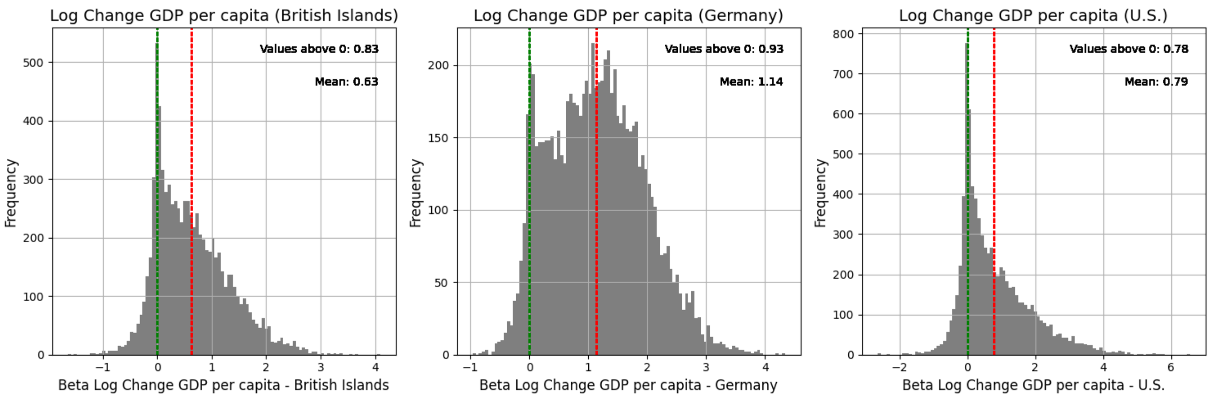


Figure 15: Slope coefficients for GDP per capita log growth

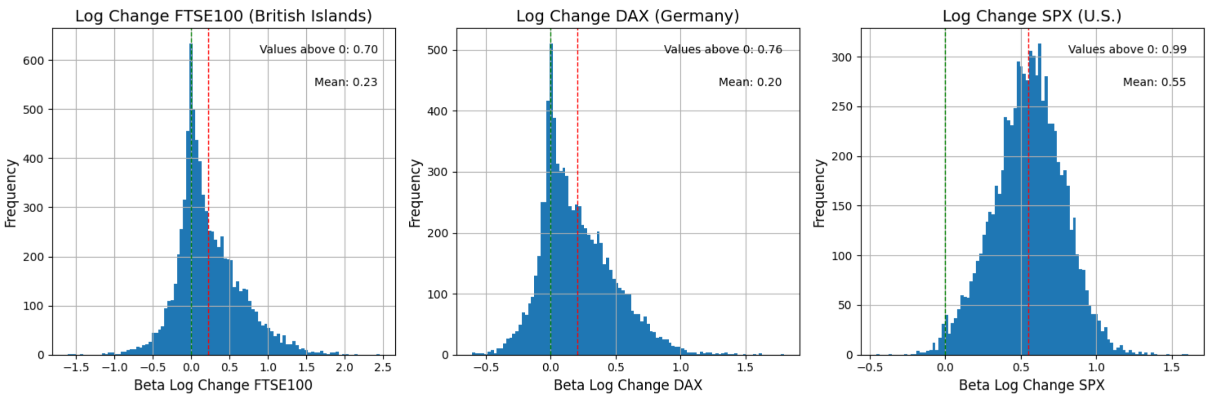


Figure 16: Slope coefficients for stock market log returns

The mean of all posteriors, both for the change in GDP and for stock market returns, is positive across all three countries. Although the effects are most pronounced for the three chosen countries, their choice is hardly cherry picking. The means of the slope posteriors are almost all positive, with the exception of only a few countries. In fact, the stock market coefficient is only negative for one country, India. Again, details can be found in Appendix I.

Further, we evaluate how important the chosen covariates are using the Bayesian R-squared. We focus on the best fitting model that utilises the Generalised Pareto Distribution. When including only a constant and the change in GDP, we find those two variables explain, on average across countries, 56% of the variance in the log scale changes, with the lowest median R^2 being 50% for China, and the highest being 71% for the British Islands. Adding the stock market index returns, we obtain an explained variance of log scale changes of 63% across countries, with the lowest median being 51% for Brazil, and again the highest median being in the British Islands with 81%, closely followed by the United States. In brief, a majority share of the variance in changes in log scale of the top wealth distribution can be explained by just a constant, the growth in GDP, and the local stock market. Detailed summary statistics can be found in Appendix J, Tables 14 and 15.

5 Implications

This section explores some of the practical economic interpretations and implications of the results described above, that go beyond the purely statistical view.

To begin, we comment on the estimated elasticity figures of billionaire wealth to GDP growth and the stock market. The latter also allows us to comment on the long term outlook on billionaire wealth, and ultimately judge whether Oxfam's prediction of the existence of the first trillionaire by 2034 (Riddell et al., 2024) is realistic or not.

We then show the consequences of our findings on the best fitting models for the top wealth distribution, which directly relate to the effectiveness of taxation policies. Finally, we move to a closer analysis of the estimated model parameters, with a particular focus on the tail index estimates from the Generalised Pareto models. The latter translate directly into the fatness of the wealth tail, telling us how likely it is to observe extreme wealth in a country. We can relate this to individual countries' political and economical environments.

5.1 Consequences from covariate results

We now move to a closer interpretation of the estimated elasticities of the billionaire wealth scale parameters to GDP per capita and the corresponding regional stock markets. As the slope coefficients do not differ greatly between the related Weibull or Generalised Pareto models, we focus on the latter.

We begin with regional stock market indices. Here, we observed in the results that, besides India's scale elasticity to the NIFTY, which is only slightly negative, all other slope coefficients are positive in mean and median. This is what one would expect: billionaires are almost always shareholders of large corporations, which are consequently constituents of the important stock market indices. In France, for example, Oxfam estimates that the ten richest French families control 29 per cent of the national stock market (Jacobs, 2015). Hence, as stock indices tick up, so does billionaire wealth, and thus the scale coefficient increases.

Although the slopes are clearly positive, their magnitude is surprising. Under the assumption that all billionaire wealth exists in the form of companies that are also listed on the exchanges, one would expect unit elasticity: As the stock market ticks up, billionaire wealth would tick up proportionately. Such is a result that Gandhi and Walton (2012) find with exploratory analysis of aggregate Indian billionaire wealth. Further, Freund and Oliver (2016) find that only about one in eight billionaire's wealth can be classified as 'diversified' across assets. However, our results show a very low elasticity of the scale to the stock market, with only the US and corresponding SPX exceeding a median coefficient of 0.5. Thus, the distribution of top wealth is inelastic to returns of the stock market. Two explanations may be considered for this finding. For one, billionaires may be diversified geographically, or not even do business in their country of origin - little prevents a French citizen from founding a business in the United States, and thus not owning stock in the french CAC40. However, as indices correlate strongly, this cannot be the full explanation. One may thus take this finding as an indication that top wealth is very well hedged. Especially people with inherited fortunes, often less connected to the source of the family's wealth, may naturally consider growing, but also protecting their wealth, via appropriate financial instruments.

Moving onto the GDP coefficients, again one finds they are positive across all countries. Again, this is expected, similarly to as we discussed before for the stock index covariate, and in line with the findings from Gandhi and Walton (2012).

There is a feature in the posteriors that is specific to only a few countries. The posteriors of the GDP and stock market coefficients of the three countries displayed in Figures 15 and 16 have not only a positive mean, but a very clear positive skew. They are the only countries exhibiting this behaviour. Hence, for the UK, Germany and the US, while it seems clear that economic growth and positive stock market returns have a positive effect on the scale of the billionaire distribution, it is unclear how strong this effect really is. Why one observes this posterior calls for speculation, but the most likely explanation lies in the asymmetry of reactions. As already shown by Gandhi and Walton (2012), aggregate billionaire wealth can react disproportionately in recessions, such as for example in 2008. In our case, our data contain two recessions, the financial crisis and Covid-19, and it is likely that these two very strong observations of a shift in scale skew the posterior distribution, inducing the uncertainty. From a practical perspective, this has an important implication for policy makers considering a wealth tax. Most forms of taxation are proportional to economic output, that is, GDP. Think, for example, of VAT or income tax, which are levied in proportion to all spent or earned. If GDP drops by roughly 5%, one expects VAT revenue to drop by roughly 5% too. This seems to be very different for a wealth tax, if wealth reacts disproportionately to GDP. Therefore, governments should recall that even though a given wealth tax may help balancing state budgets in regular times, its revenues will fall short in times of recessions, and thus force a government to run even larger short term deficits.

Finally, a brief comment on the constant coefficients. They are all very close to and centred around zero. In other words, there is no evidence of 'natural' growth of the scale of top wealth simply as time progresses, without economic or capital markets growth causing it. Popov (2018) and Prinz and Bollacke (2018) show that cross-sectional differences in the number and wealth of billionaires can mainly be explained by GDP per capita, but our research is the first to demonstrate that this relationship also holds true across time within the same country. Further, we also show using residuals that the implemented covariates explain over 60% of the variance in the scale coefficient changes. It is thus unlikely that the remaining variance can be explained by further factors such as GDP and capital markets that grow over time, but rather by factors unrelated to time, such as for example changes in taxation.

Last but not least, we can attempt to answer the question teased in the introduction: will there be a trillionaire by 2034? To answer this question, we make some simplifying assumptions. We restrict ourselves to billionaires in the United States. We assume the US economy grows about 3.4% per year, and that the S&P500 delivers a 5.7% per cent return every year after 2022.

Further, we assume the number of billionaires in the U.S. grows by a number of just over 26 per year. These numbers are historical averages over the range of our dataset. This leads to an expected 1024 billionaires in the U.S. in 2034.

We now estimate a time series model from 2008 up to and including 2022 in the U.S., and make a twelve step ahead forecast of the wealth distribution in 2034 using the estimated posteriors. In a simulation exercise, we draw many times a total of 1024 observations from the predicted posterior, and report whether a trillionaire is present. Over ten thousand simulations, we find that at least one trillionaire exists in about 19% of the simulations. Considering that about half of the world's billionaires currently live in the U.S., one may say that the existence of a trillionaire by 2034 is by no means guaranteed, but very much realistic. The figure is also a solid reality check for the model, as it is roughly in line with the expert judgements from the Oxfam study.

One should note that, unlike suggested by Riddell et al. (2024), the possibility of the existence of a trillionaire does not imply an increase in inequality, or a loss of economic prosperity. Quite the opposite, our research has shown that not only does billionaire wealth react in-elastically to growth, but it is indeed very much dependent on economic growth. The existence of a trillionaire in 2034 is very dependent on whether the world economy grows as quickly as it has over the past two decades.

5.2 Consequences from the shape

To understand why it matters whether the top wealth distribution is Pareto, Weibull, or Generalised Pareto, one has to understand which underlying economic processes result in the respective top wealth distributions.

Teulings and Toussaint (2023) have indeed already argued that, as the Gompertz (log of Weibull) distribution has an exponentially increasing hazard rate, it lends itself to interpretation in the sense of a capacity constraint in network models. Notably, as proven by Tishby et al. (2016), the distribution of self-avoiding random walks in Erdös–Rényi–Gilbert networks is Gompertz. If one views a network as an economy with nodes as individuals, then one could view the length of such a walk as a capacity constraint to entrepreneurs - a company can only sell its product to each individual a finite number of times. If the assumption of Weibull were to be found fitting, it would thus imply an exponentially increasing hazard rate on wealth. In other words, it would indicate that we are unlikely to observe much more extreme net worth in the future, possibly due to economic capacity constraints.

Our results, however, have shown that the Generalised Pareto distribution is overall the better fit for top wealth. In particular, we have shown that when considering only a subset of billionaires, or equivalently, the wealth distribution in the early 2000s when billionaires were rarer, the Pareto distribution does become a much better fit. As discussed, this is not a surprise. As outlined by Charpentier and Flachaire (2022), the GPD behaves like a Pareto distribution above a higher threshold. Hence, when selecting a smaller elite of billionaires, a Generalised Pareto distribution will resemble the conventional Pareto counterpart.

From an economic perspective, the Pareto distribution lends itself as the result of much simpler and realistic economic processes than the Weibull distribution. Notably, these include simple diffusion processes, as pointed out by Gabaix (2009). Of particular interest is a random difference equation model.

There are two ways for an individual to grow wealth. Either, one generates a return on one's own wealth (e.g. a growth in a stock portfolio), or one earns some exogenous form of income, like salary. Denote such a model with wealth W as $W_{t+1} = A_{t+1} \cdot W_t + B_{t+1}$, where A_{t+1} and B_{t+1} are the return on existing wealth and exogenous income, both random, respectively.

Two important results follow. First, under some loose regularity conditions, Kesten (1973) and later Goldie (1991) show that as one lets wealth go to infinity, the shape of the excess wealth distribution follows a power law (i.e., Pareto). Translated into the real economy, this is a strong theoretical foundation to expect the top wealth distribution to be Pareto, (virtually) irrespective of what the return distribution on wealth is, and more importantly, also irrespective of the income distribution. In other words, no matter what income redistribution policy a government adapts, the shape of the top wealth distribution should always resemble a Pareto distribution.

The second, even more practically relevant result from the work of Kesten (1973) and Goldie (1991), is that the shape parameter of the limiting Pareto distribution is independent, to a large extent, of the distribution of B_t, and dependent on the distribution of A_t. What does this mean in practice? Essentially, any perturbation to the income distribution, for example via a form of taxation, has no influence on the shape parameter of the top wealth distribution. So while higher income tax may be effective at reducing inequality in the more general population, it should be unsuccessful at reducing the frequency and magnitude of extreme net worth. This is consistent with empirical research, as for example Berman et al. (2016) have shown that changes in the income distribution have no significant impact on the wealth distribution. On the other hand, the theoretical results suggest that taxes like a wealth tax or a capital gains tax should indeed have an impact on the shape of the top wealth distribution. As a simple example, if one imposes a 1% annual wealth tax on everyone, the expected value of A_t decreases by 0.01. That is, any tax affecting the return on wealth will result in less extreme wealth observations within the subset of billionaires.

