Supplementary Information (SI) for

Empirical confirmation of bosonic wealth statistics in Bitcoin UTXOs

Chanhee Park, Claudio J. Tessone*, Yu Zhang, and Jeong-Hyuck Park*

claudio.tessone@uzh.ch park@sogang.ac.kr

Bayesian Consistency of Geometric Distribution Across 63 UTXOs over 72 Monthly Snapshots

Figures SI 1–SI 63 display the Bayesian analysis of 63 representative UTXO values i, covering the period from January 2018 to December 2023, which corresponds to a total of 72 monthly snapshots. In total, $63 \times 72 = 4{,}536$ monthly observations were analysed.

a. For each i, panels $\mathbf{a}_1$–$\mathbf{a}_4$ illustrate four representative examples of the posterior probability distributions of β_i at different times—June 1st of 2019, 2020, 2021, and 2022—together with their corresponding 1σ credible intervals.

b. Panels $\mathbf{b}_1$–$\mathbf{b}_4$ correspond to those same examples, comparing the smoothed empirical ownership distributions, $\bar{P}_\star(k)$, with the Bayesian-fitted geometric distributions, $P_\beta(k)$, parametrised by the posterior mean values $\langle\beta_i\rangle$. The horizontal axis represents the holding, k, for a given UTXO denomination i.

c. Panel **c** tracks the full monthly evolution of the mean holdings, where $m_i^\star$ denotes the empirical mean obtained from the data and $m_i(\beta_i)$ represents the theoretical mean derived from the fitted geometric distribution.

d. Panel **d** shows the Jensen–Shannon divergence, D_{JS}, quantifying the statistical distance between the smoothed empirical and fitted distributions, which remains consistently stable (around 0.04) throughout the six-year period.

These results collectively demonstrate the robustness of the geometric distribution as a consistent descriptor of UTXO holdings over time, supporting the bosonic interpretation of Bitcoin's UTXO statistics introduced in the main text.