Summarising, as our results show a better fit of the GPD over Weibull, and we showed that the very top end of wealth does indeed resemble a Pareto distribution, this is an indication that wealth can be modelled by a simple diffusion process. In turn, this implies that while income tax policies should have no impact on extreme wealth, wealth or capital gains taxes should.

The results also have direct implications not only on the adequate form, but also on the optimal rate of taxation. Teulings and Toussaint (2023) have already noted that the optimal marginal income tax as derived by Diamond (1998) and Saez (2001) on the richest individual is constant under the assumption of Paretianity, but converges to zero when assuming Weibull due to the exponentially increasing hazard rate. In the same application, assuming a GPD again implies an constant non-zero optimal tax rate. That is because while the hazard rate of the exponentiated GPD (log of GPD) increases exponentially for small log-wealth, it quickly converges to a constant as log-wealth increases. Making cardinal estimates of the optimal tax rates requires strong assumptions on the income elasticity, and is thus not the focus of this research.

What about an optimal wealth tax? Assume one decides to impose tax τ on the net worth of billionaires that gets levied once a year, on the full estate. In this simple set-up, Saez and

Zucman (2019) derive that the optimal (i.e., revenue maximising) wealth tax rate is given by

$$\tau^R = \frac{1}{1 + e_T},$$

where e_T is the average number of years that a billionaire is exposed to the tax rate, weighted by his wealth, in the T-th year that the tax is implemented. Hence, the tax rate may change over time, and converge to e_∞. In the case of a thin-tailed distribution, like Weibull, the set of billionaires will be relatively elastic, as 'small' fortunes are eradicated relatively quickly, resulting in a low e_∞. In contrast, with a heavy tailed distribution, higher weights are placed on longer lasting billionaires, resulting in a higher average e_∞. This leads to a counter-intuitive result: as we found that the billionaire wealth distribution is best described by a heavy-tailed generalised Pareto distribution, the revenue maximising wealth tax is higher than that for a thin tailed distribution. One may note that in practice, one can also compute an instantaneous estimate of e_∞ based on historical billionaire data. Indeed, Saez and Zucman (2019) have shown that for the United States in the year 2018, the revenue maximising wealth tax would have been around 6%. Of course, one may consider more complicated, possibly marginal wealth taxes.

Let us now move to a more in-depth interpretation of the estimated tail indices. We have already presented the results on the parameter estimates for the Generalised Pareto Distribution in Figure 13. A couple of straightforward observations can be made. First, the large majority of countries have similar tail index values of between 0.4 and 0.6, suggesting strong similarity in the likelihood of extreme net worth across the world. Second, one can observe that the width of the posterior distributions - a metric for the level of confidence in the parameter - is much narrower for the countries with more billionaires, like the United States, China and Germany. This is expected, as with more observations the estimate of a parameter should have less variance.

Of more economic interest are the top and bottom observations in the tail indices. On the very high end, France and Russia have a tail index posterior distribution with a median around 0.85. This is significantly higher than the tail indices in other countries. Economically, they indicate that in France and Russia, observing extreme net worth among the class of billionaires is more likely than in other countries. In other words, inequality among billionaires is most pronounced in the two nations. Note this does not imply anything on the actual magnitude of the observations, as that is determined by the scale parameter.

What can explain these estimates? For France, this result should come as a particularly important surprise, as France is one of just five OECD countries implementing a wealth tax on individuals as of 2021[7]. If anything, one should therefore expect the tail index to be particularly low. A hint for an explanation for the high value is given by the width of the posterior - it is indeed the broadest across all countries, despite a considerable number of observations. The cause can be found in the outliers: Bernard Arnault and Françoise Bettencourt, respectively main shareholders of LVMH and L'Oréal. The gap between them and the remainder of French billionaires is unusually large: Bernard Arnault has over twice the wealth of Françoise Bettencourt, who in turn has around three times the wealth of the next richest French billionaire. In this sense, it is reasonable to conclude that the high tail index in France is a result of two outliers, rather than structural factors in the French economy. Indeed, dropping the top two outliers from the sample

[7] https://stats.oecd.org/Index.aspx?DataSetCode=RS_GBL

results in a tail index around 0.6, similar to the rest of the world.

The conclusion is different for Russia. Many publications highlight the evolution of the Russian elite after the falling apart of the Soviet Union. In particular, Treisman (2007) and Djankov (2015) emphasise how under Russian President Vladimir Putin, the former oligarchs have lost property, whilst assets were redistributed to people close to the President, often with a background in the Russian or formerly Soviet secret services. While Treisman (2016) also highlights that the modern Russian billionaires are more integrated into the market economy than they used to be two decades ago, he acknowledges a substantial proportion of wealth is still a relic of political forces. Therefore, one can easily conclude that the exceptional shape of Russia's top wealth distribution is not a result of usual market forces, but rather a consequence of manual interference by political leadership.

On the opposite side, the by far lowest tail index can be observed for the Alps, which for top wealth essentially corresponds to a population of Swiss billionaires. At first sight, the especially low tail index may defy the stereotype of the ultra-rich Swiss, but recall that this only concerns the inequality amongst billionaires. In other words, extreme net worths are unlikely to be observed in Switzerland. This cannot be explained by supposed tax evasion from foreign nationals migrating to Switzerland, as this paper's geographical classifications are based on citizenship, not residency. The literature proposes two explanations for unexpectedly high 'equality' in the tail of the Swiss wealth distribution. For one, Baselgia and Martínez (2024) show that an unusually high proportion of over 60% of Swiss billionaires are heirs, a ratio twice as high as for example in the United States. However, it is unclear whether this is a cause or a result of the low tail index: maybe concentration of generational wealth limits its expansion, but an economy with a low propensity to extreme wealth also results in a naturally higher share of heirs. A better explanation of the flat wealth tail may be Switzerland's wealth tax, which although avoidable for foreign nationals, is more difficult to circumvent for Swiss citizens. This is supported by empirical studies. For instance, Brülhart et al. (2022) estimate that indeed, wealth in Switzerland is very elastic to changes in the wealth tax rate, as can be investigated using inter-cantonal differences in the tax policy. While this is no conclusive proof that a Swiss-model wealth tax succeeds in reducing extreme wealth, it is a significant indication thereof.

6 Conclusion

To conclude, we summarise the essence of our findings. First and foremost, all parts of the results show that the Pareto distribution is a poor fit for billionaire wealth data, across all countries. These results are independent of the structure of the underlying tested model, whether Bayesian or frequentist. Paretianity is extensively tested for using both the $\mathcal{R}_k$ statistics from Teulings and Toussaint (2023), as well as the conventional Kolmogorov-Smirnov test. However, as highlighted before, this result is very sensitive to the threshold. When increasing the threshold substantially beyond the one billion dollar mark, the empirical hazard rate of the log-transform again appears linear, and the Pareto distribution provides good fit.

Furthermore, we provide a detailed investigation of measurement errors in the context of top wealth. We compare independent measurements from Bloomberg and Forbes and find substantial differences in wealth estimations for the same individuals. By sampling errors from kernel density estimates of the measurement error distribution, we see that such measurement errors invalidate essential properties of the $\mathcal{R}_k$, such as a mean at 1. Nevertheless, conclusions rejecting Paretianity withstand even the most substantial contamination by measurement error.

Moving on, we introduced a model using the Generalised Pareto Distribution (GPD) for excess wealth beyond a billion $US. Whilst this is a standard approach in the more general context of Extreme Value Theory, it has hardly been applied to the wealth distribution. We find that even with simple frequentist maximum likelihood methods, the GPD is a better fit to the data than both the Pareto distribution and the Weibull distribution.

Next, we examined a set of Bayesian models with informative priors on the shape of the studied distributions. It is clear that when regularisation across time is imposed on the shape parameters of the Weibull and Generalised Pareto distributions, both models fit the data remarkably well. In general, whether in a frequentist or Bayesian context, the Weibull distribution provides by far the best in and out of sample predictions for mean billionaire wealth.

Finally, we constructed Bayesian time series models in which the scale parameters follow a log-autoregressive process with macroeconomic covariates. We find that both GDP per capita growth and stock index returns have significant positive impact on the scale of the wealth distribution, with the slope coefficients having high upside entropy. Hence, changes in the wealth distribution of billionaires can be explained in large parts by changes in GDP per capita and returns of the regional stock market. One year ahead predictive distribution forecasts using these models provide good out of sample fit to future wealth data, and can be used to forecast mean wealth better than a simple random walk model. Answering the question from the opener, we conclude that having a trillionaire by 2034 is by no means guaranteed, but still realistic.

It is worth mentioning the limitations of the analysis. A clear methodological issue of this research is correlation across countries. Our own analysis shows this: the key factors driving changes in the wealth distribution are similar in all sub-regions of the analysis. If changes in the distribution across countries are correlated, so will be the outcomes of any statistical tests. Further regularisation could solve this. For example, one may consider imposing cross-country correlation restrictions on coefficients, with priors based, for instance, on empirical correlations, geographical proximity, or the economic closeness between countries as measured by trade flows. One could for instance introduce a model where the shape parameters of countries are not independently drawn from a common hyper-prior, but instead from a multivariate distribution with a covariance structure.

Further, whilst the forecasting component of the research is not its centrepiece, it is still valid to criticise it on several levels. First, as we mentioned ourselves, the covariates employed are not known a priori in a forecasting environment. GDP may be forecasted well, but only in standard economic conditions. Financial crisis and Covid-19 are not predictable. Further, the data at hand forced us to focus out-of-sample forecasting on the years 2019 to 2022 only, and so it is difficult to generalise the findings from the exercise. Performing the same covariate analysis but with variables known in advance, possibly lags of the ones in this paper, used on a broader time horizon with Forbes data prior to 1997 would be desirable. One should further consider models allowing for asymmetric reactions to positive and negative growth.

The findings of this research have direct implications for policy. In particular, they suggest that only taxes on capital stock, or simply said, wealth directly, can have an impact on the shape of the top wealth distribution. We have further outlined the importance that the fatness of the tail of the distribution has on the revenue maximising wealth tax rate. There are also very important differences across countries, particularly in the tail indices. Crucially, we point out the low tail index in Switzerland, and proposed it may be due to Switzerland being the most prominent example of a developed country with an important wealth tax.

Some of these results definitely merit follow-up. In particular, the large discrepancy in the slope estimates for GDP growth and stock returns across countries merits a separate study in its own. Possible explanations include, for example, discrepancies in taxation schemes across countries. If wealth, or additions to it, are taxed at a higher rate, one would expect the elasticity of a billionaire's wealth to a change in macroeconomic conditions, or to a change in the valuation of his business, to be attenuated heavily. That is, the elasticity of the scale of the wealth distribution may be strongly driven by tax rates. This would require to take a more thorough approach to the Forbes data, as one should then determine where a billionaire's wealth generating activities are located, what exact tax scheme they are subject to, and make adjustments for changes in exchange rates.

Finally, a natural follow-up to forecasting entire distributions would be to instead forecast the wealth distribution by modelling the individual. A problem with forecasting distributions in our context is that the observation threshold is fixed in nature: we only look at billionaires, but not at anyone with a net worth of $999 million. This is an issue for policy makers too. Whilst this research can be used to forecast the mean wealth of billionaires, it cannot forecast the index, that is, how many billionaires there will be. This number, however, is just as essential when forecasting tax revenue. Moreover, over a very long time horizon, as a larger share of the population becomes billionaires, one should expect the fit of the Generalised Pareto distribution to deteriorate.

In the view of the lack of public data, it would thus be reasonable to perform simulation experiments of a simple economy that reproduce the empirical top wealth distribution. This is not a new idea. In particular, Coelho et al. (2005) have constructed a family network model, in which total wealth is constant, agents only spend money on raising their children, and expenses are redistributed to other agents in a fashion favouring agents with higher wealth. In this simulation setting, the upper 5% of the artificial society indeed exhibits a wealth distribution with a Pareto shape, and the estimated Pareto index matches the measurements of real UK population data. Similar Paretian results were obtained by Bouchaud and Mézard (2000), who add a stochastic wealth growth component in a similar setting as Coelho et al. (2005). As the sizes of the networks in these studies are about four orders of magnitude too small to produce results that in the real world would include billionaires, one should consider similar set-ups with more computational capacity.

References

Atkinson, A. B., & Harrison, A. J. (1979). The analysis of trends over time in the distribution of personal wealth in britain. In *Annales de l'insee* (pp. 89–108).

Bagchi, S., & Svejnar, J. (2015). Does wealth inequality matter for growth? the effect of billionaire wealth, income distribution, and poverty. *Journal of Comparative Economics*, *43*(3), 505–530.

Balkema, A. A., & De Haan, L. (1974). Residual life time at great age. *The Annals of probability*, *2*(5), 792–804.

Baselgia, E., & Martínez, I. Z. (2024). Using rich lists to study the super-rich and top wealth inequality: Insights from switzerland. *Review of Economics and Statistics*, 1–45.

Berman, Y., Ben-Jacob, E., & Shapira, Y. (2016). The dynamics of wealth inequality and the effect of income distribution. *PloS one*, *11*(4), e0154196.

Blanchet, T., Fournier, J., & Piketty, T. (2022). Generalized pareto curves: theory and applications. *Review of Income and Wealth*, *68*(1), 263–288.

Bouchaud, J.-P., & Mézard, M. (2000). Wealth condensation in a simple model of economy. *Physica A: Statistical Mechanics and its Applications*, *282*(3-4), 536–545.

Brülhart, M., Gruber, J., Krapf, M., & Schmidheiny, K. (2022). Behavioral responses to wealth taxes: Evidence from switzerland. *American Economic Journal: Economic Policy*, *14*(4), 111–150.

Capehart, K. W. (2014). Is the wealth of the world's billionaires not paretian? *Physica A: Statistical Mechanics and its Applications*, *395*, 255–260.

Charpentier, A., & Flachaire, E. (2022). Pareto models for top incomes and wealth. *The Journal of Economic Inequality*, *20*(1), 1–25.

Chu, J., Dickin, O., & Nadarajah, S. (2019). A review of goodness of fit tests for pareto distributions. *Journal of Computational and Applied Mathematics*, *361*, 13–41.

Coelho, R., Néda, Z., Ramasco, J. J., & Santos, M. A. (2005). A family-network model for wealth distribution in societies. *Physica A: Statistical Mechanics and its Applications*, *353*, 515–528.

Diamond, P. A. (1998). Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates. *American Economic Review*, 83–95.

Diebolt, J., El-Aroui, M.-A., Garrido, M., & Girard, S. (2005). Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling. *Extremes*, *8*, 57–78.

Djankov, S. (2015). *Russia's economy under putin: From crony capitalism to state capitalism* (Tech. Rep.). 1750 Massachusetts Avenue, NW Washington, DC 20036: Peterson Institute for International Economics Washington.

Dombry, C., Padoan, S. A., & Rizzelli, S. (2023). Asymptotic theory for bayesian inference and prediction: from the ordinary to a conditional peaks-over-threshold method. *arXiv preprint arXiv:2310.06720*.

Drăgulescu, A., & Yakovenko, V. M. (2001). Exponential and power-law probability distributions of wealth and income in the united kingdom and the united states. *Physica A: Statistical Mechanics and its Applications*, *299*(1-2), 213–221.

Flanigan, J., & Freiman, C. (2022). Wealth without limits: In defense of billionaires. *Ethical theory and moral practice*, *25*(5), 755–775.

Freund, C., & Oliver, S. (2016). The origins of the superrich: the billionaire characteristics database. *Peterson Institute for International Economics Working Paper*(16-1).

Gabaix, X. (2009). Power laws in economics and finance. *Annu. Rev. Econ.*, *1*(1), 255–294.

Gandhi, A., & Walton, M. (2012). Where do india's billionaires get their wealth? *Economic and Political Weekly*, 10–14.

Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for bayesian regression models. *The American Statistician*.

Goldie, C. M. (1991). Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 126–166.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 1163–1174.

Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, *29*(3), 339–349.

Jacobi, A., & Tzur, J. (2020). Quality of weibull, dagum and burr xii in estimating wealth distribution across countries and over time. *Dagum and Burr XII in Estimating Wealth Distribution Across Countries and Over Time (March 25, 2020)*.

Jacobs, D. (2015). Extreme wealth is not merited. *Op. cit*.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, *53*(282), 457–481.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, *131*, 208–248.

Klass, O. S., Biham, O., Levy, M., Malcai, O., & Solomon, S. (2006). The forbes 400 and the pareto wealth distribution. *Economics Letters*, *90*(2), 290–295.

Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *34*(1), 1–18.

Pareto, V. (1896). *Cours d'économie politique* (Vol. 1). Librairie Droz.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, 119–131.

Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

Popov, V. (2018). Why some countries have more billionaires than others? explaining variety in the billionaire intensity of gdp. *Investment Migration Research Paper 2018*, *3*.

Prinz, A., & Bollacke, S. (2018). Reassessing the relationship between billionaires, billionaires' wealth and per-capita gdp: A gini regression analysis. *Billionaires' Wealth and Per-Capita GDP: A Gini Regression Analysis (January 26, 2018)*.

Riddell, R., Ahmed, N., Maitland, A., Lawson, M., & Taneja, A. (2024). Inequality inc. how corporate power divides our world and the need for a new era of public action. *TODO*.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies*, *68*(1), 205–229.

Saez, E., & Zucman, G. (2019). Progressive wealth taxation. *Brookings Papers on Economic Activity*, *2019*(2), 437–533.

Stiglitz, J. E. (1969). Distribution of income and wealth among individuals. *Econometrica: Journal of the Econometric Society*, 382–397.

Teulings, C. N., & Toussaint, S. (2023). Top wealth is distributed weibull, not pareto. *TODO*.

Tishby, I., Biham, O., & Katzav, E. (2016). The distribution of path lengths of self avoiding walks on erdős–rényi networks. *Journal of Physics A: Mathematical and Theoretical*, *49*(28), 285002.

Treisman, D. (2007). Putin's silovarchs. *Orbis*, *51*(1), 141–153.

Treisman, D. (2016). Russia's billionaires. *American Economic Review*, *106*(5), 236–241.

Vermeulen, P. (2018). How fat is the top tail of the wealth distribution? *Review of Income and Wealth*, *64*(2), 357–387.

A Regions and sub-regions

Table 4: Countries by Region

Region	**Countries**
North America	United States, Canada
Europe	Germany, United Kingdom, Ireland, Cyprus, Czech Republic, Czechia, Denmark, Austria, Belgium, Spain, France, Greece, Italy, Netherlands, Norway, Poland, Portugal, Sweden, Switzerland, Liechtenstein, Lithuania, Monaco, Estonia, Finland, Slovakia, Romania, Hungary, Bulgaria, Guernsey, Iceland
China	China, Hong Kong, Macau, Macao
East Asia	Thailand, Malaysia, Singapore, Taiwan, Philippines, Indonesia, South Korea, Japan, Australia, Vietnam, New Zealand
India	India
Central Eurasia	Russia, Kazakhstan, Ukraine, Armenia, Georgia
South America	Brazil, Chile, Argentina, Peru, Venezuela, Colombia, Uruguay, Guatemala, Panama, Barbados, Belize, Mexico
Middle East	Turkey, Egypt, Israel, Saudi Arabia, United Arab Emirates, Kuwait, Qatar, Oman, Lebanon

Table 5: Countries by Sub-Region

Sub-Region	**Countries**
U.S.	United States
Canada	Canada
Germany	Germany
British Islands	United Kingdom, Ireland
Scandinavia	Denmark, Norway, Sweden, Finland
France	France, Monaco
Alps	Switzerland, Liechtenstein, Austria
Italy	Italy
China	China, Hong Kong
Southeast Asia	Thailand, Malaysia, Singapore
Asian Islands	Taiwan, Philippines, Indonesia
South Korea	South Korea
Japan	Japan
Australia	Australia
India	India
Russia	Russia
Brazil	Brazil
Israel + Turkey	Israel, Turkey

B Supplementary summary statistics

Table 6: Changes in the number of billionaires in the Forbes dataset over time

Year	N	Change	New	Left list	Remained	Share remained
1997	2	0	0	0	0	0.00
1998	1	-1	1	2	0	0.00
1999	8	7	7	0	1	0.12
2000	8	0	5	5	3	0.38
2001	335	327	332	5	3	0.01
2002	333	-2	27	29	306	0.92
2003	332	-1	38	39	294	0.89
2004	432	100	101	1	331	0.77
2005	530	98	101	3	429	0.81
2006	628	98	108	10	520	0.83
2007	761	133	147	14	614	0.81
2008	908	147	223	76	685	0.75
2009	738	-170	39	209	699	0.95
2010	1011	273	354	81	657	0.65
2011	1209	198	255	58	953	0.79
2012	1226	17	212	195	1013	0.83
2013	1426	200	304	104	1121	0.79
2014	1645	219	360	141	1284	0.78
2015	1826	181	347	167	1477	0.81
2016	1811	-15	255	270	1554	0.86
2017	2043	232	375	143	1666	0.82
2018	2208	165	392	226	1815	0.82
2019	2153	-55	275	331	1876	0.87
2020	2095	-58	257	315	1836	0.88
2021	2755	660	851	191	1902	0.69
2022	2668	-87	418	505	2248	0.84
2023	2640	-28	286	314	2352	0.89

Note: ‘N’ refers to the number of billionaires. ‘Left list’ indicates how many billionaires were removed from a list in a year. ‘Remained’ indicates how many billionaires were also on the list in the previous year.

Table 7: Summary statistics of the Bloomberg 500 richest people dataset, 2021

region	N	Mean Net Worth	Mean Log Net Worth
Central Eurasia	28.0	14.1	2.5
China	95.0	15.5	2.6
East Asia	39.0	13.0	2.4
Europe	112.0	13.7	2.3
India	17.0	19.9	2.7
Middle East	8.0	8.8	2.1
North America	170.0	19.6	2.5
South America	14.0	16.6	2.6
Rest of World	15.0	13.9	2.3

Note: 'N' refers to the number of billionaires.

C Definitions of distributions

This section defines the distributions and corresponding notation as they are used in this paper.

C.1 Pareto distribution

The cumulative distribution function (CDF) of the Pareto distribution is defined as:

$$F(x; x_m, \alpha) = 1 - \left(\frac{x}{x_m}\right)^{-\frac{1}{\alpha}}$$

where x_m is the scale parameter and α is the shape parameter.

The probability density function (PDF) of the Pareto distribution is defined as:

$$f(x; x_m, \alpha) = \frac{1}{\alpha} \frac{x_m^{\frac{1}{\alpha}}}{x^{\frac{1}{\alpha}+1}}$$

for $x \geq x_m$.

C.2 Exponential distribution

The cumulative distribution function (CDF) of the Exponential distribution is defined as:

$$F(x; \alpha) = 1 - e^{-\frac{x}{\alpha}}$$

where α is the scale parameter.

The probability density function (PDF) of the Exponential distribution is defined as:

$$f(x; \alpha) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}}$$

for $x \geq 0$.

The hazard function, or hazard rate, of the Exponential distribution is given by:

$$h(x;\alpha) = \frac{1}{\alpha}$$

The cumulative hazard function of the Exponential distribution is given by:

$$H(x;\alpha) = -\log(1 - F(x;\alpha)) = \frac{x}{\alpha}$$

C.3 Weibull distribution

The cumulative distribution function (CDF) of the Weibull distribution is defined as:

$$F(x;\gamma,\alpha) = 1 - \exp\left(\frac{1 - x^{\gamma}}{\alpha\gamma}\right)$$

The probability density function (PDF) of the Weibull distribution is defined as:

$$f(x;\gamma,\alpha) = \frac{1}{\alpha}x^{\gamma-1}\exp\left(\frac{1 - x^{\gamma}}{\alpha\gamma}\right)$$

for $x \geq 1$.

C.4 Gompertz distribution

The cumulative distribution function (CDF) of the Gompertz distribution is defined as:

$$F(x;\gamma,\alpha) = 1 - \exp\left(\frac{1 - \exp(\gamma x)}{\alpha\gamma}\right)$$

The probability density function (PDF) of the Gompertz distribution is defined as:

$$f(x;\gamma,\alpha) = \frac{1}{\alpha}\exp\left(\gamma x - \frac{\exp(\gamma x) - 1}{\alpha\gamma}\right)$$

The hazard function, or hazard rate, of the Gompertz distribution is given by:

$$h(x;\gamma,\alpha) = \frac{1}{\alpha}\exp\left(\gamma x\right)$$

The cumulative hazard function of the Gompertz distribution is given by:

$$H(x;\gamma,\alpha) = -\log(1 - F(x;\gamma,\alpha)) = \frac{\exp(\gamma x) - 1}{\alpha\gamma}$$

C.5 Generalised Pareto distribution

The cumulative distribution function (CDF) of the Generalised Pareto distribution is defined as:

$$F(x;\gamma,\sigma,\mu) = 1 - \left(1 + \frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}$$

for $\gamma \neq 0$. When $\gamma = 0$, it simplifies to:

$$F(x;\sigma,\mu) = 1 - \exp\left(-\frac{x-\mu}{\sigma}\right)$$

The probability density function (PDF) of the Generalised Pareto distribution is defined as:

$$f(x;\gamma,\sigma,\mu) = \frac{1}{\sigma}\left(1 + \frac{\gamma(x-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}-1}$$

for $\gamma \neq 0$. When $\gamma = 0$, it simplifies to:

$$f(x;\sigma,\mu) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)$$

C.6 Exponentiated generalised Pareto distribution

The cumulative distribution function (CDF) of the Exponentiated Generalised Pareto distribution is defined as:

$$F(x;\gamma,\sigma,\mu) = 1 - \left(1 + \frac{\gamma(\exp(x)-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}}$$

for $\gamma \neq 0$. When $\gamma = 0$, it simplifies to:

$$F(x;\sigma,\mu) = 1 - \exp\left(-\frac{\exp(x)-\mu}{\sigma}\right)$$

The probability density function (PDF) of the Exponentiated Generalised Pareto distribution is defined as:

$$f(x;\gamma,\sigma,\mu) = \frac{\exp(x)}{\sigma}\left(1 + \frac{\gamma(\exp(x)-\mu)}{\sigma}\right)^{-\frac{1}{\gamma}-1}$$

for $\gamma \neq 0$. When $\gamma = 0$, it simplifies to:

$$f(x;\sigma,\mu) = \frac{\exp(x)}{\sigma}\exp\left(-\frac{\exp(x)-\mu}{\sigma}\right)$$

The hazard function, or hazard rate, of the Exponentiated Generalised Pareto distribution is given by:

$$h(x;\gamma,\sigma,\mu) = \frac{f(x;\gamma,\sigma,\mu)}{1 - F(x;\gamma,\sigma,\mu)}$$

The cumulative hazard function of the Exponentiated Generalised Pareto distribution is given by:

$$H(x;\gamma,\sigma,\mu) = -\log(1 - F(x;\gamma,\sigma,\mu)) = \begin{cases} \frac{1}{\gamma}\log\left(1 + \frac{\gamma(\exp(x)-\mu)}{\sigma}\right), & \gamma \neq 0 \\ \frac{\exp(x)-\mu}{\sigma}, & \gamma = 0 \end{cases}$$

C.7 Gamma distribution

The gamma distribution can be parameterised in terms of a shape parameter α and an inverse scale parameter β. A random variable X that is gamma-distributed with shape α and rate β is denoted

$$X \sim \Gamma(\alpha, \beta) \equiv \text{Gamma}(\alpha, \beta)$$

The corresponding probability density function in the shape-rate parameterisation is

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1} e^{-\beta x} \beta^{\alpha}}{\Gamma(\alpha)} \quad \text{for } x > 0, \quad \alpha, \beta > 0,$$

where $\Gamma(\alpha)$ is the Gamma function. For all positive integers, $\Gamma(\alpha) = (\alpha - 1)!$.