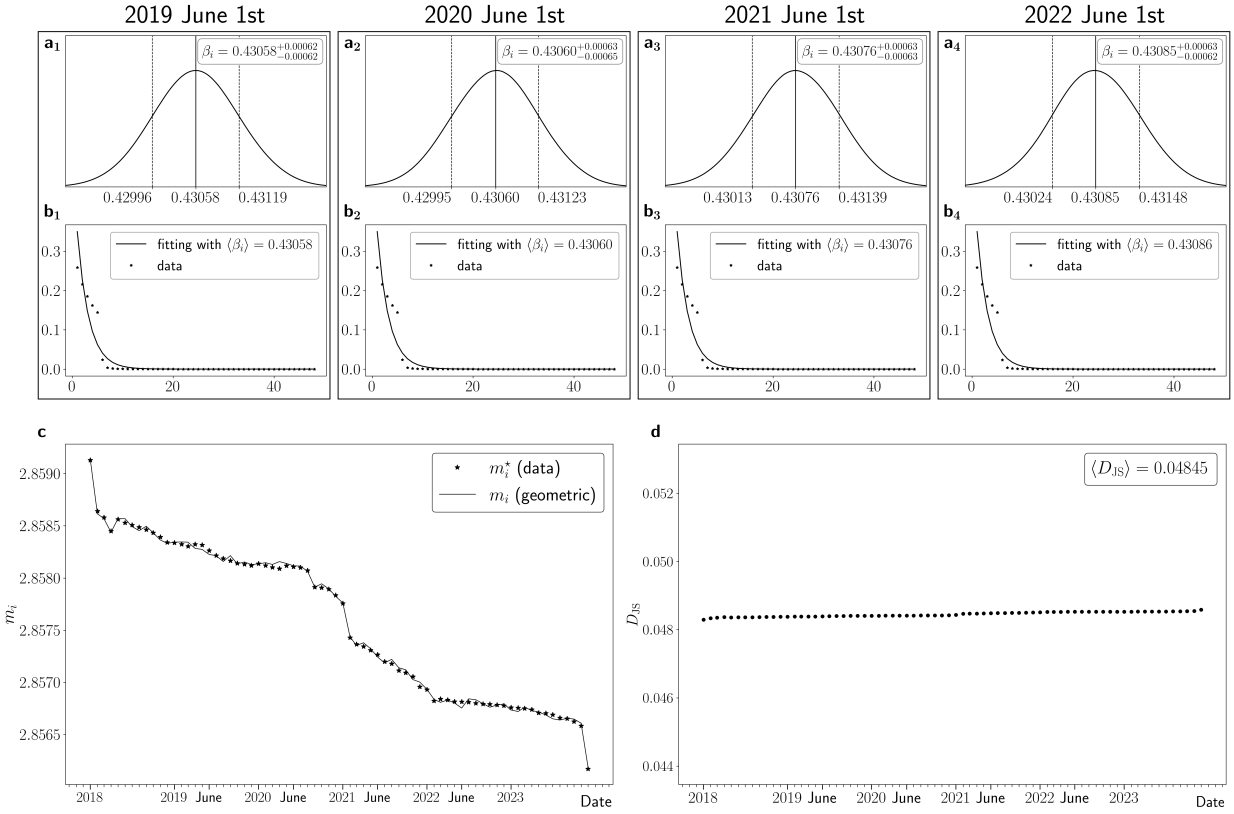

Figure SI 1: $i = 1$

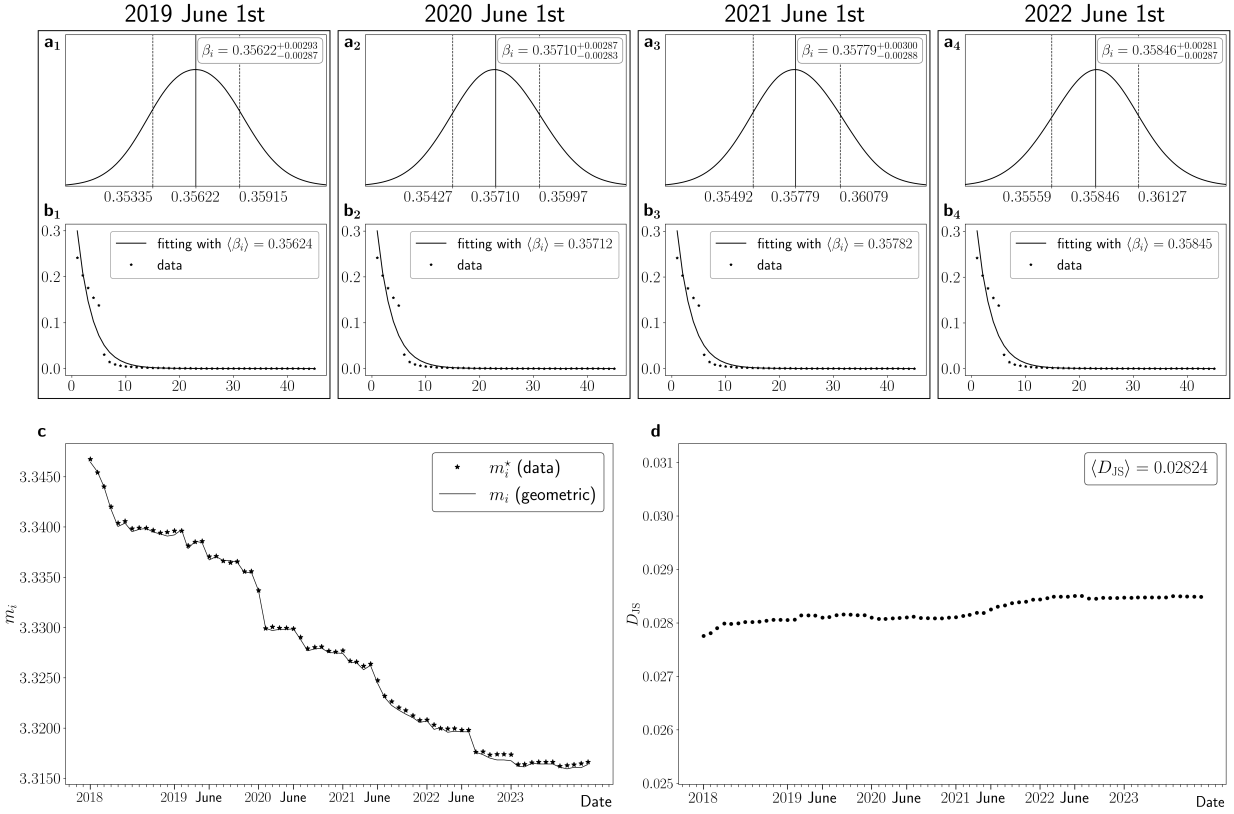

Figure SI 2: $i = 4$

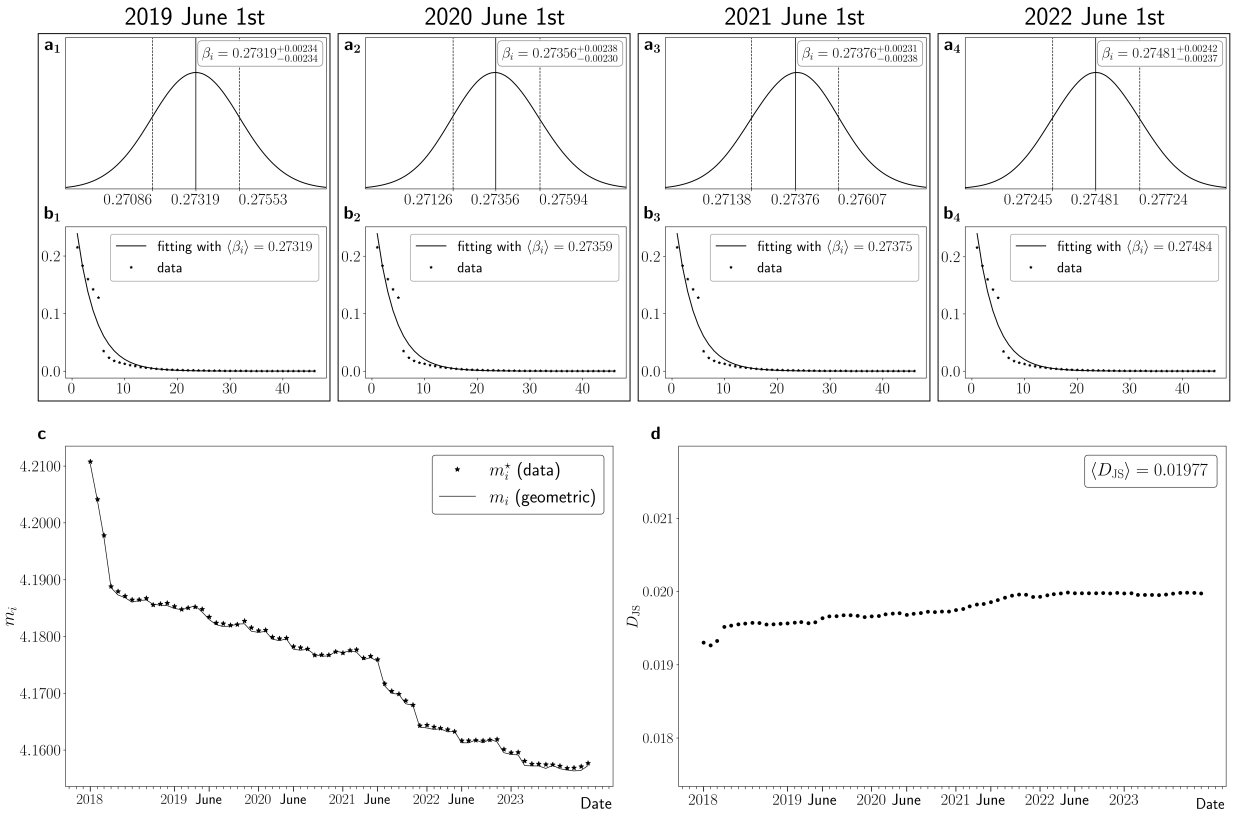

Figure SI 3: $i = 10$

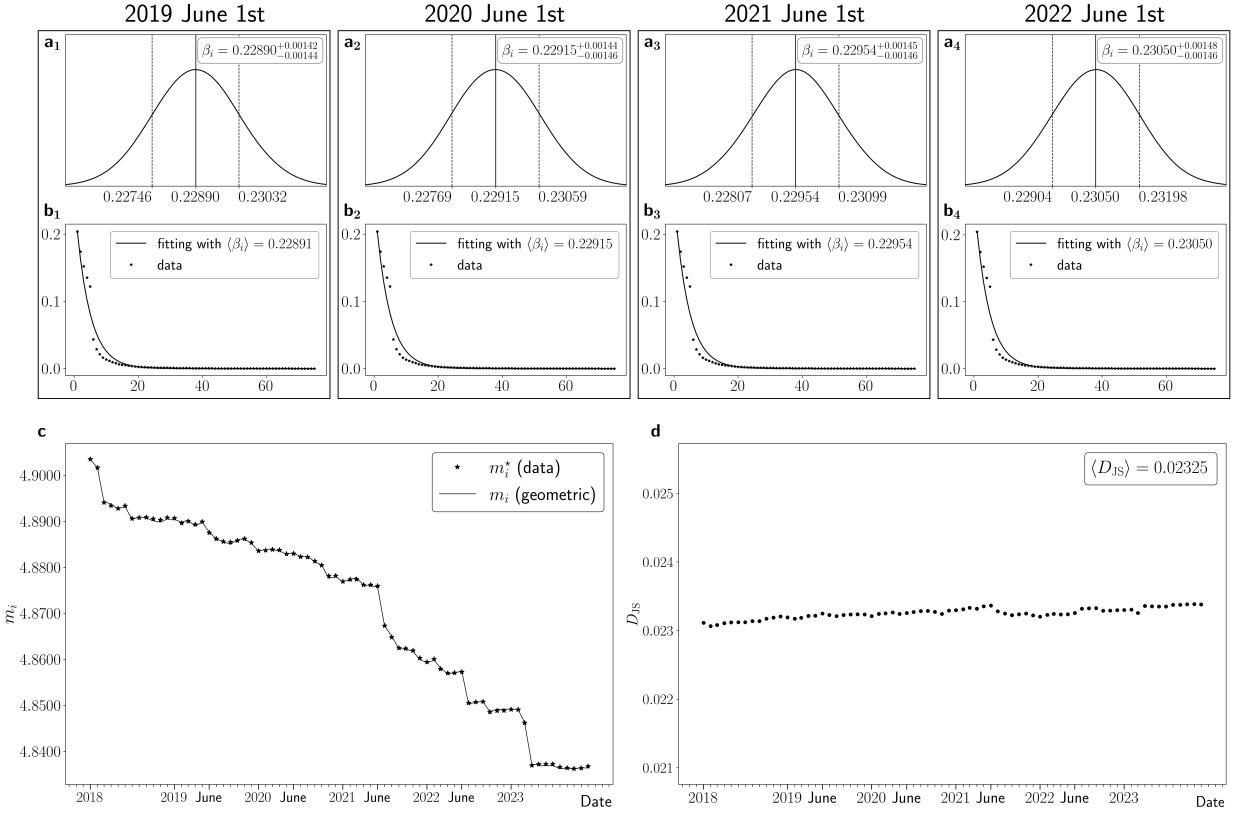

Figure SI 4: $i = 100$

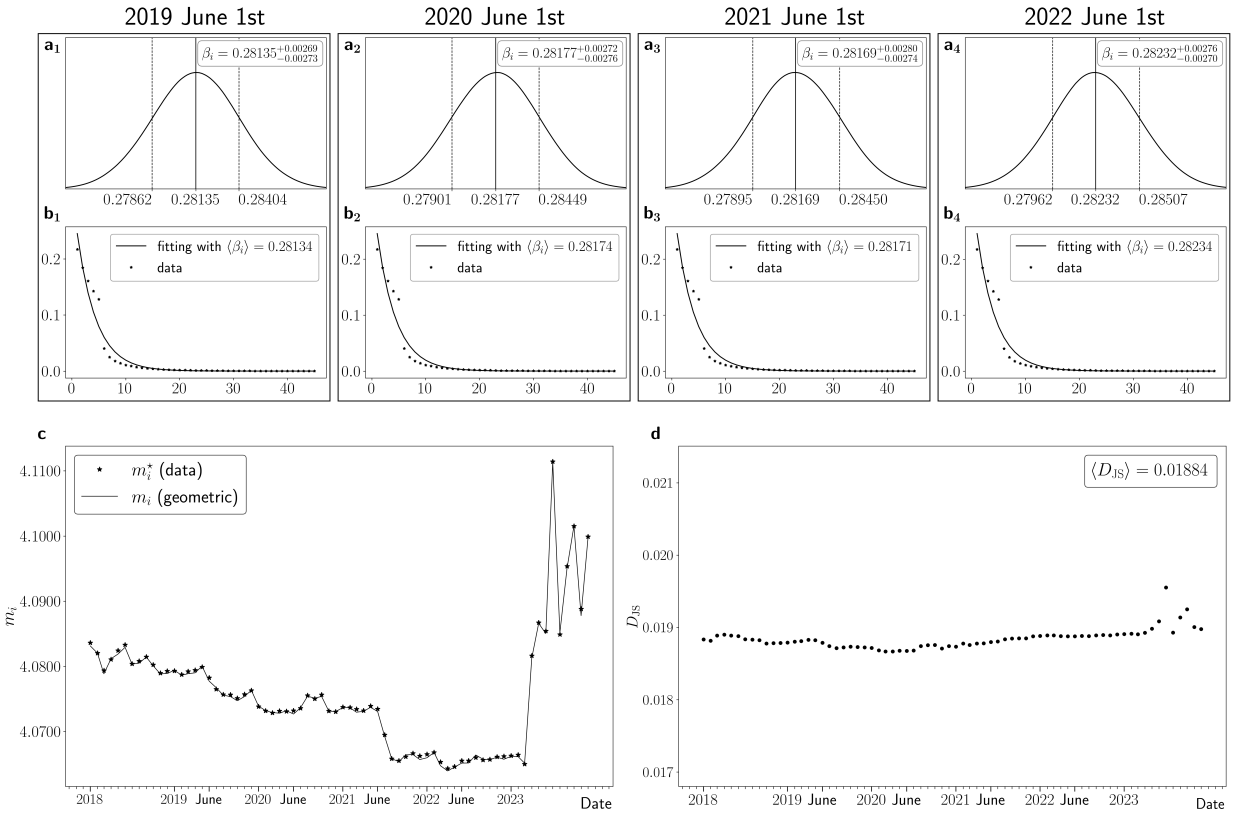