C.8 Inverse Gamma distribution

The inverse gamma distribution's probability density function is defined over the support $x > 0$

$$f(x; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} \exp\left(-\frac{\beta}{x}\right)$$

with shape parameter α and scale parameter β. Here $\Gamma(\cdot)$ denotes the gamma function.

D Derivation of the posterior distribution - Pareto

For this derivation only, we parameterise the Pareto distribution with $\theta = \frac{1}{\alpha}$ to prevent confusion with the α parameter of the Gamma prior, and to simplify derivations. We prove that a Gamma distribution is a conjugate prior for θ, which implies that the Inverse Gamma distribution is a conjugate prior for γ. Let $y_i \overset{\text{iid}}{\sim} \text{Pareto}(\theta, x_m)$, with the probability density function

$$f(y_i \mid \theta, x_m) = \frac{\theta x_m^{\theta}}{y_i^{\theta+1}}$$

So the joint likelihood function for $y = (y_1, y_2, \ldots, y_n)$ is

$$\begin{aligned} f(y \mid \theta, x_m) &= \prod_{i=1}^{n} f(y_i \mid \theta, x_m) \\ &= \prod_{i=1}^{n} \left(\frac{\theta x_m^{\theta}}{y_i^{\theta+1}}\right) \\ &= \theta^n x_m^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)} \\ &= \theta^n x_m^{n\theta} \left(\prod_{i=1}^{n} y_i\right)^{-(\theta+1)} \end{aligned}$$

Assume $\theta \sim \text{Gamma}(\alpha, \beta)$, with the probability density function

$$f(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

The posterior distribution $f(\theta \mid y)$ is proportional to the product of the likelihood function and the prior distribution:

$$
\begin{aligned}
f(\theta \mid y) &\propto f(y \mid \theta, x_m) f(\theta) \\
&= \theta^n x_m^{n\theta} \left(\prod_{i=1}^{n} y_i \right)^{-(\theta+1)} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\
&\propto \theta^{n+\alpha-1} x_m^{n\theta} e^{-\beta\theta} \left(\prod_{i=1}^{n} y_i \right)^{-\theta-1} \\
&= \theta^{n+\alpha-1} x_m^{n\theta} e^{-\beta\theta} \left(\prod_{i=1}^{n} y_i \right)^{-\theta-1} \\
&= \theta^{n+\alpha-1} x_m^{n\theta} e^{-\beta\theta} e^{-\theta \sum_{i=1}^{n} \log y_i - \sum_{i=1}^{n} \log y_i} \\
&= \theta^{n+\alpha-1} e^{-\theta(\beta + \sum_{i=1}^{n} \log(y_i/x_m))} x_m^{n\theta} e^{-\sum_{i=1}^{n} \log y_i}
\end{aligned}
$$

Thus, the posterior distribution is also a Gamma distribution:

$$
\theta \mid y \sim \text{Gamma}\left(\alpha + n, \beta + \sum_{i=1}^{n} \log\left(\frac{y_i}{x_m} \right) \right)
$$

E Run configurations

Table 8: Covariates included in each group by start year and end years

Group	Start Year	End Years	Constant	GDP	MSCI World	FTSE100	SSE	CAC40	DAX	NIFTY	MOEX	OMX40	SPX
Alps	2013	2021	X	X	X								
Asian Islands	2010	2019, 2020, 2021	X	X	X								
Australia	2013	2021	X	X	X								
Brazil	2011	2020, 2021	X	X	X								
British Islands	2007	2019, 2020, 2021	X	X		X							
Canada	2010	2019, 2020, 2021	X	X	X								
China	2007	2019, 2020, 2021	X	X			X						
France	2013	2021	X	X				X					
Germany	2005	2019, 2020, 2021	X	X					X				
India	2009	2019, 2020, 2021	X	X						X			
Israel + Turkey	2010	2019, 2020, 2021	X	X	X								
Italy	2014	2021	X	X	X								
Japan	2014	2021	X	X	X								
Russia	2015	2021	X	X							X		
Scandinavia	2016	2021	X	X								X	
South Korea	2015	2021	X	X	X								
Southeast Asia	2013	2021	X	X	X								
U.S.	2005	2019, 2020, 2021	X	X									X

Note: 'GDP' refers to the log growth in GDP per capita. All indices refer to the log return in the respective index. Each model is estimated once with all three covariates, and once without the stock index return for mean wealth forecasting purposes. The start year indicates the first training year for a model, and the end year the last training year. When multiple end years are indicated, the model is estimated multiple times with different estimation windows, once for each end year. This is such that there is at least one one-step-ahead forecast available for each year. The year 2023 is never forecasted for, as GDP growth data for it is not yet available across the board.

F Frequentist estimation summary statistics

Table 9: Summary statistics of the frequentist scale parameter

	n_years	mean_α_{pareto}	min_α_{pareto}	max_α_{pareto}	mean_$\alpha_{Weibull}$	min_$\alpha_{Weibull}$	max_$\alpha_{Weibull}$	mean_σ_{GPD}	min_σ_{GPD}	max_σ_{GPD}
Southeast Asia	11.00	0.93	0.86	1.08	1.19	1.06	1.42	1.34	1.17	1.75
India	17.00	0.97	0.78	1.37	1.27	0.97	1.98	1.43	0.96	2.72
Germany	21.00	1.12	0.86	1.31	1.44	1.03	1.73	1.95	1.26	2.66
Alps	11.00	1.07	0.89	1.26	1.35	1.10	1.67	1.82	1.27	2.40
Brazil	13.00	0.92	0.74	1.12	1.17	0.93	1.45	1.32	0.85	1.92
China	18.00	0.77	0.59	0.97	0.98	0.73	1.28	0.92	0.57	1.38
France	11.00	1.31	1.02	1.64	1.85	1.40	2.46	2.52	1.40	4.03
Scandinavia	11.00	1.02	0.90	1.23	1.31	1.13	1.60	1.63	1.24	2.34
Japan	14.00	0.88	0.77	1.01	1.13	1.01	1.33	1.14	0.80	1.51
U.S.	23.00	0.97	0.75	1.22	1.27	0.95	1.65	1.44	0.87	2.19
British Islands	18.00	0.94	0.69	1.17	1.17	0.82	1.51	1.43	0.81	2.13
Italy	11.00	0.97	0.11	0.86	1.20	1.10	1.60	1.44	1.07	2.18
Canada	15.00	0.92	0.83	1.09	1.15	1.05	1.38	1.32	1.06	1.87
Israel + Turkey	18.00	0.61	0.41	0.84	0.73	0.47	1.04	0.68	0.31	1.13
Asian Islands	14.00	0.78	0.67	0.87	0.95	0.79	1.09	1.01	0.79	1.20
Australia	11.00	0.84	0.69	1.07	1.06	0.84	1.41	1.10	0.75	1.69
Russia	19.00	1.04	0.90	1.45	1.37	1.18	1.90	1.65	1.13	3.28
South Korea	11.00	0.67	0.57	0.77	0.82	0.69	0.93	0.73	0.56	0.97

G Replication

As this researched was produced in the context of a Bachelor's thesis, it contains a mandatory replication of parts of the results of the assigned paper from Teulings and Toussaint (2023). These are presented in this section. The sub-section headings refer to the numbering of figures and tables in the original paper. The methodology is the same as in the original paper.

G.1 Figure 2 - $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$ histograms

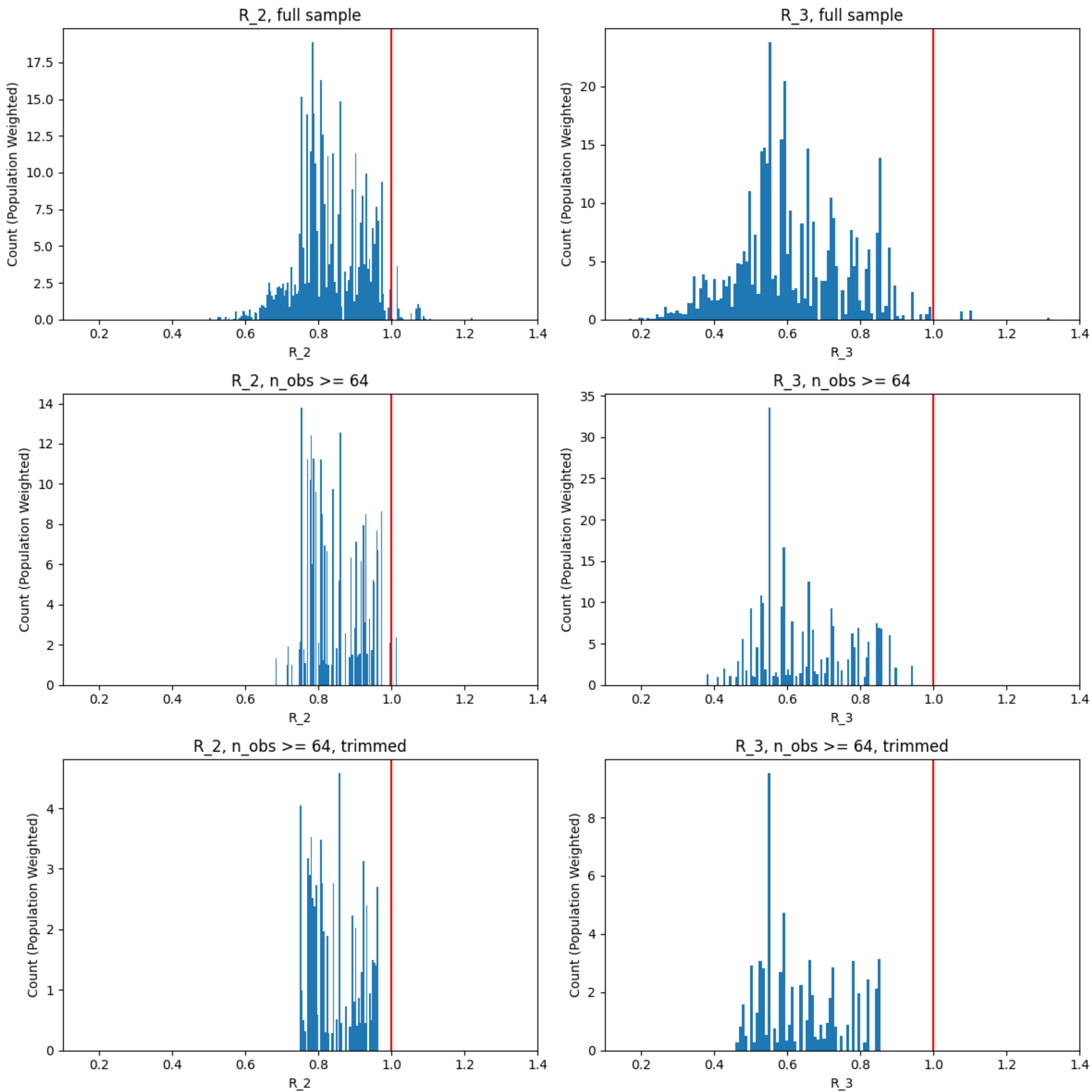


Figure 17: Distribution of test statistics for $\hat{\mathcal{R}}_2$ and $\hat{\mathcal{R}}_3$

G.2 Table 5 - Mean wealth predictions

Table 10: Pareto and Weibull Estimates of Billionaire Wealth in 2018

	Mean Data	Mean Weibull	Mean Pareto	Alpha_Weibull	Alpha_Pareto
U.S.	5.29	5.48	∞	1.52	1.16
Canada	3.23	3.22	5.96	1.04	0.83
Germany	4.70	5.58	∞	1.53	1.18
British Islands	3.83	4.26	∞	1.29	1.01
Scandinavia	3.51	4.06	226.69	1.24	1.00
France	7.44	8.12	∞	1.89	1.37
Alps	3.80	4.66	∞	1.37	1.10
Italy	3.96	4.29	∞	1.29	1.02
China	3.31	3.10	5.00	1.01	0.80
Southeast Asia	3.33	3.51	9.43	1.12	0.89
Japan	3.95	3.82	12.42	1.19	0.92
Asian Islands	3.17	3.22	6.18	1.04	0.84
South Korea	2.88	2.78	3.80	0.91	0.74
Japan	3.95	3.82	12.42	1.19	0.92
Australia	2.74	2.75	3.86	0.90	0.74
India	3.70	3.90	24.11	1.21	0.96
Russia	4.05	4.07	21.35	1.25	0.95
Brazil	4.20	4.42	∞	1.32	1.03
Israel + Turkey	2.17	2.24	2.64	0.73	0.62

G.3 Table 1 - Summary statistics

Table 11: Summary Statistics of Billionaire Wealth by Region and Sub-Region

	R2	R3	Log wealth mean	Billionaires per capita	Billionaire count
North America	0.85	0.65	0.94	1.25	440.81
Europe	0.77	0.49	1.05	0.48	247.71
China	0.93	0.76	0.82	0.13	185.86
East Asia	0.80	0.53	0.78	0.16	129.48
India	0.82	0.57	0.99	0.04	54.71
Central Eurasia	0.82	0.56	0.93	0.35	75.62
South America	0.82	0.59	0.99	0.11	56.81
Middle East	0.86	0.65	0.78	0.24	56.14
U.S.	0.85	0.66	0.95	1.30	413.29
Canada	0.76	0.49	0.89	0.78	27.52
Germany	0.71	0.41	1.13	0.83	67.90
British Islands	0.76	0.48	0.87	0.55	38.14
Scandinavia	0.76	0.46	1.20	1.15	30.00
France	0.77	0.49	1.33	0.40	25.14
Alps	0.64	0.32	1.10	1.37	23.24
Italy	0.76	0.49	1.08	0.38	22.90
China	0.93	0.76	0.82	0.13	185.52
Southeast Asia	0.79	0.55	0.88	0.28	29.62
Asian Islands	0.76	0.48	0.69	0.10	37.76
South Korea	0.83	0.62	0.71	0.36	18.24
Japan	0.80	0.54	0.87	0.18	22.67
Australia	0.83	0.58	0.71	0.76	18.19
India	0.82	0.57	0.99	0.04	54.71
Russia	0.81	0.55	0.97	0.46	66.76
Brazil	0.81	0.56	0.84	0.14	29.24
Israel + Turkey	0.88	0.63	0.60	0.41	34.71
Rest of World	0.74	0.44	1.01	0.00	10.19

Note: Billionaires per capita are denoted in the number of billionaires per a million people. Values are averaged across time.

G.4 Table 2 - WLS regressions, Pareto test

Table 12: WLS Regressions, Pareto Test

	$\hat{R}_2$			$\hat{R}_3$		
Model:	R2 full	R2 n64	R2 trimmed	R3 full	R3 n64	R3 trimmed
Variables						
Constant	0.82***	0.85***	0.85***	0.58***	0.65***	0.64***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics						
Observations	378	75	67	378	75	67
RMSE	0.13	0.08	0.07	0.20	0.13	0.12
Theoretical RMSE	1	1	1	$\sqrt{10}$	$\sqrt{10}$	$\sqrt{10}$

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

G.5 Table 4 - Weibull predictions

Table 13: Regression Results of $\hat{R}_2$ and $\hat{R}_3$ on Gompertz Parameters

	$\hat{R}_2$		$\hat{R}_3$
Model:	(1)	(2)	(3)
Variables			
Constant	0.76***	0.50***	-0.74***
	(0.02)	(0.02)	(0.08)
log $\hat{\alpha}$	-0.10**	-0.19***	
	(0.03)	(0.02)	
log $\hat{\gamma}$	-0.07***	-0.12***	
	(0.01)	(0.01)	
$\hat{R}_2$			1.63***
			(0.10)
Weights	$\sqrt{N}$	$\sqrt{N}$	$\sqrt{N}$
Fit statistics			
Observations	67	67	67
R^2	0.95	0.99	0.95
Adjusted R^2	0.95	0.99	0.95
RMSE	0.06	0.05	0.10
F-Stat	0.32	5.38	

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

H Weibull γ estimates

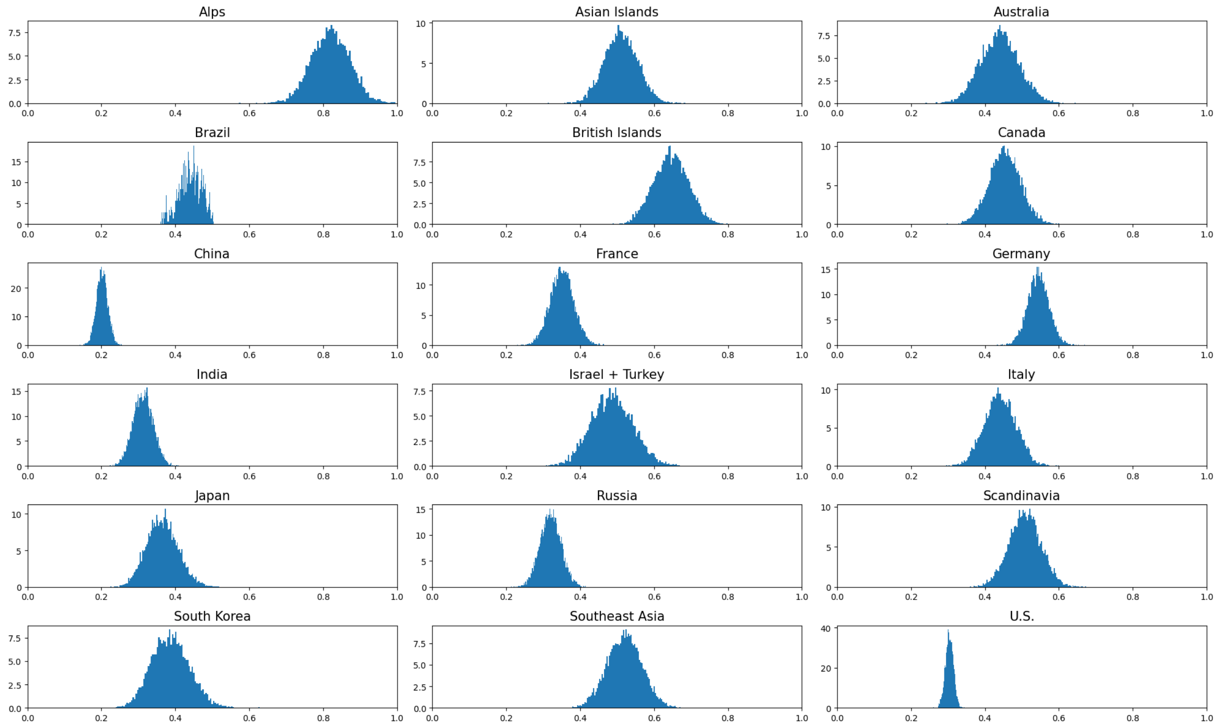


Figure 18: Weibull γ posterior distributions across sub-regions, regularised models

I Posterior distribution plots

I.1 Pareto posteriors

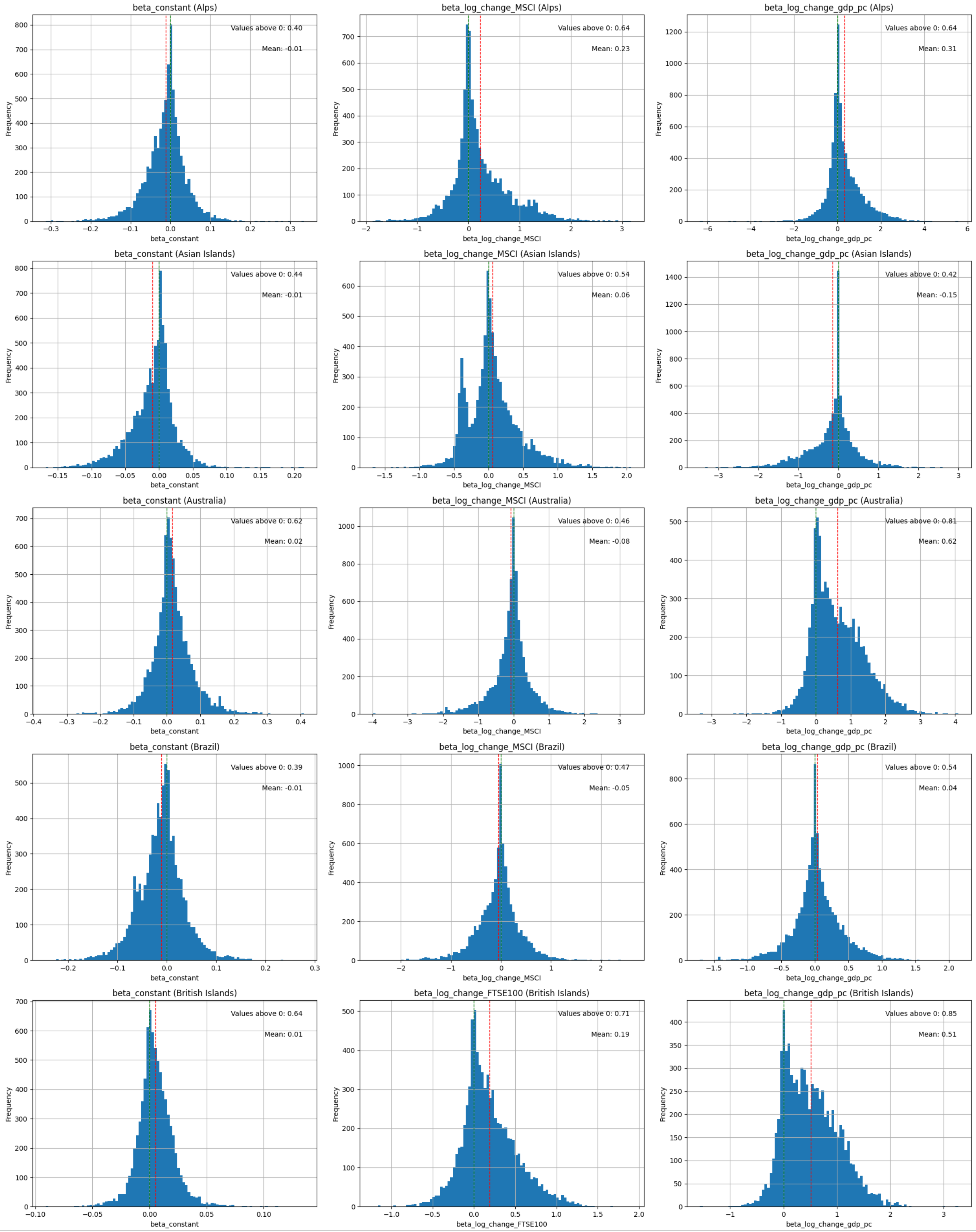


Figure 19: Posterior parameter estimates for the Pareto time series model (1/4)

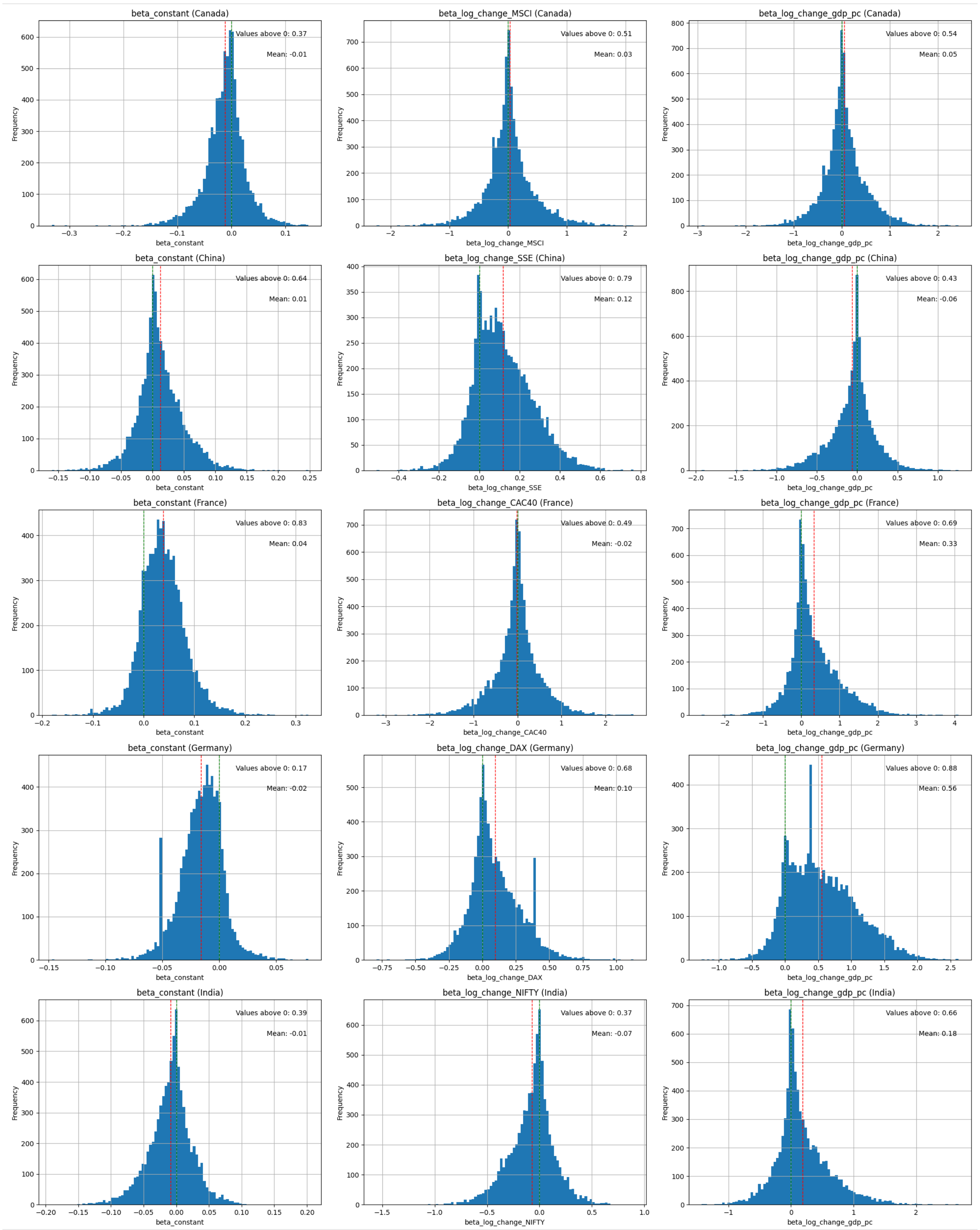


Figure 20: Posterior parameter estimates for the Pareto time series model (2/4)