Figure SI 5: $i = 200$

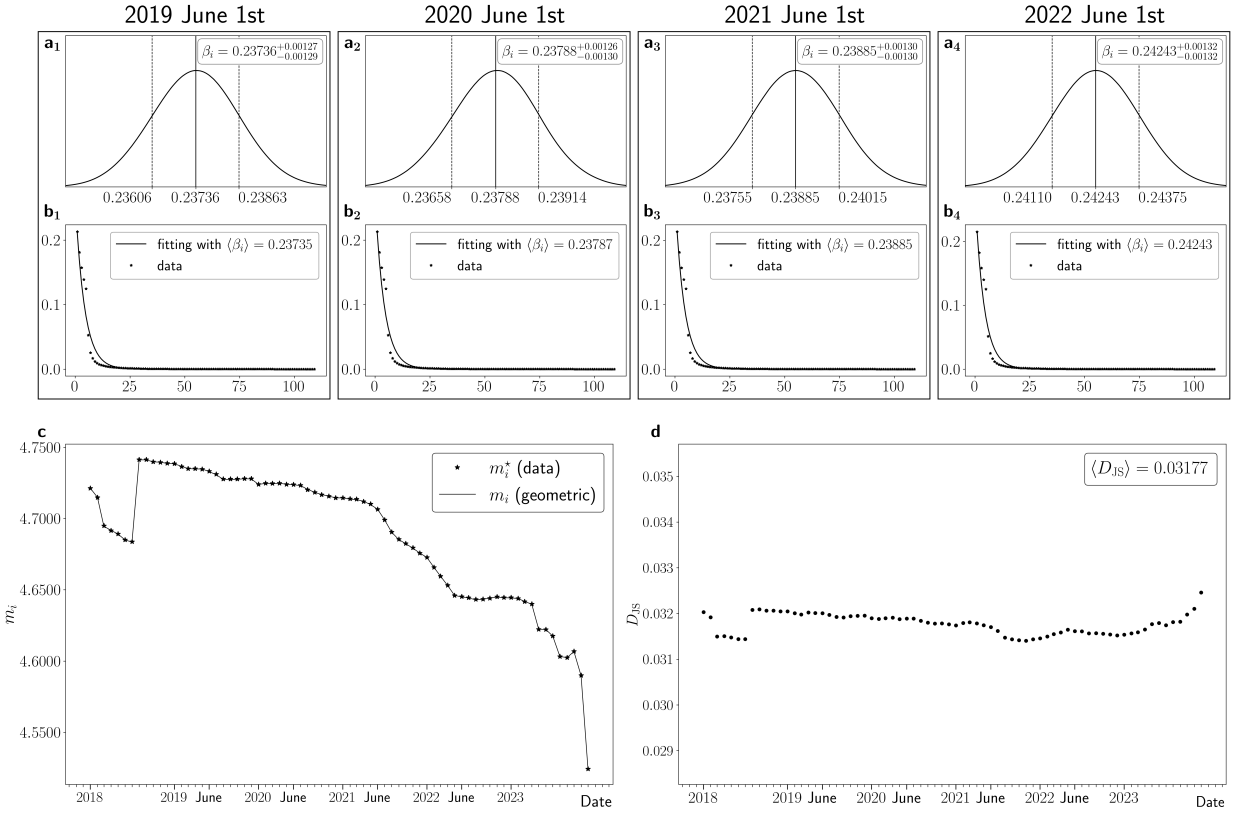

Figure SI 6: $i = 400$

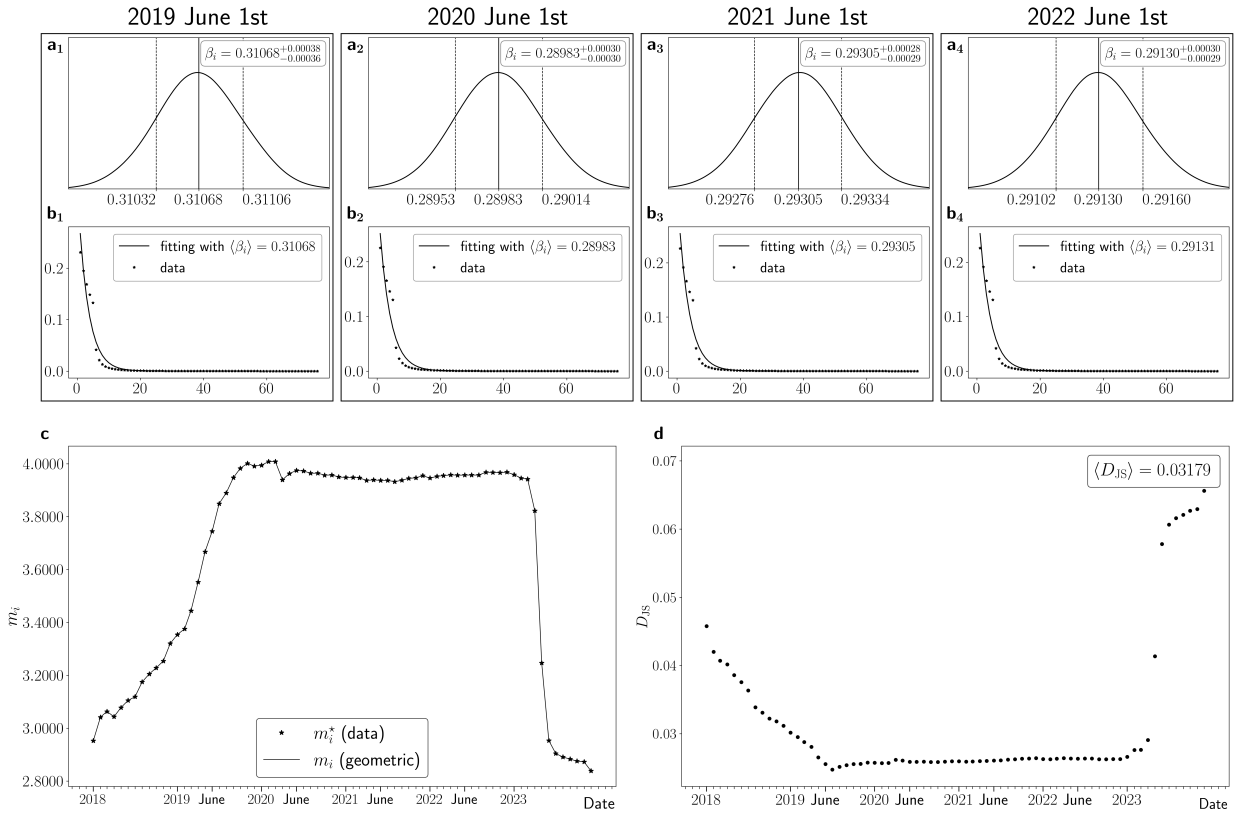

Figure SI 7: $i = 546$

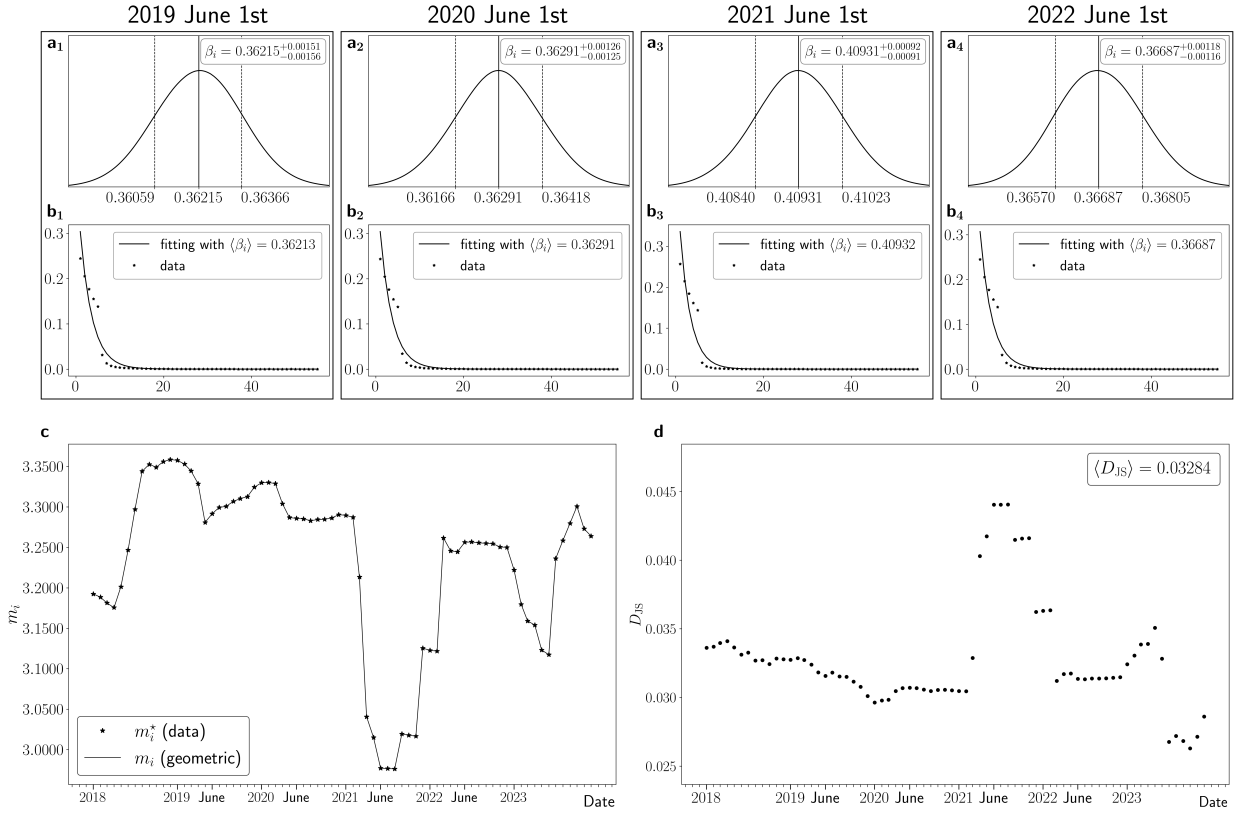

Figure SI 8: $i = 600$

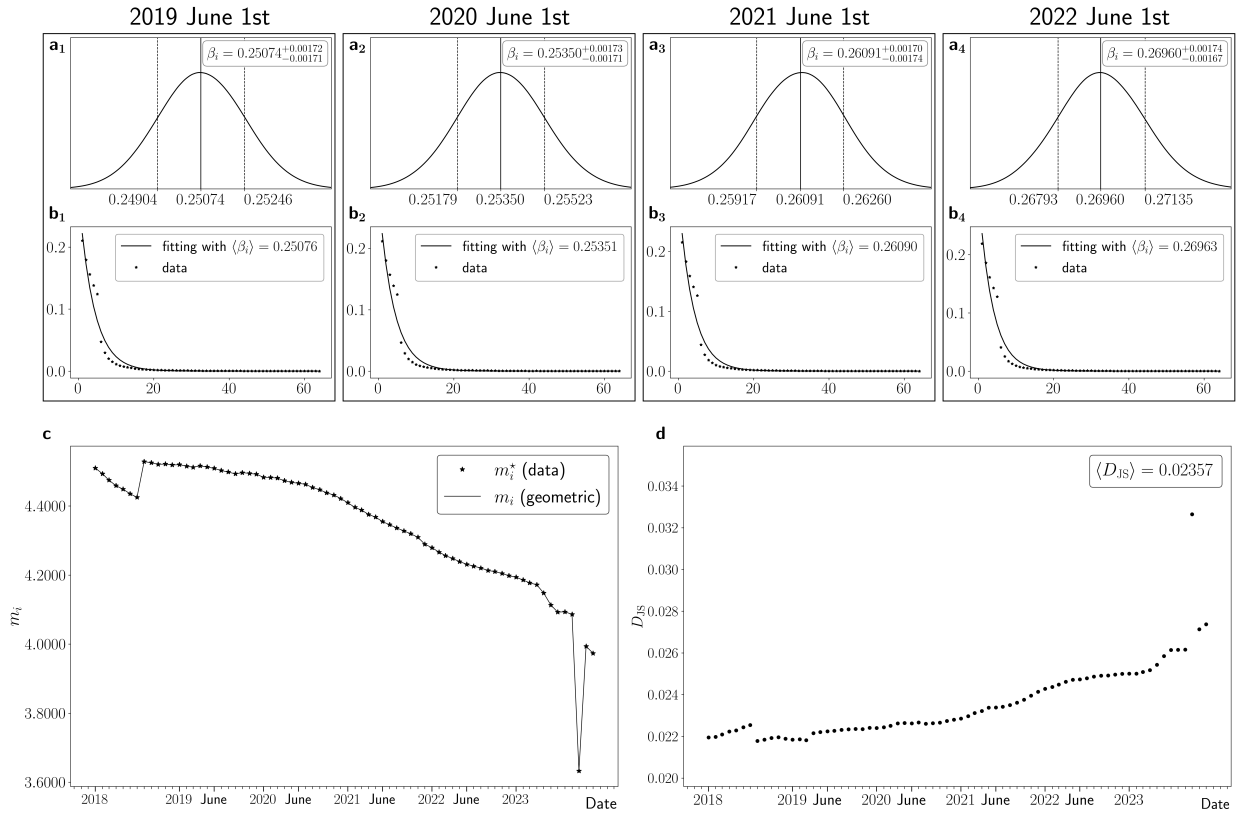

Figure SI 9: $i = 800$