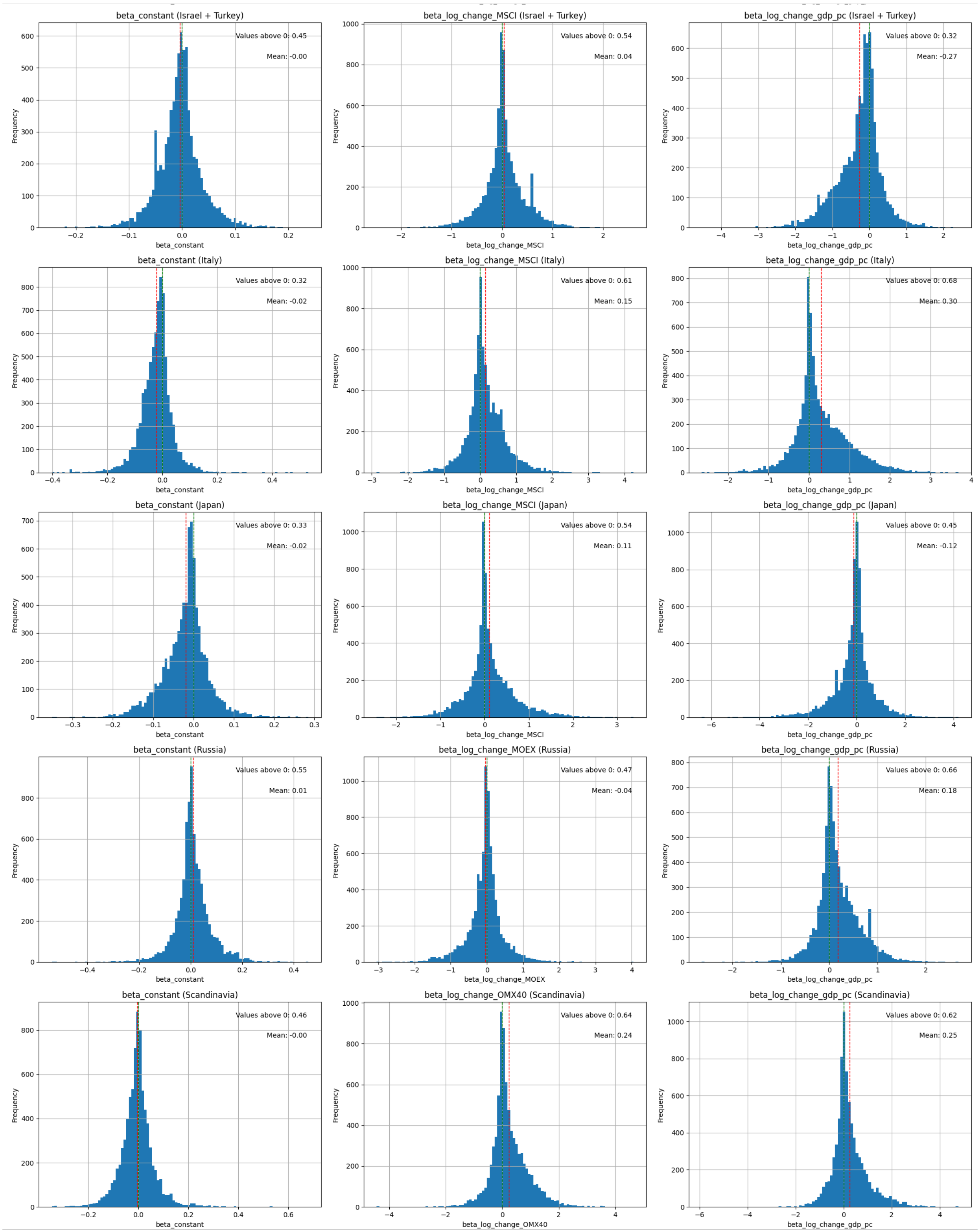


Figure 21: Posterior parameter estimates for the Pareto time series model (3/4)

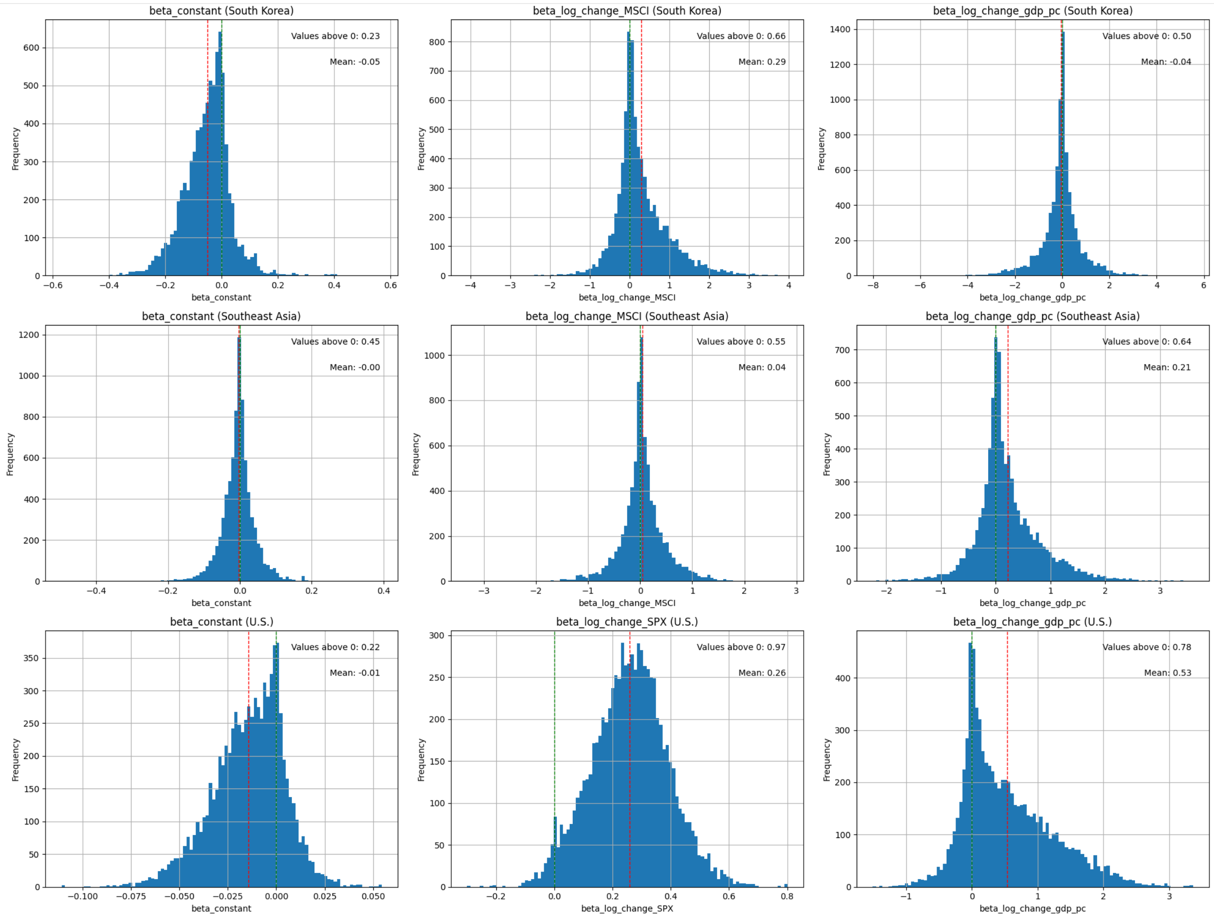


Figure 22: Posterior parameter estimates for the Pareto time series model (4/4)

I.2 Weibull posteriors

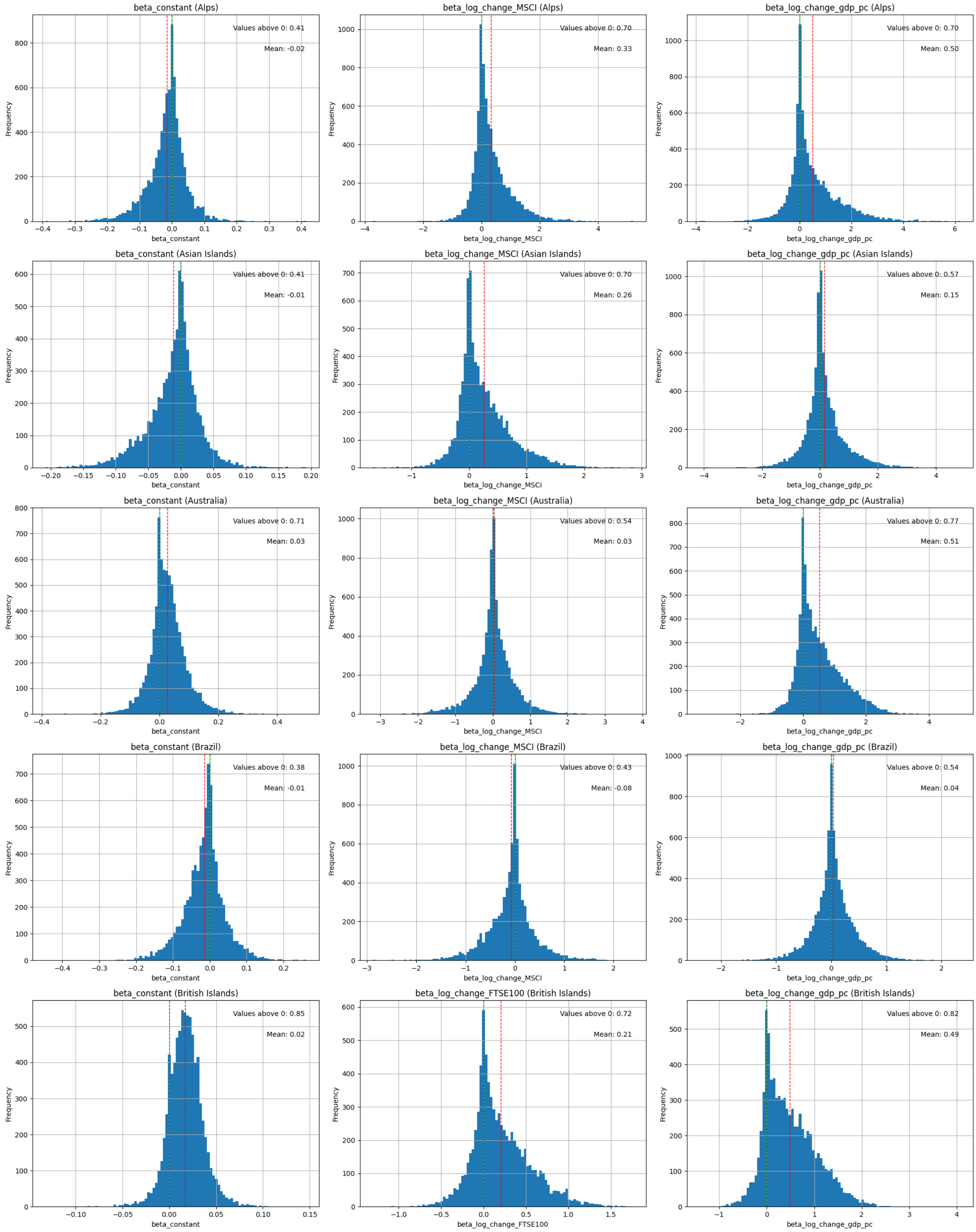


Figure 23: Posterior parameter estimates for the Weibull time series model (1/4)

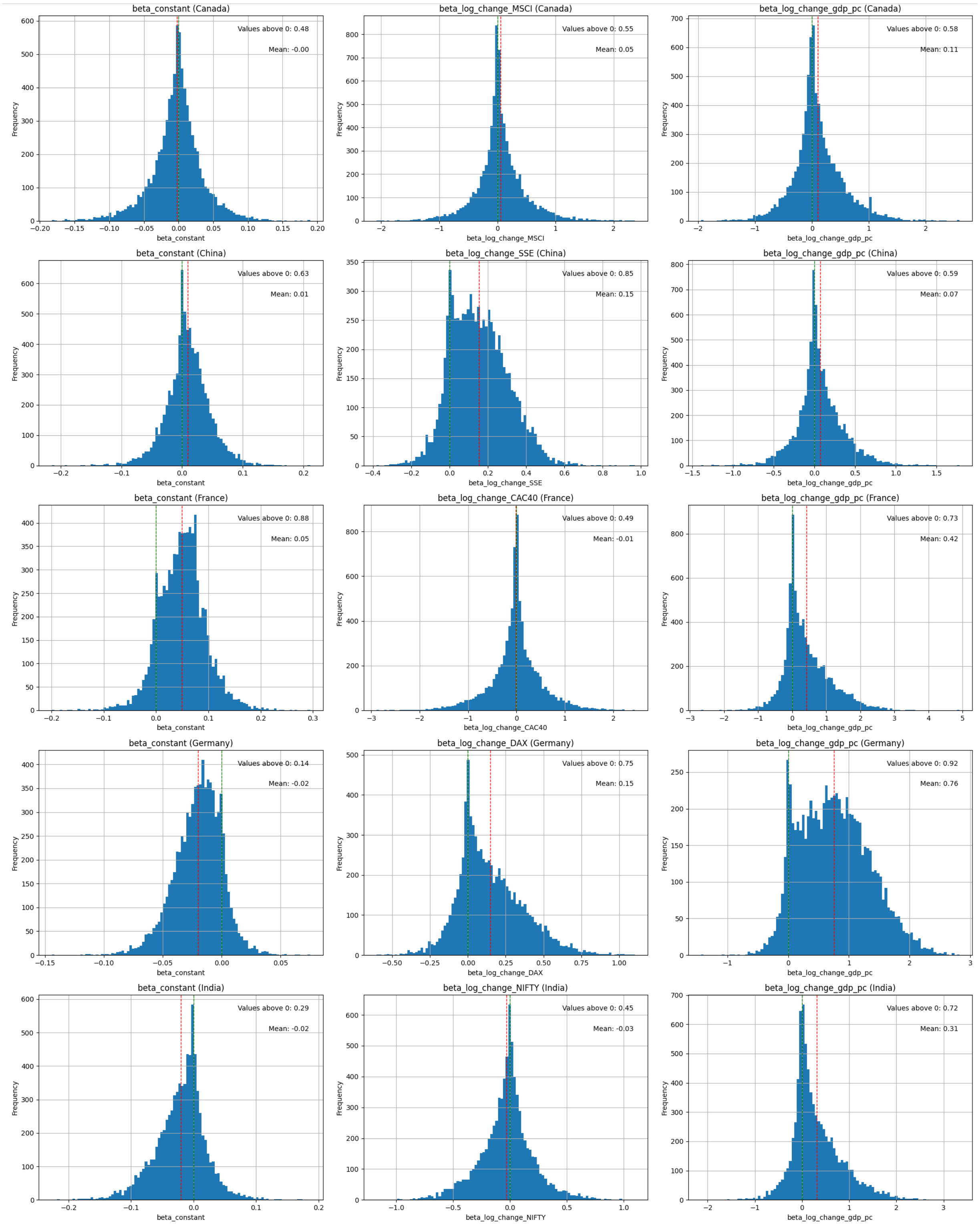


Figure 24: Posterior parameter estimates for the Weibull time series model (2/4)