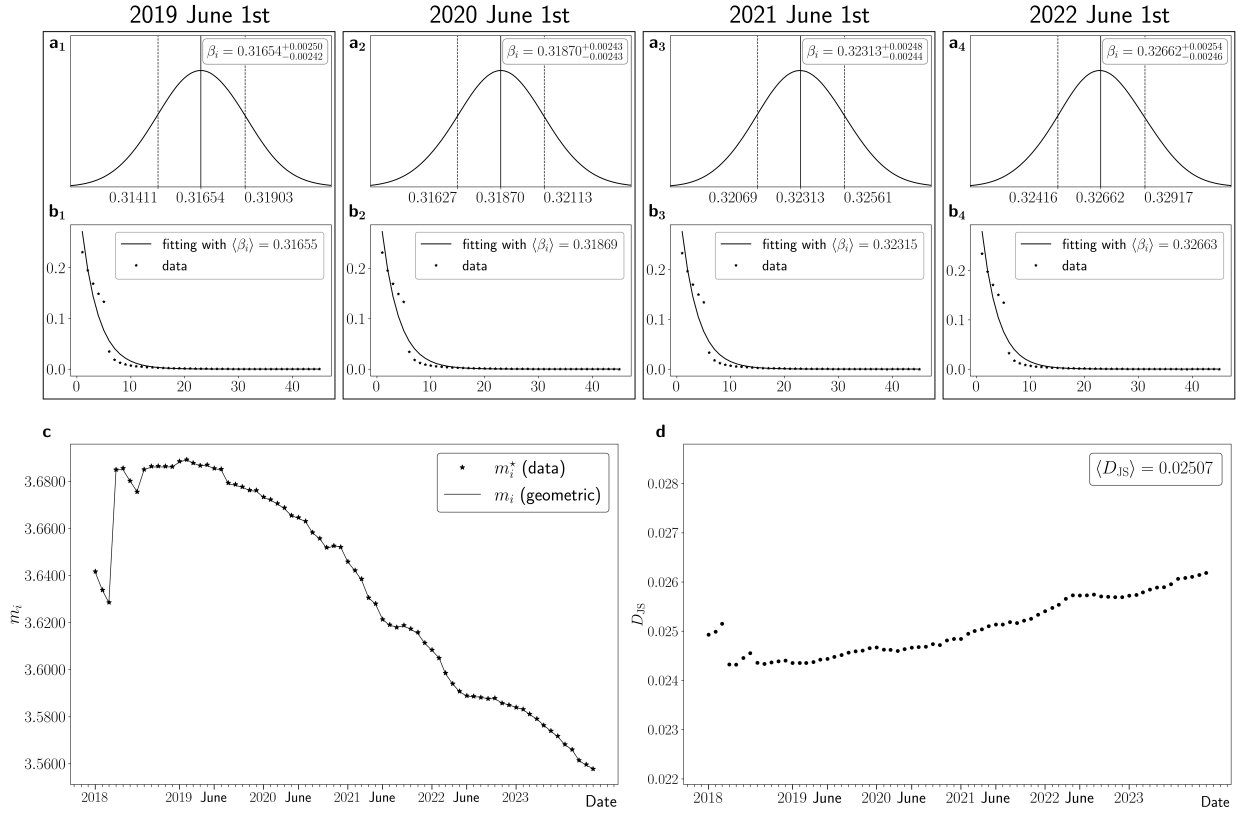

Figure SI 10: $i = 880$

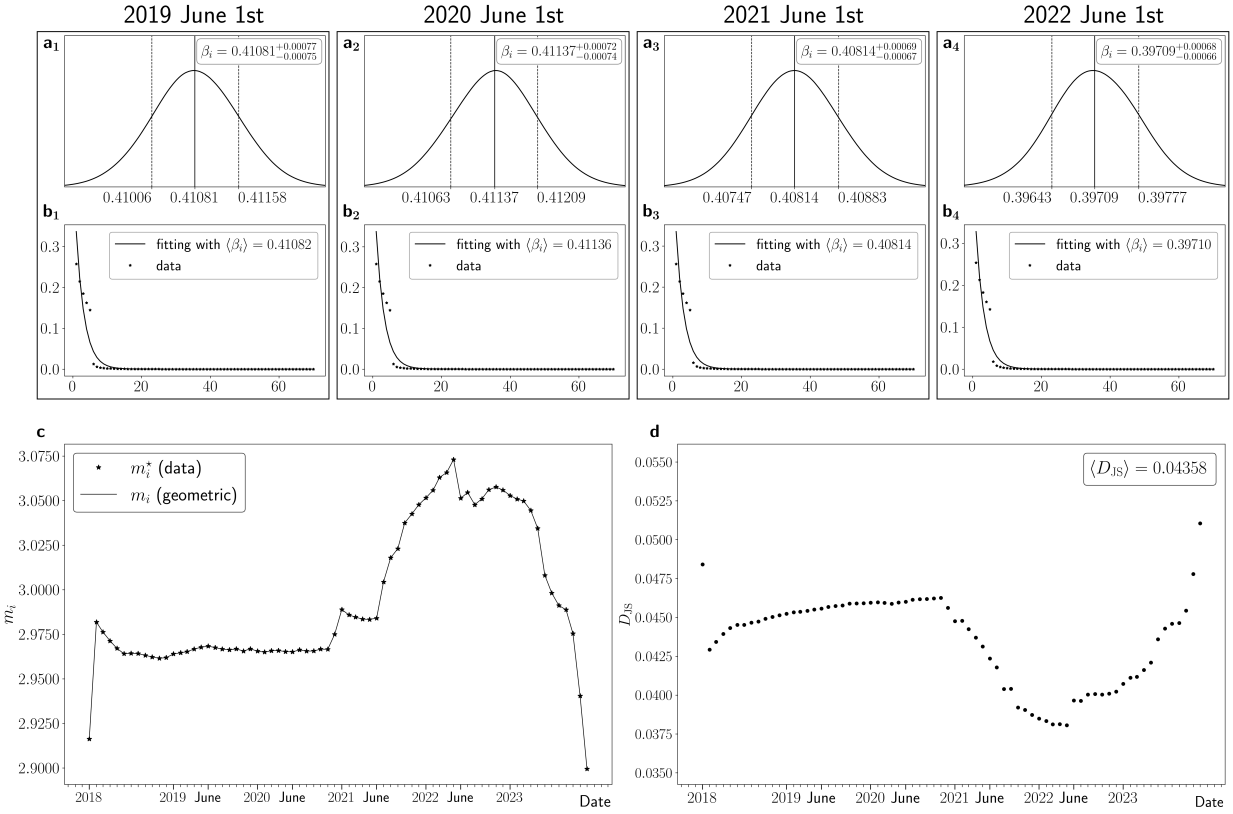

Figure SI 11: $i = 1000$

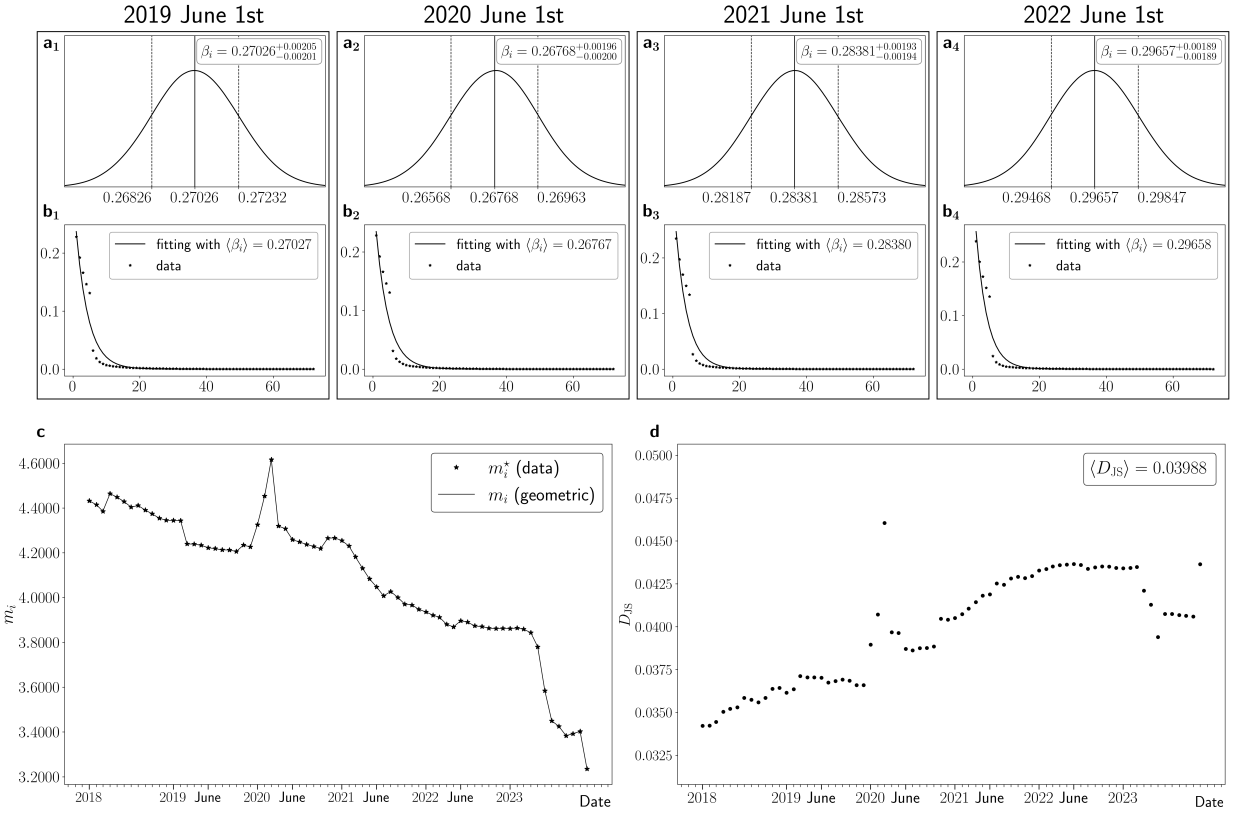

Figure SI 12: $i = 1200$

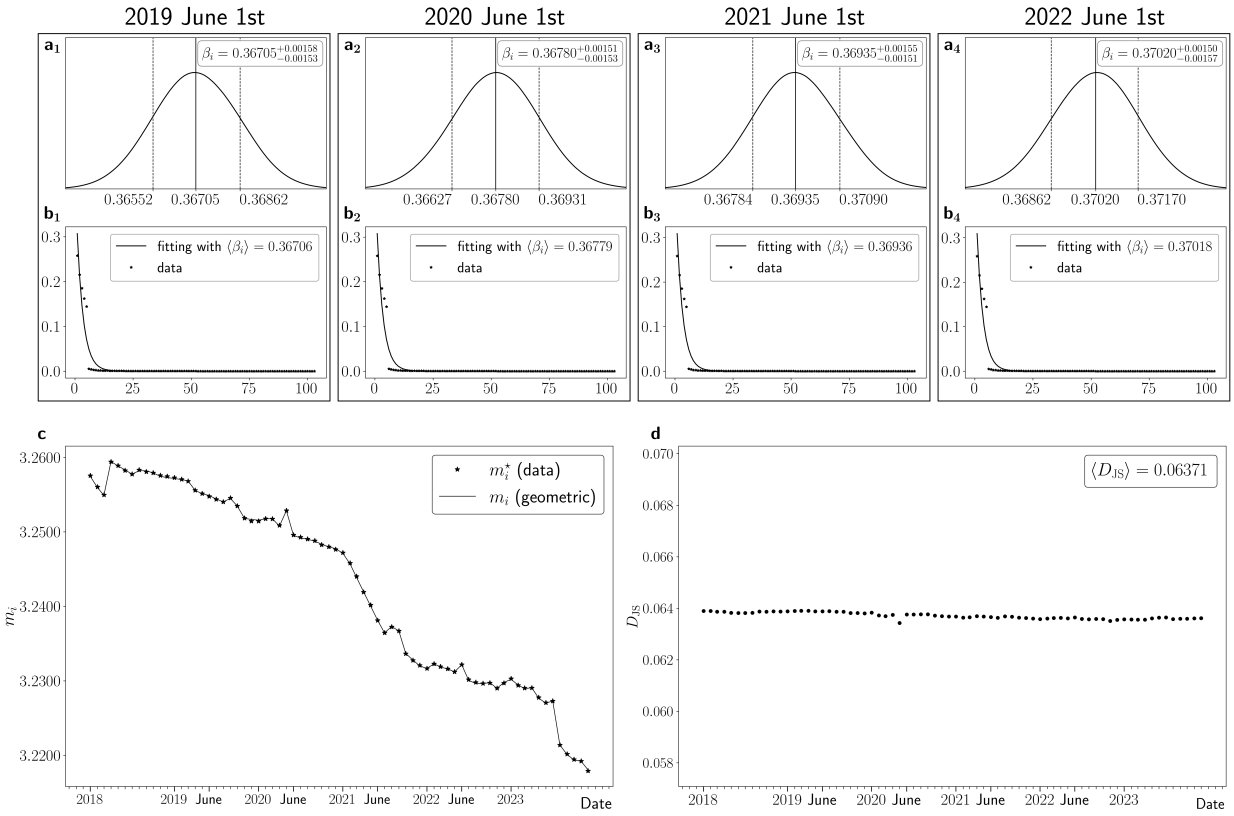

Figure SI 13: $i = 1250$

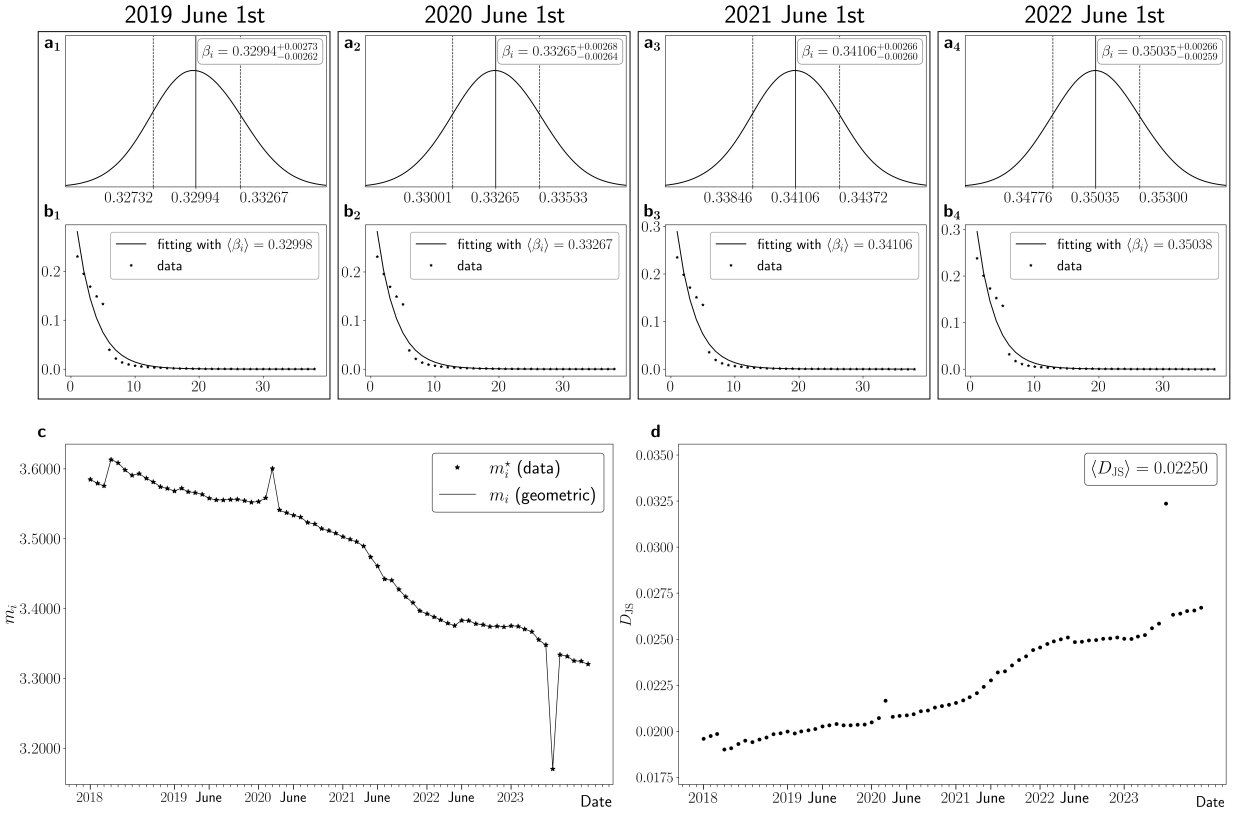

Figure SI 14: $i = 1600$

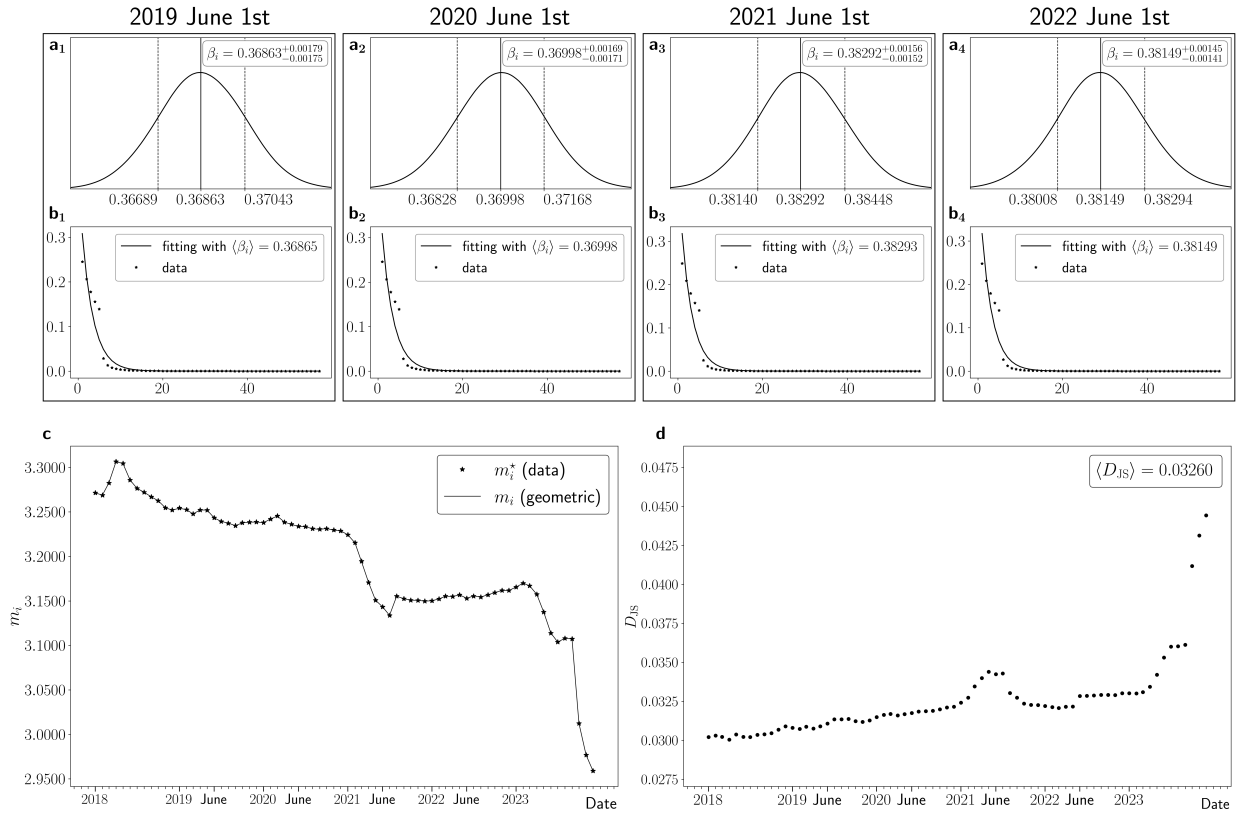

Figure SI 15: $i = 2000$

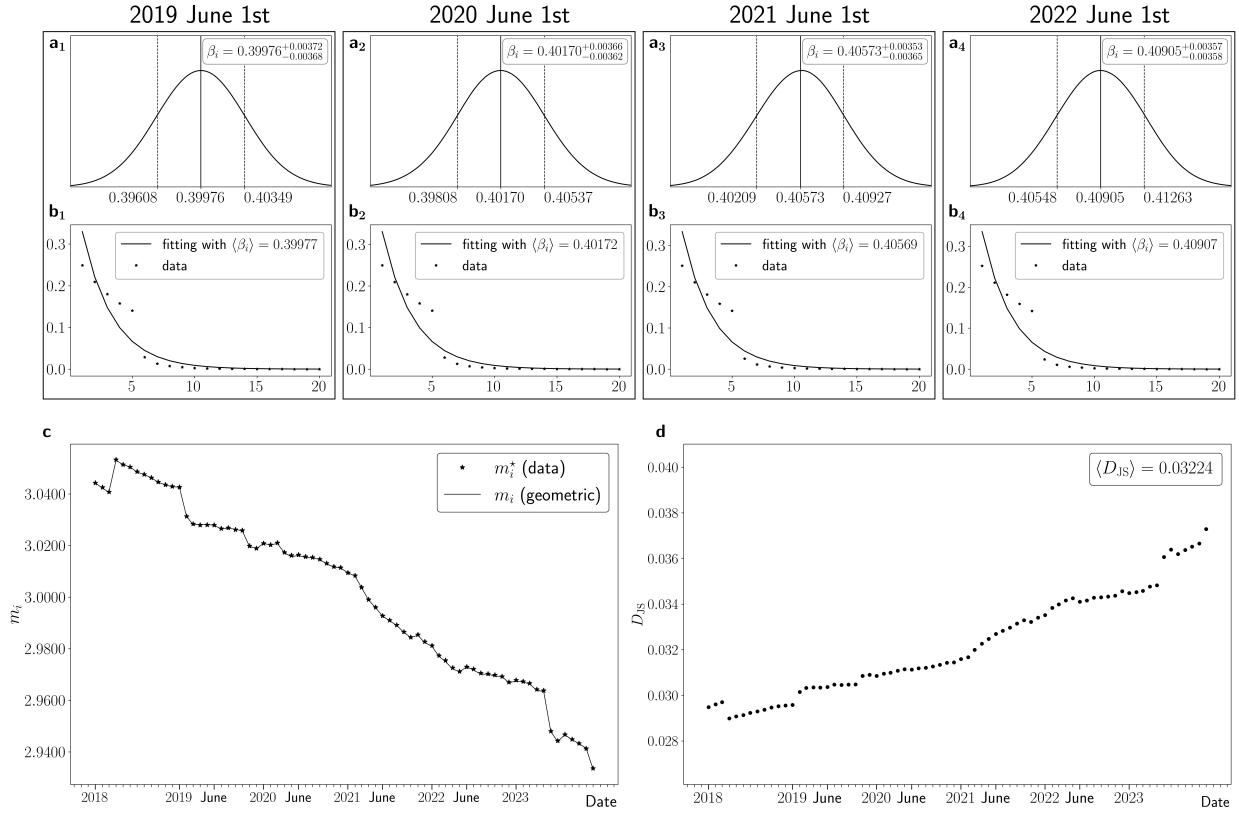

Figure SI 16: $i = 2400$

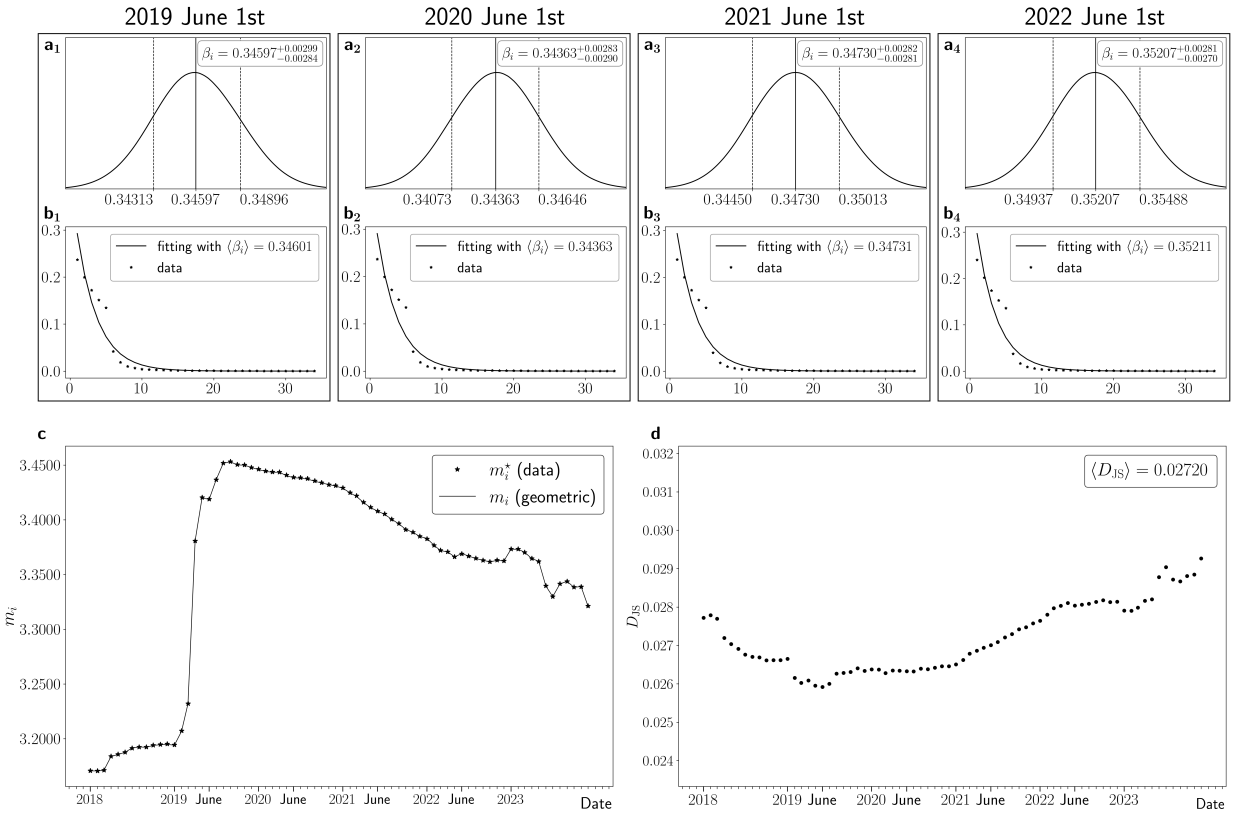

Figure SI 17: $i = 2700$

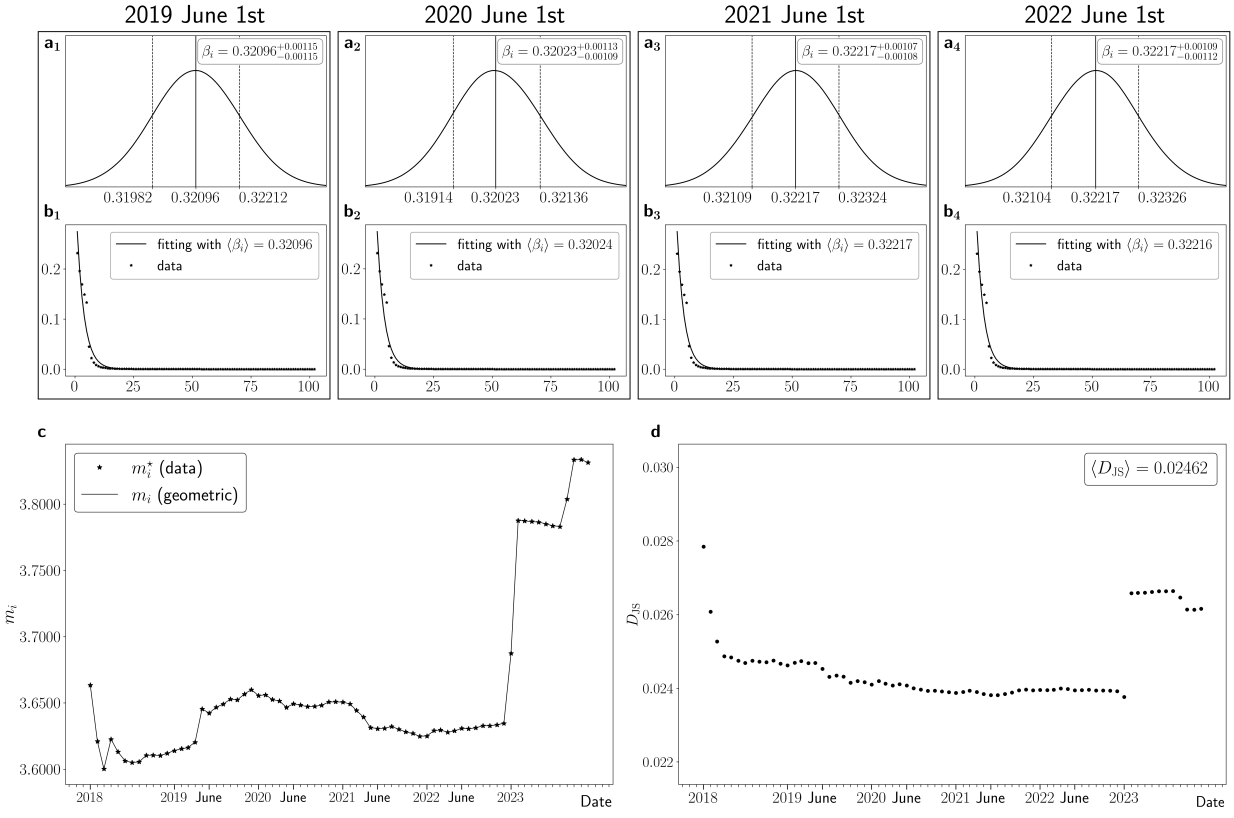

Figure SI 18: $i = 2730$