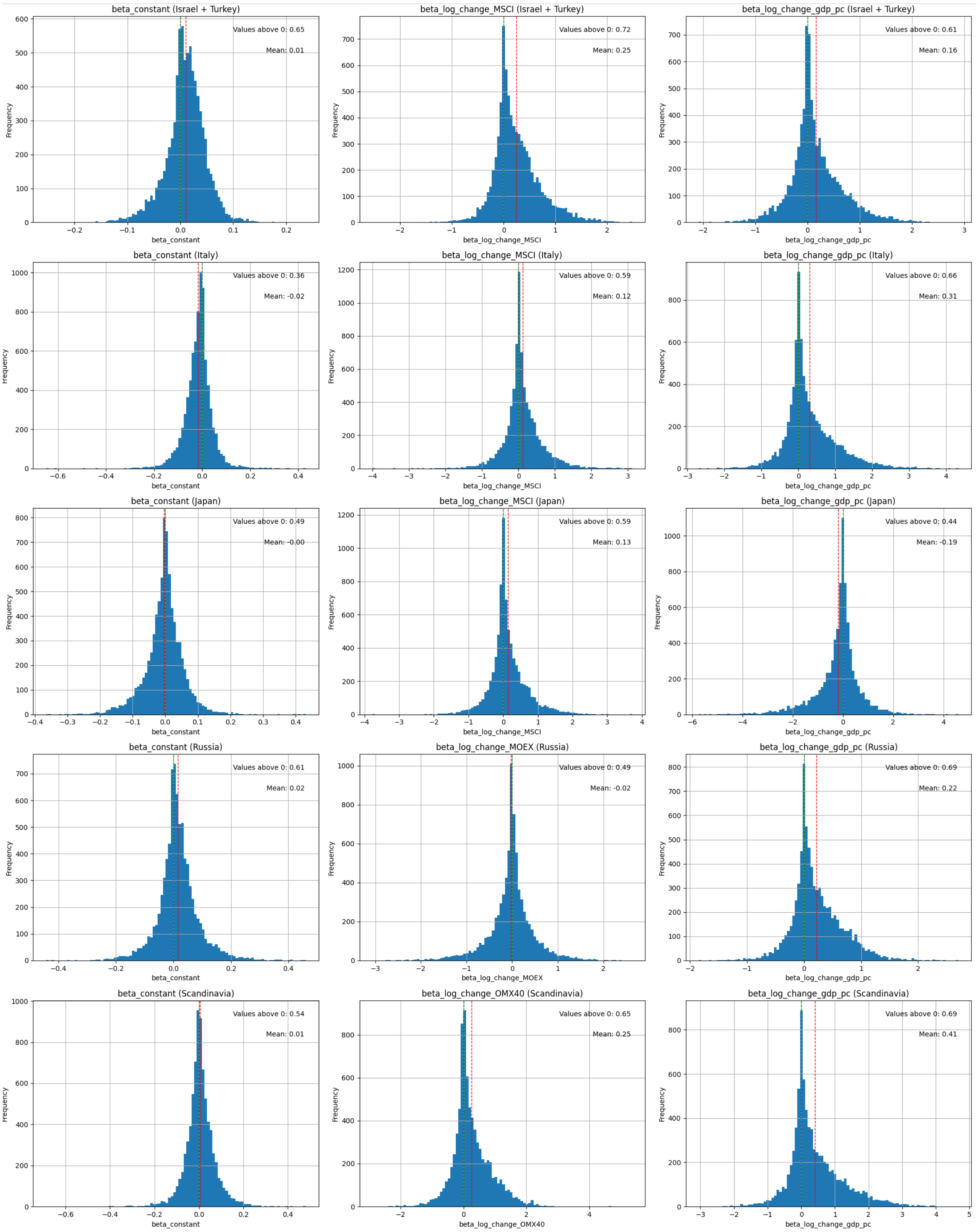


Figure 25: Posterior parameter estimates for the Weibull time series model (3/4)

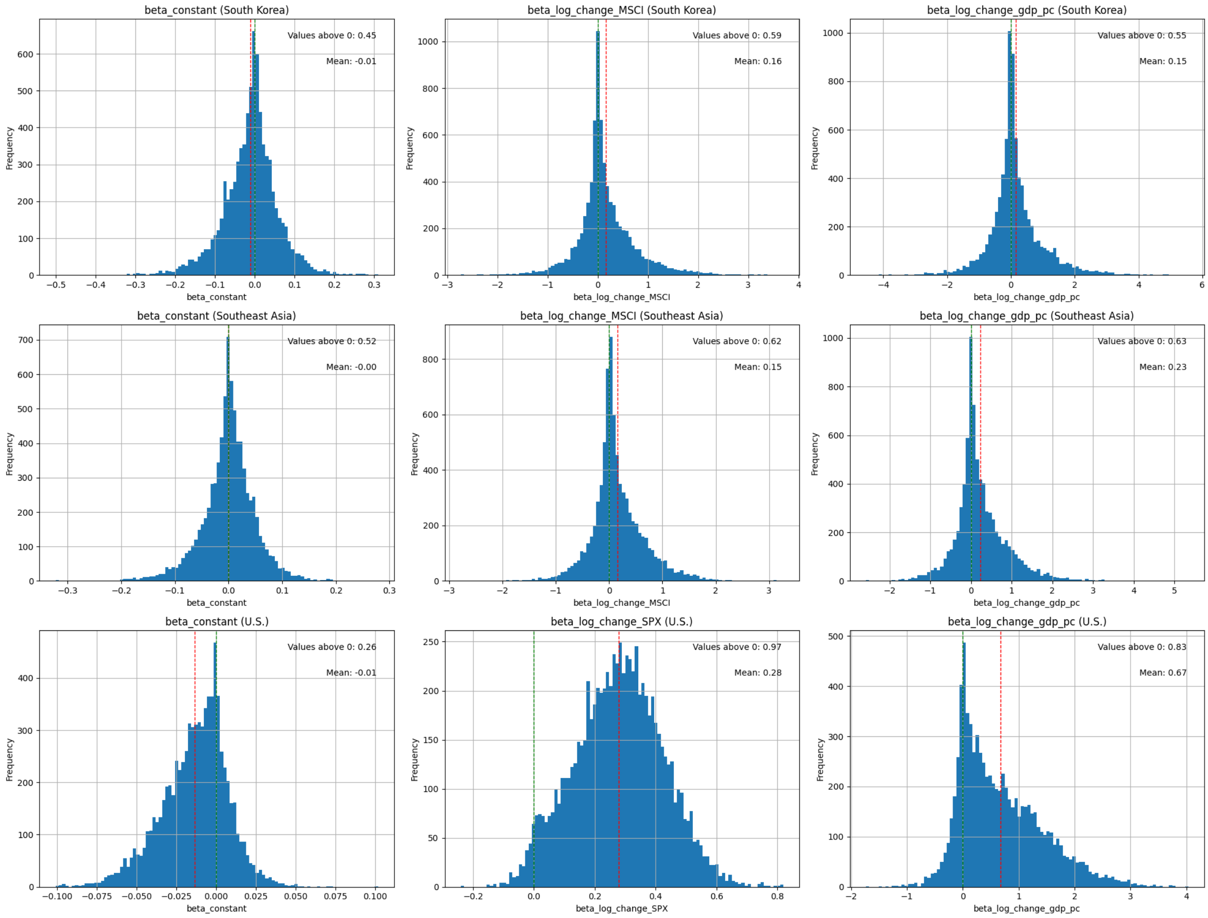


Figure 26: Posterior parameter estimates for the Weibull time series model (4/4)

I.3 Generalised Pareto posteriors

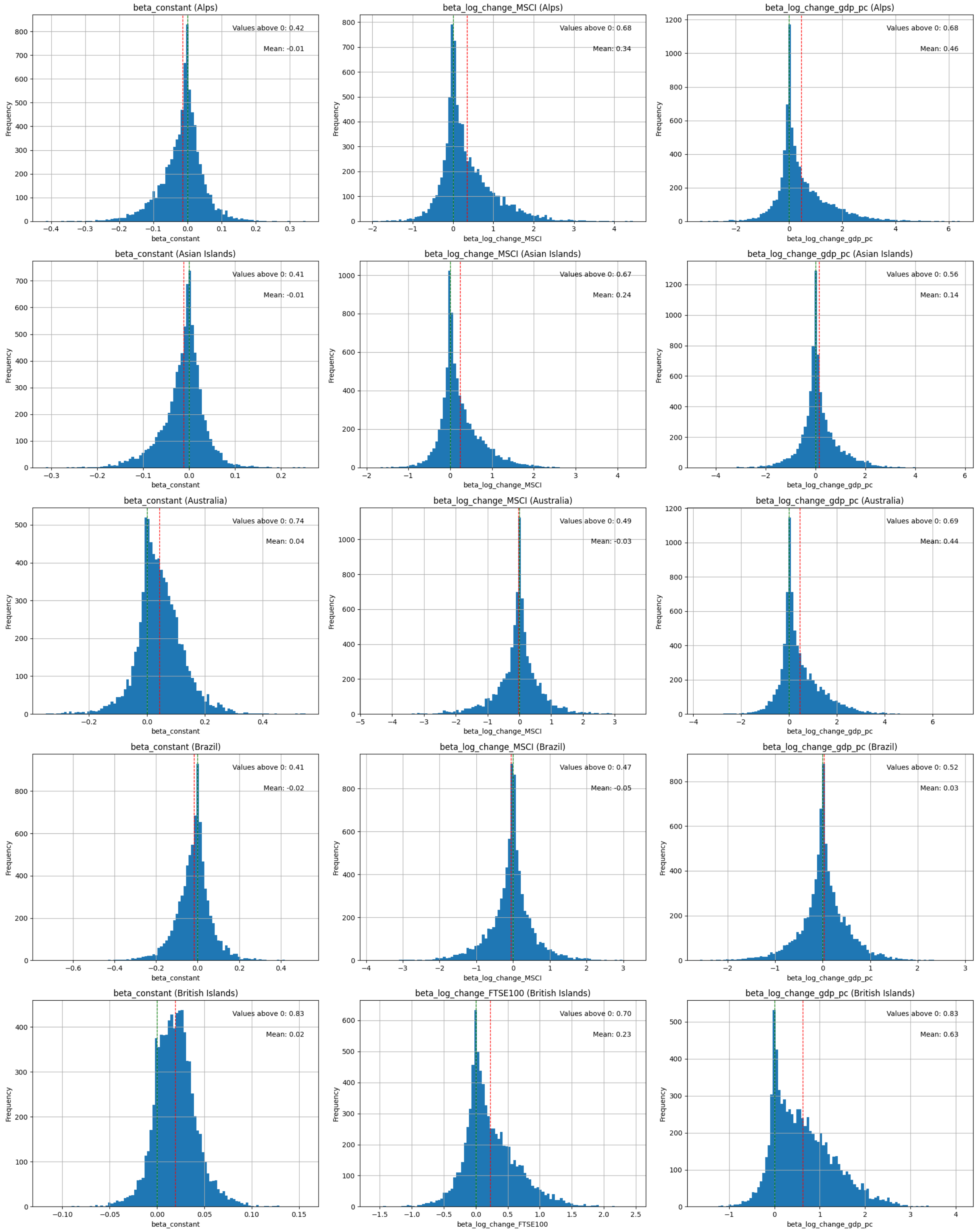


Figure 27: Posterior parameter estimates for the Generalised Pareto time series model (1/4)