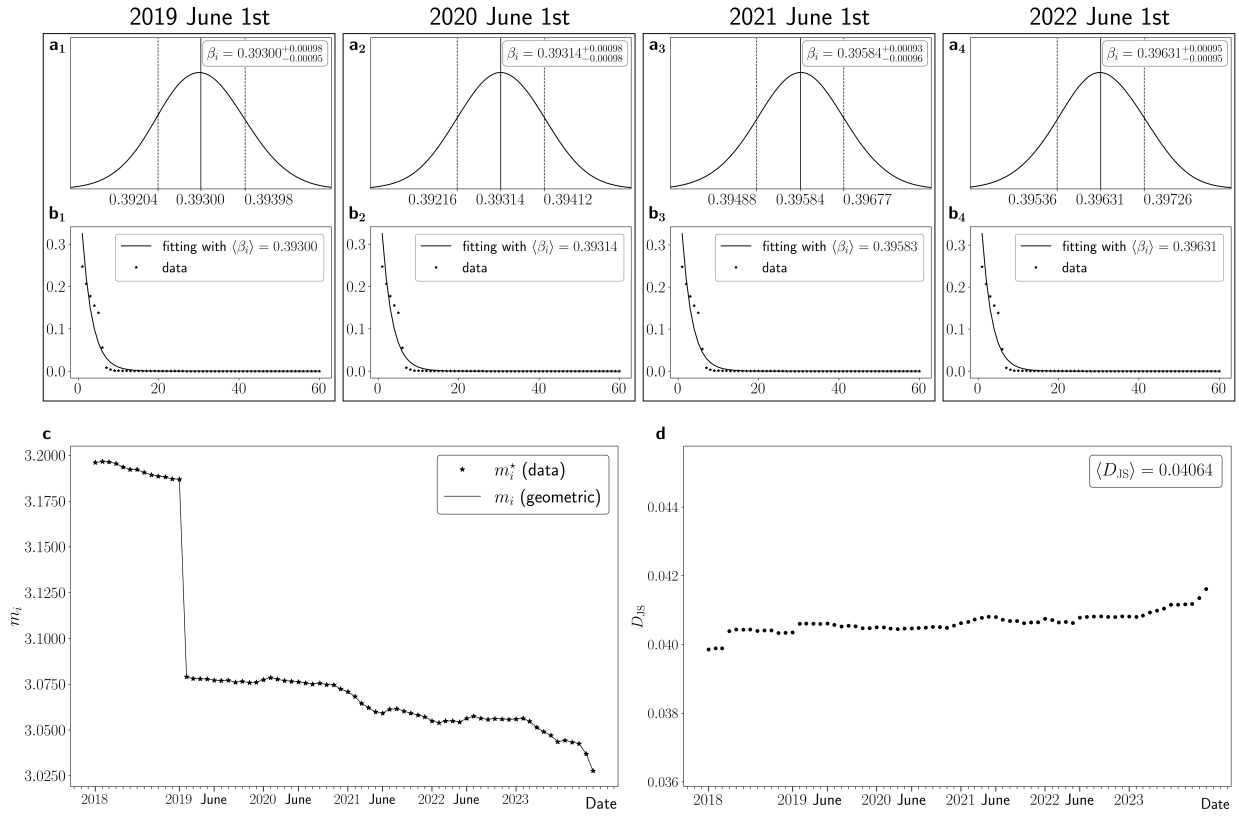

Figure SI 19: $i = 3000$

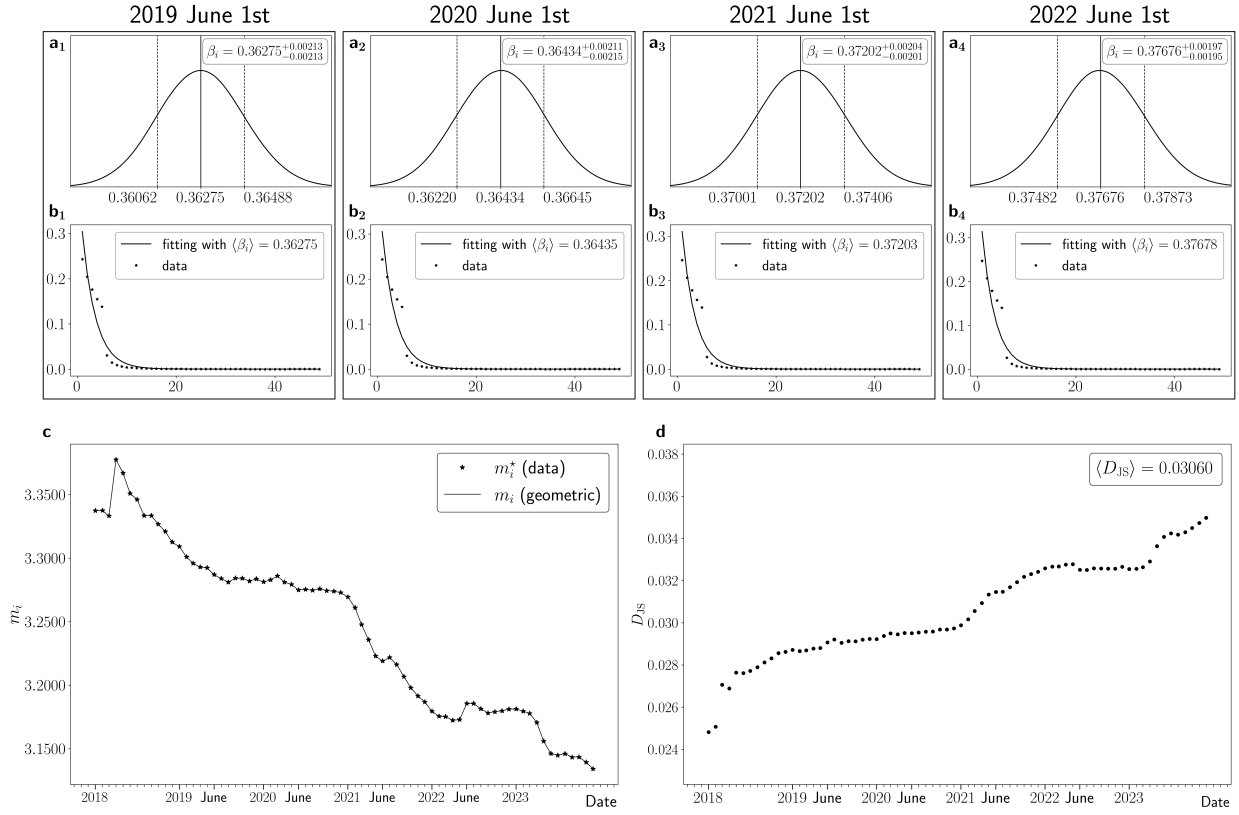

Figure SI 20: $i = 4000$

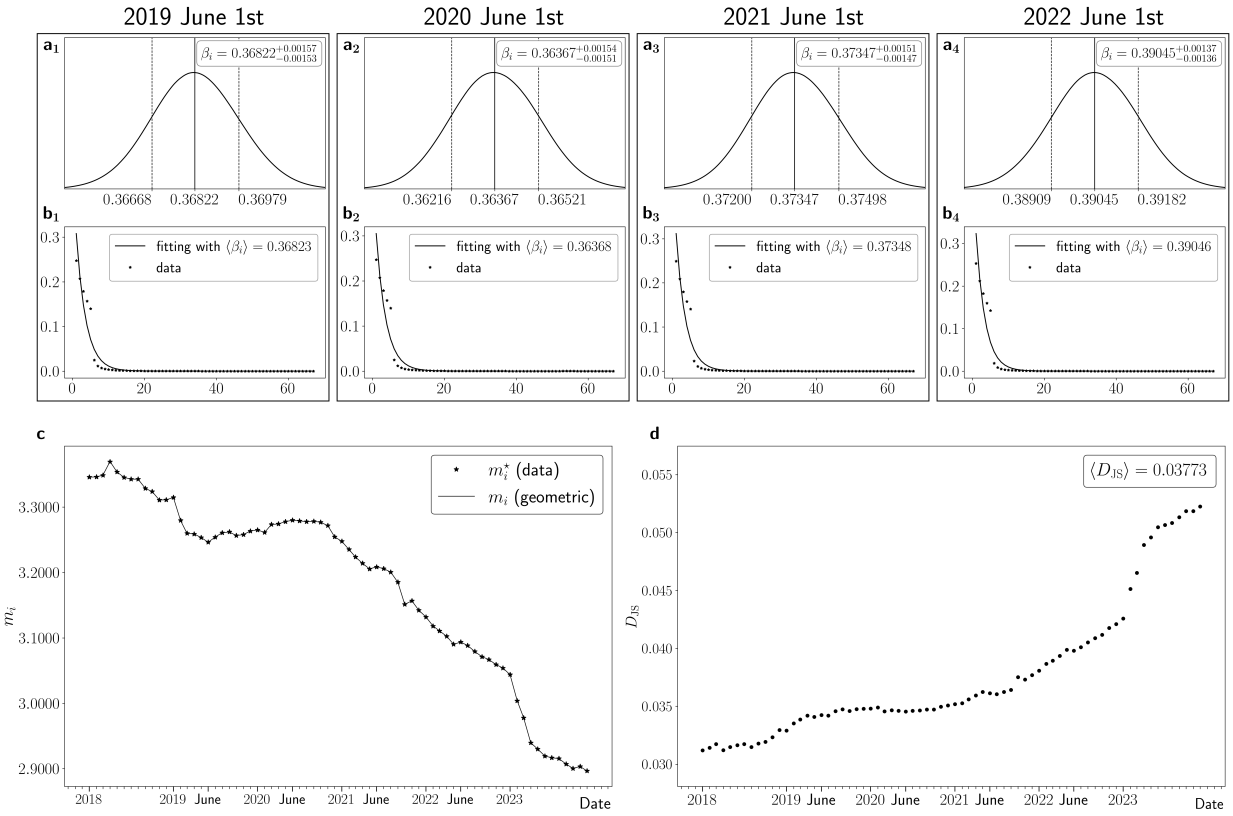

Figure SI 21: $i = 5000$

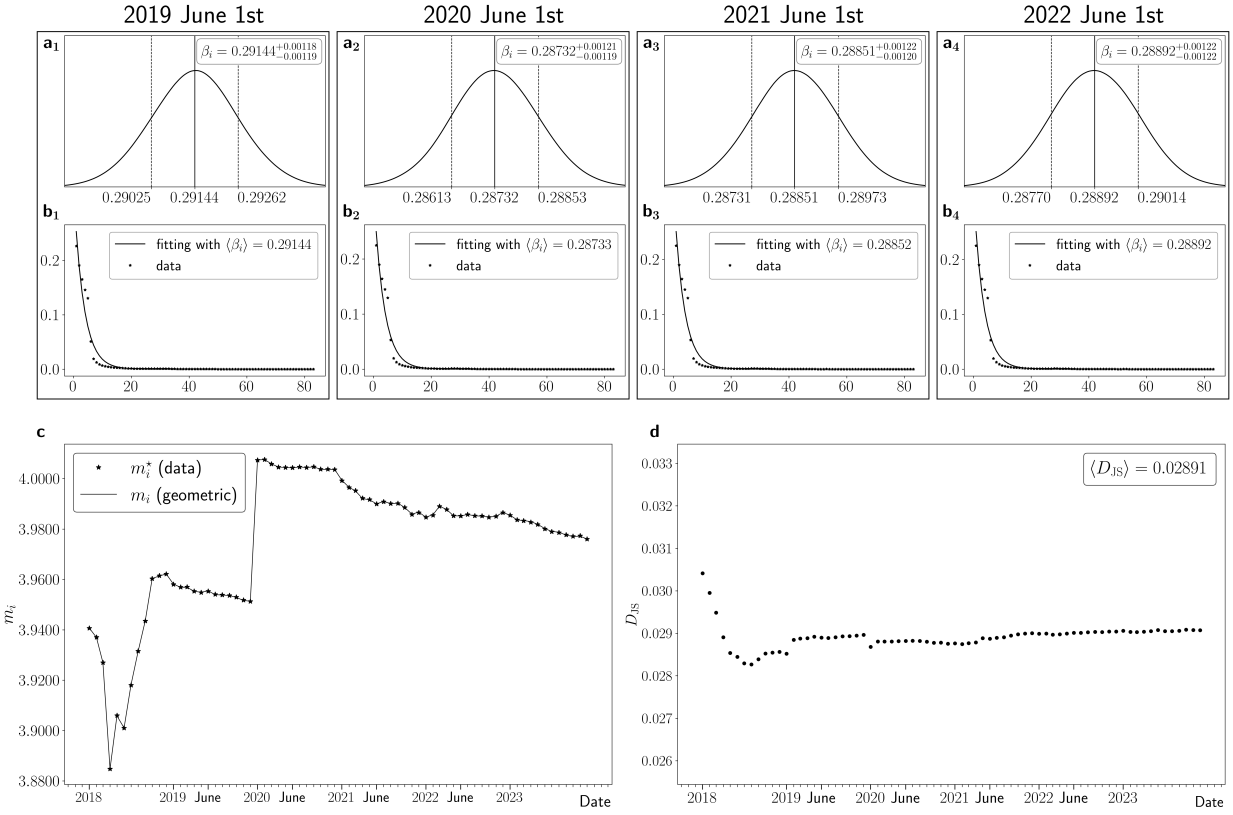

Figure SI 22: $i = 5430$

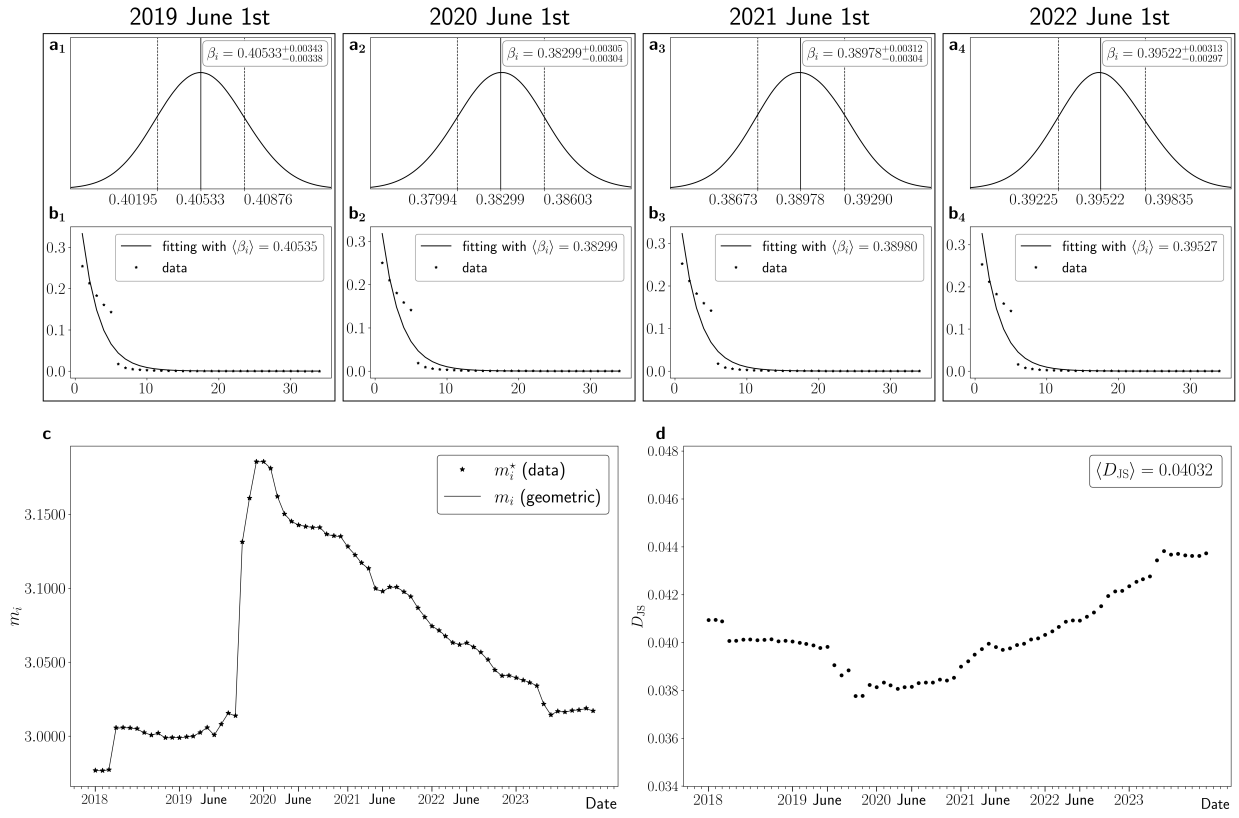

Figure SI 23: $i = 5460$

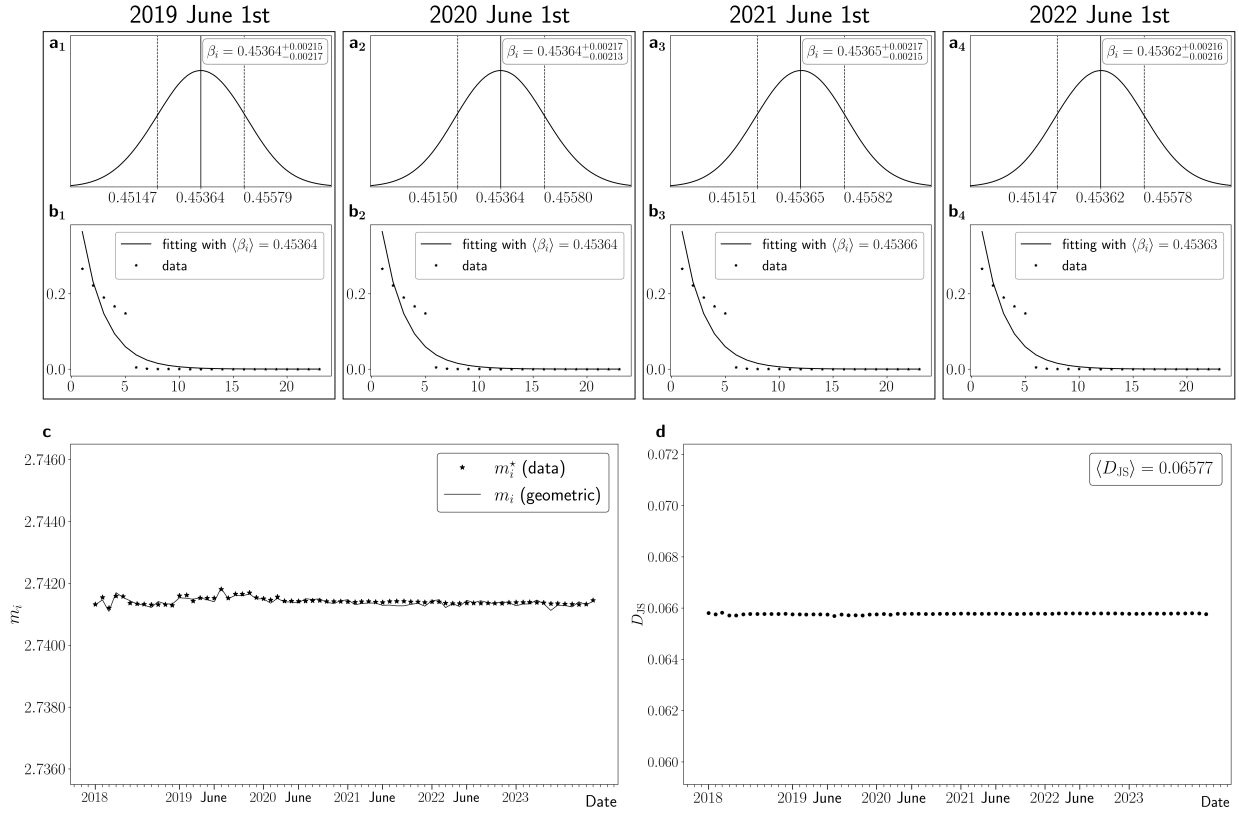

Figure SI 24: $i = 5480$

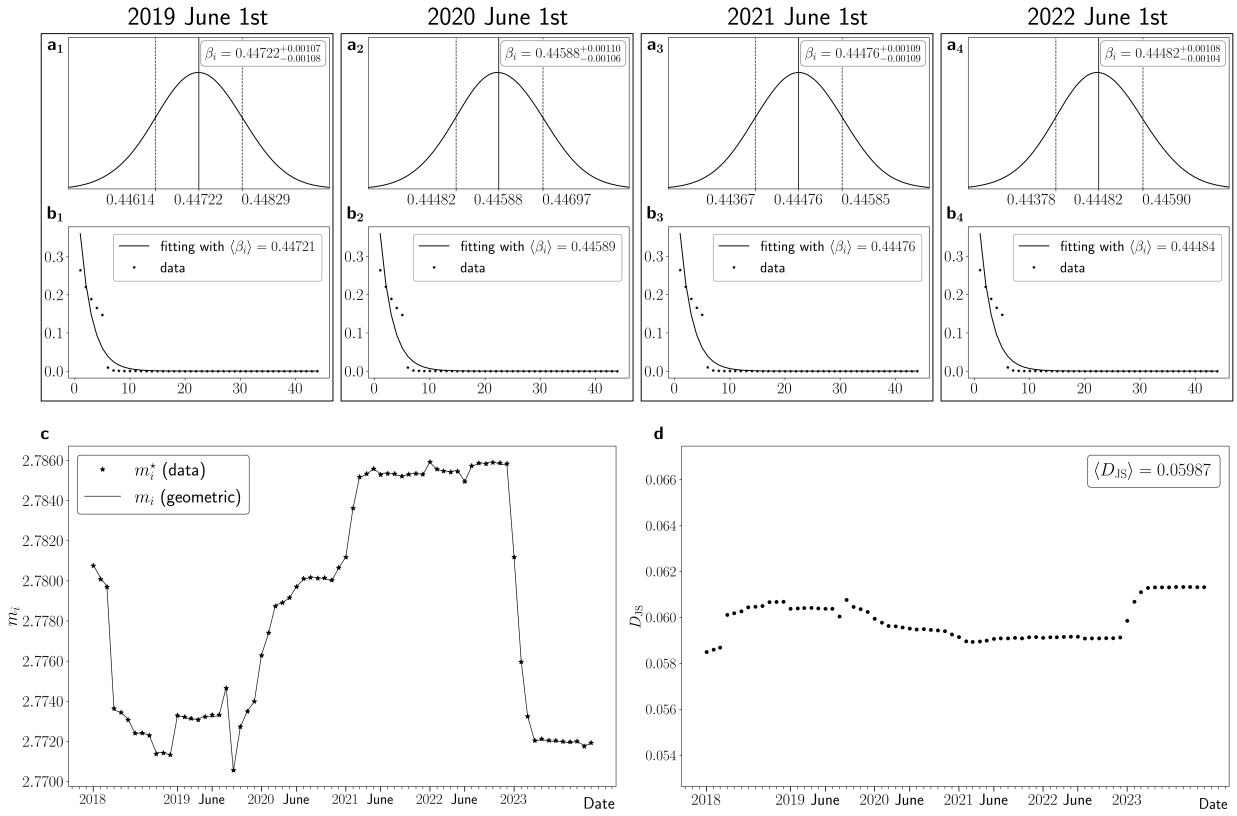

Figure SI 25: $i = 5500$

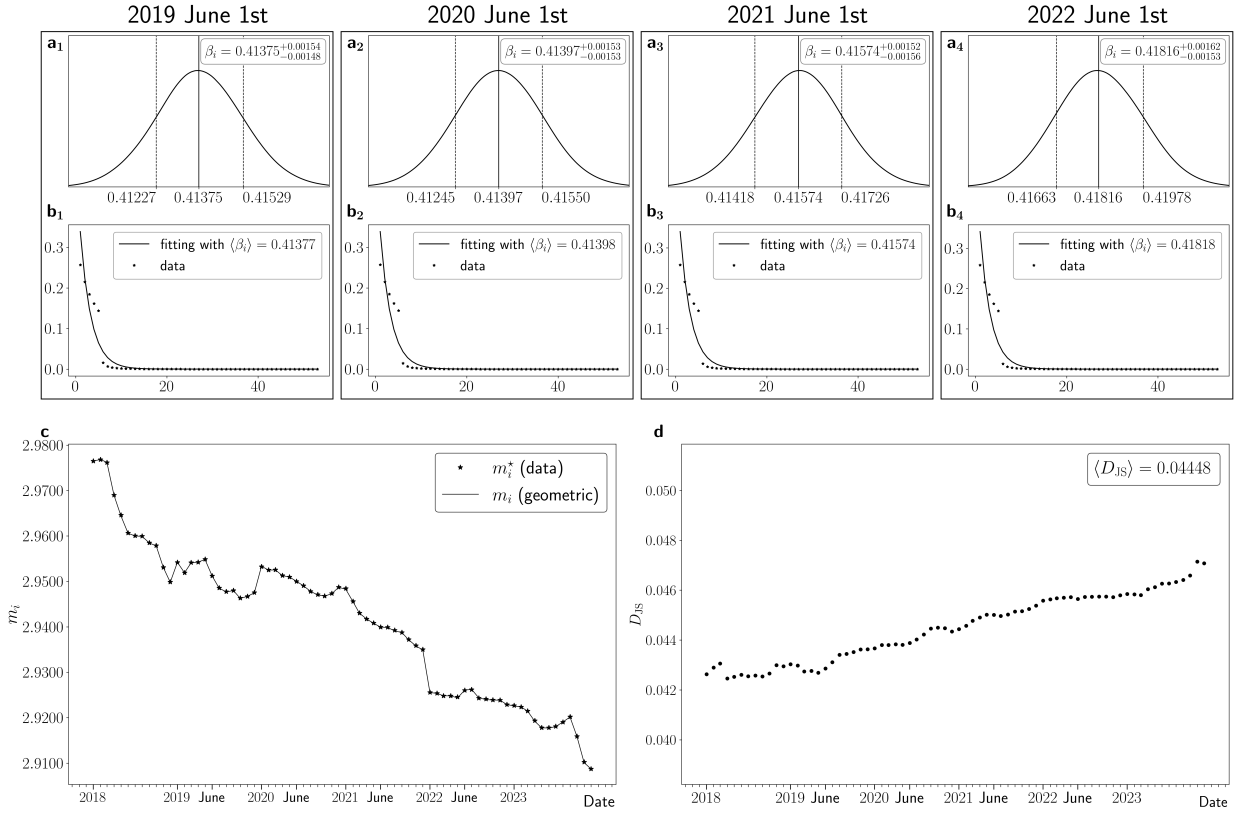

Figure SI 26: $i = 6000$

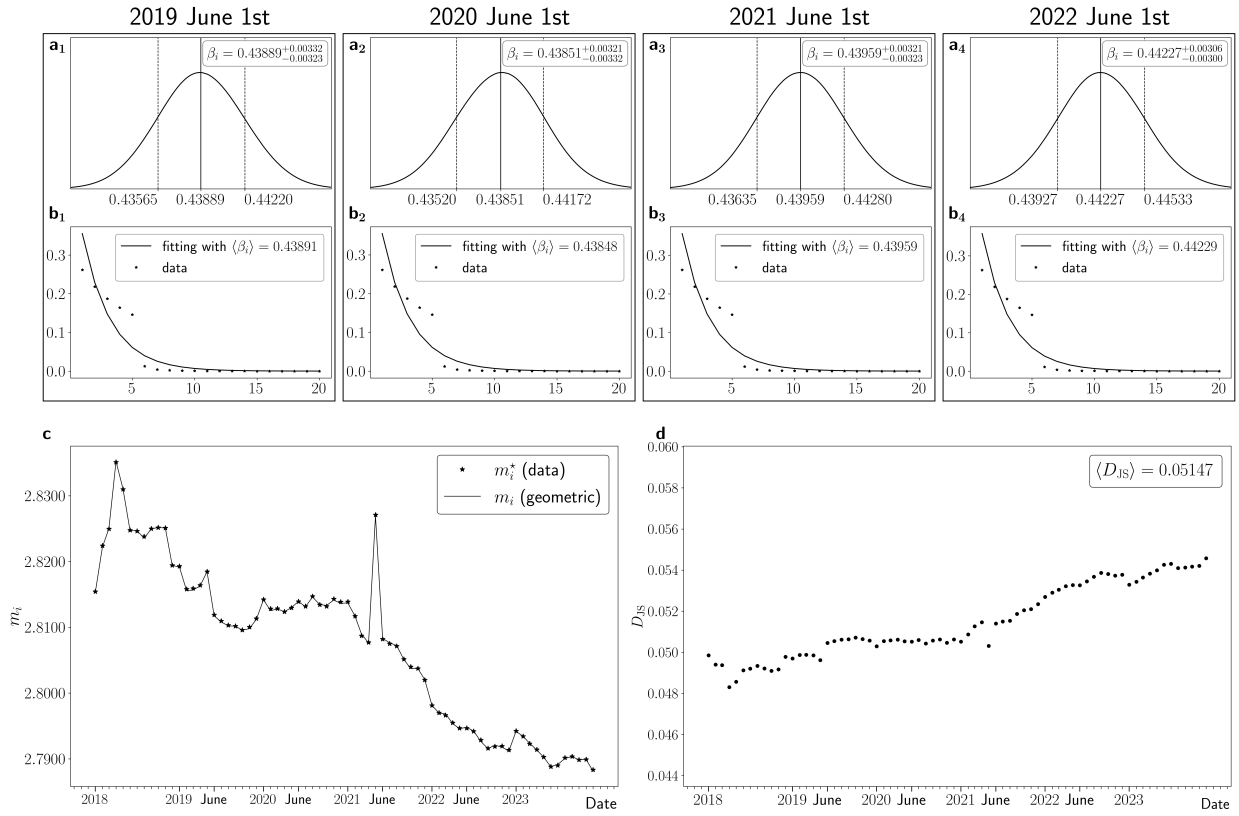