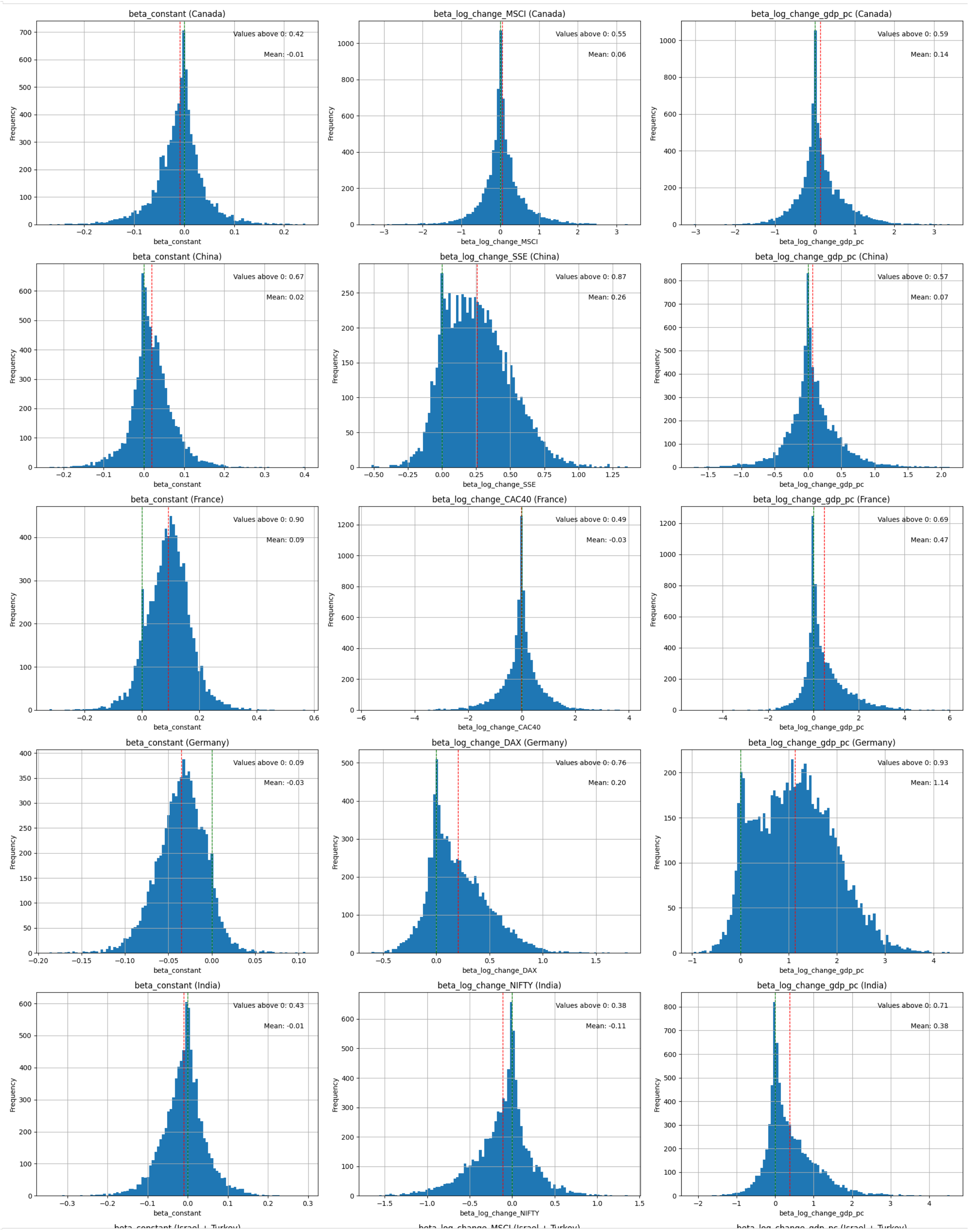


Figure 28: Posterior parameter estimates for the Generalised Pareto time series model (2/4)

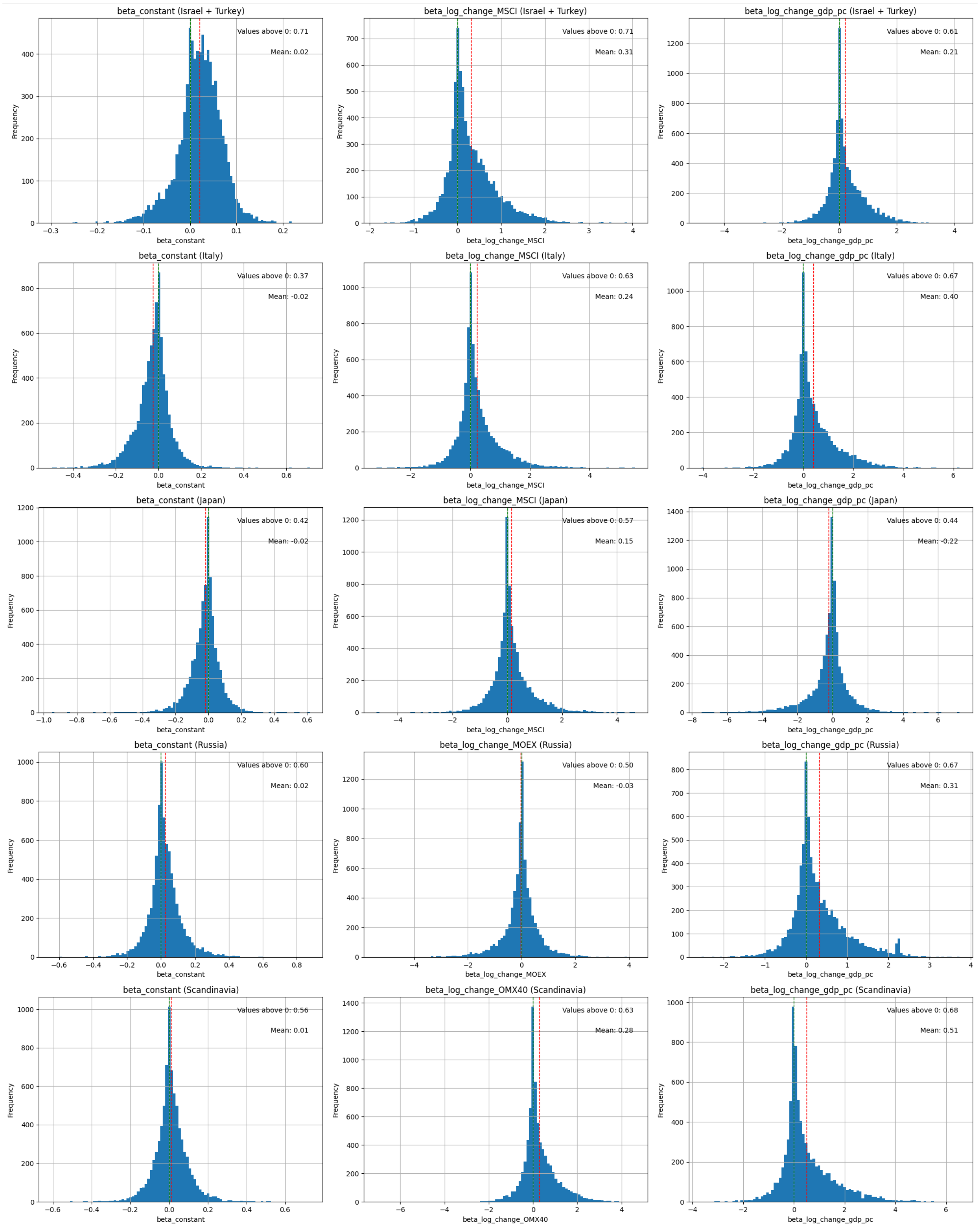


Figure 29: Posterior parameter estimates for the Generalised Pareto time series model (3/4)

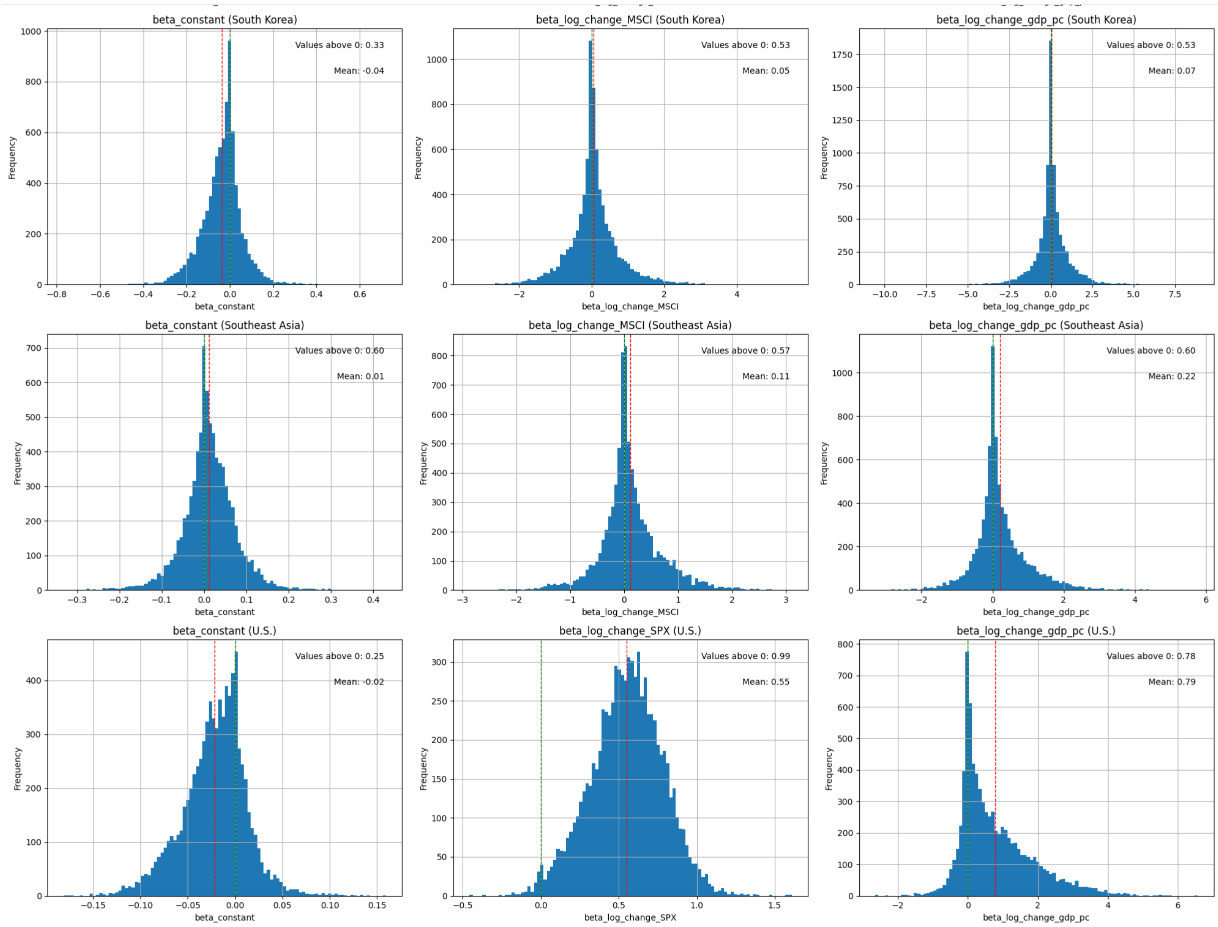


Figure 30: Posterior parameter estimates for the Generalised Pareto time series model (4/4)

J Summary statistics - alternative R^2

Table 14: Summary statistics of the R^2 values of the Generalised Pareto time series model with two covariates

	count	mean	std	min	25%	median	75%	max
Alps	8000	0.62	0.17	0.10	0.50	0.53	0.71	1.00
Asian Islands	8000	0.59	0.15	0.26	0.50	0.52	0.63	1.00
Australia	8000	0.62	0.17	0.08	0.50	0.53	0.72	1.00
Brazil	8000	0.54	0.11	0.23	0.50	0.50	0.53	1.00
British Islands	8000	0.72	0.18	0.32	0.54	0.71	0.90	1.00
Canada	8000	0.63	0.18	0.25	0.50	0.55	0.75	1.00
China	8000	0.50	0.02	0.25	0.50	0.50	0.51	0.80
France	8000	0.61	0.17	0.14	0.50	0.54	0.70	1.00
Germany	8000	0.81	0.17	0.36	0.67	0.87	0.96	1.00
India	8000	0.58	0.14	0.22	0.50	0.52	0.62	1.00
Israel + Turkey	8000	0.60	0.16	0.18	0.50	0.52	0.67	1.00
Italy	8000	0.62	0.17	0.09	0.50	0.54	0.73	1.00
Japan	8000	0.57	0.14	0.12	0.50	0.51	0.60	1.00
Russia	8000	0.61	0.18	0.09	0.50	0.54	0.72	1.00
Scandinavia	8000	0.62	0.19	0.04	0.50	0.54	0.75	1.00
South Korea	8000	0.56	0.13	0.09	0.49	0.51	0.58	1.00
Southeast Asia	8000	0.58	0.15	0.13	0.50	0.52	0.62	1.00
U.S.	8000	0.67	0.11	0.34	0.59	0.67	0.75	0.97
Average	8000	0.61	0.15	0.18	0.51	0.56	0.69	0.99

Table 15: Summary statistics of the R^2 values of the Generalised Pareto time series model with three covariates

	count	mean	std	min	25%	median	75%	max
Alps	8000	0.68	0.18	0.15	0.52	0.64	0.84	1.00
Asian Islands	8000	0.66	0.17	0.29	0.51	0.61	0.81	1.00
Australia	8000	0.64	0.18	0.15	0.50	0.58	0.77	1.00
Brazil	8000	0.53	0.10	0.24	0.49	0.51	0.54	1.00
British Islands	8000	0.78	0.18	0.30	0.62	0.81	0.95	1.00
Canada	8000	0.65	0.17	0.23	0.51	0.59	0.77	1.00
China	8000	0.61	0.11	0.26	0.51	0.59	0.69	0.98
France	8000	0.65	0.18	0.17	0.51	0.59	0.80	1.00
Germany	8000	0.85	0.15	0.33	0.75	0.90	0.98	1.00
India	8000	0.64	0.17	0.24	0.51	0.58	0.75	1.00
Israel + Turkey	8000	0.68	0.18	0.24	0.52	0.64	0.84	1.00
Italy	8000	0.66	0.19	0.03	0.51	0.60	0.83	1.00
Japan	8000	0.62	0.17	0.10	0.50	0.55	0.72	1.00
Russia	8000	0.65	0.20	0.06	0.50	0.60	0.82	1.00
Scandinavia	8000	0.67	0.20	0.02	0.51	0.64	0.86	1.00
South Korea	8000	0.64	0.20	0.03	0.50	0.58	0.81	1.00
Southeast Asia	8000	0.63	0.18	0.13	0.50	0.57	0.76	1.00
U.S.	8000	0.78	0.08	0.42	0.73	0.79	0.84	0.97
Average	8000	0.67	0.17	0.19	0.54	0.63	0.80	1.00