Figure SI 27: $i = 7000$

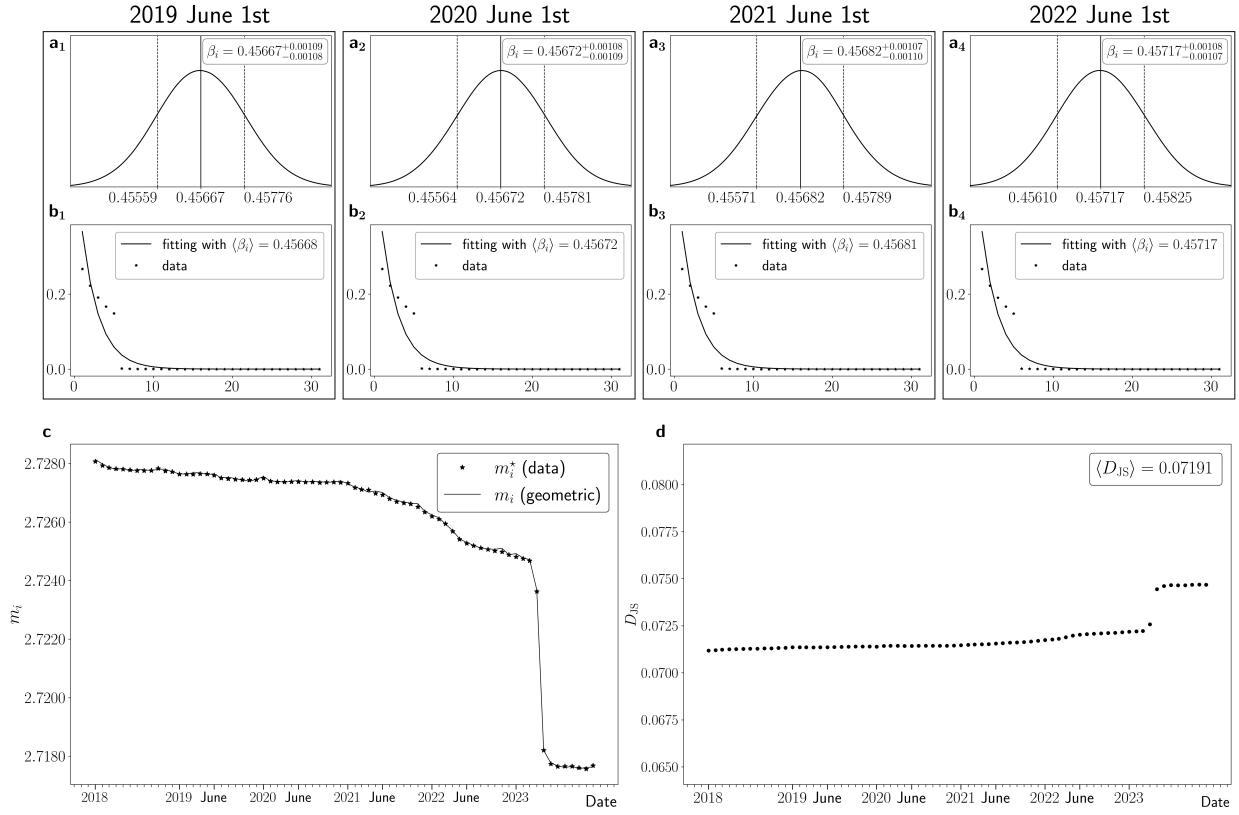

Figure SI 28: $i = 7800$

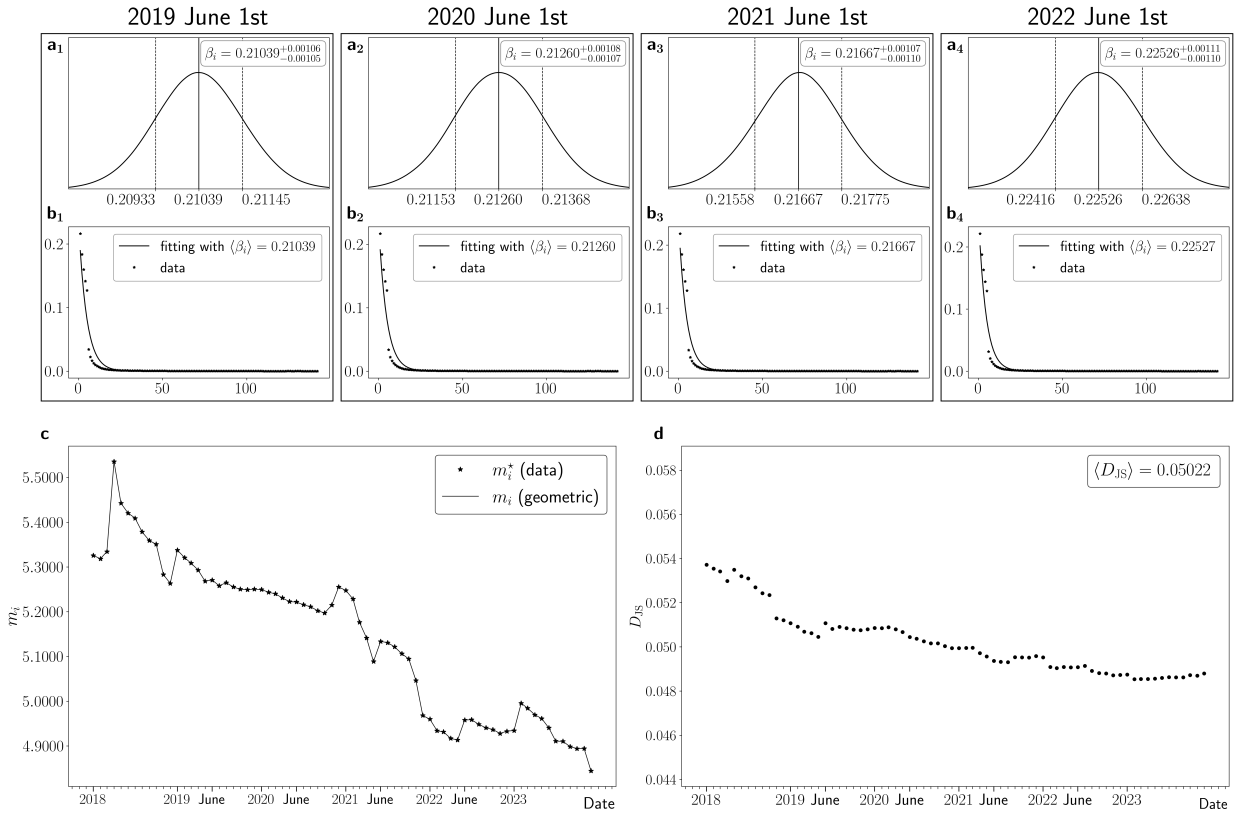

Figure SI 29: $i = 8000$

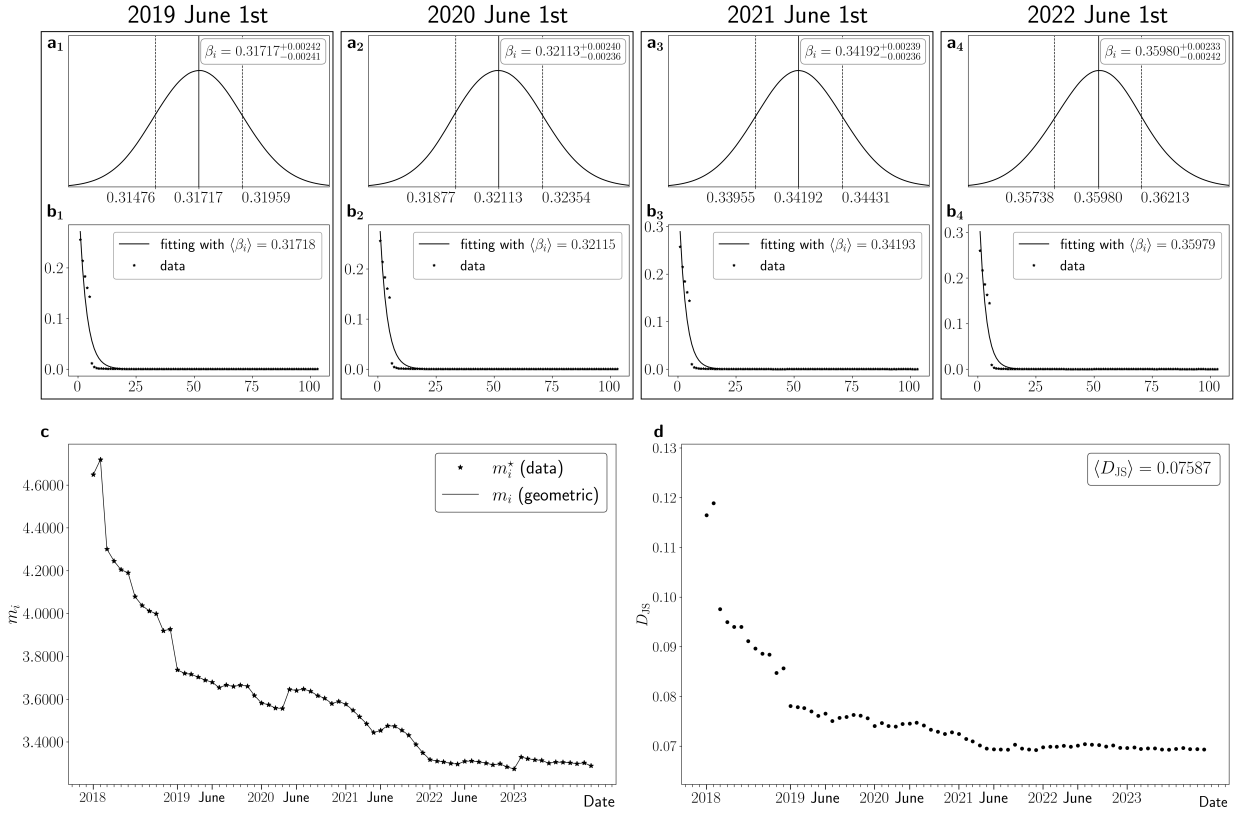

Figure SI 30: $i = 9000$

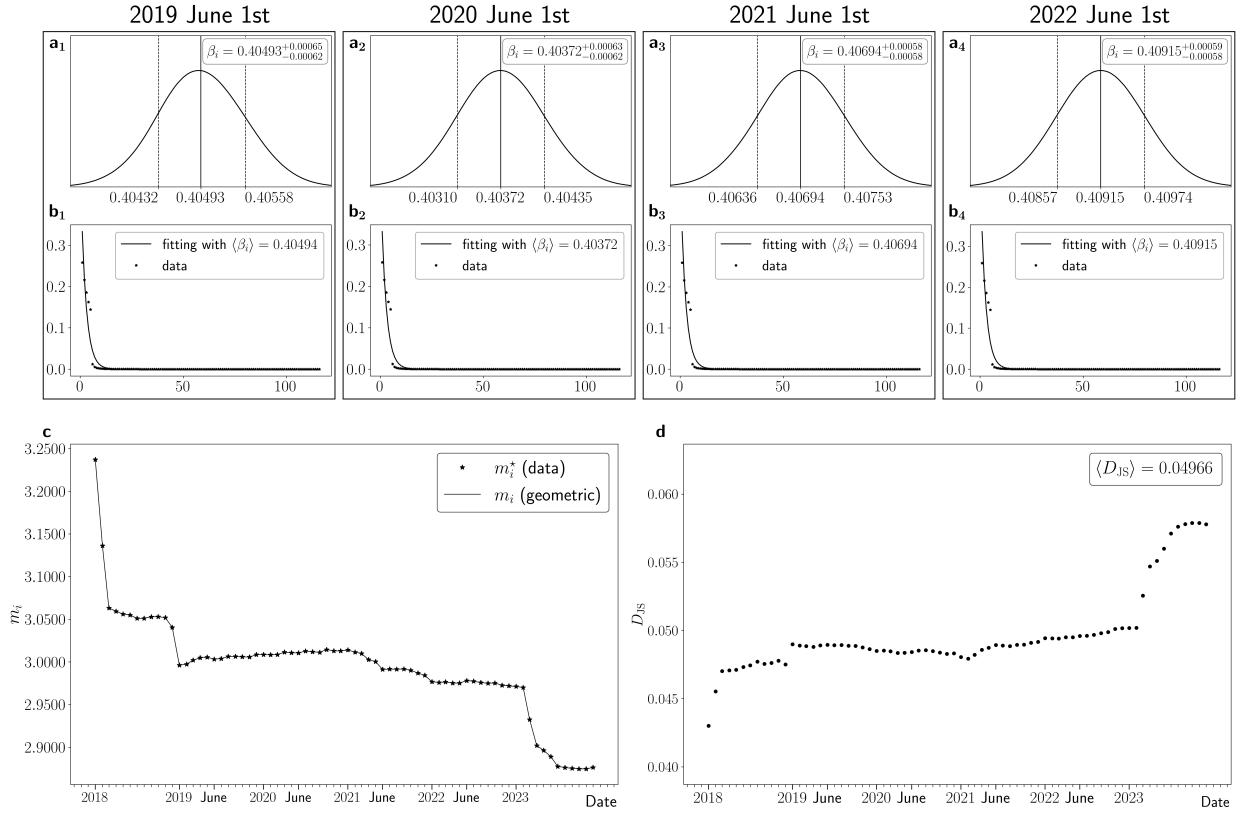

Figure SI 31: $i = 10000$

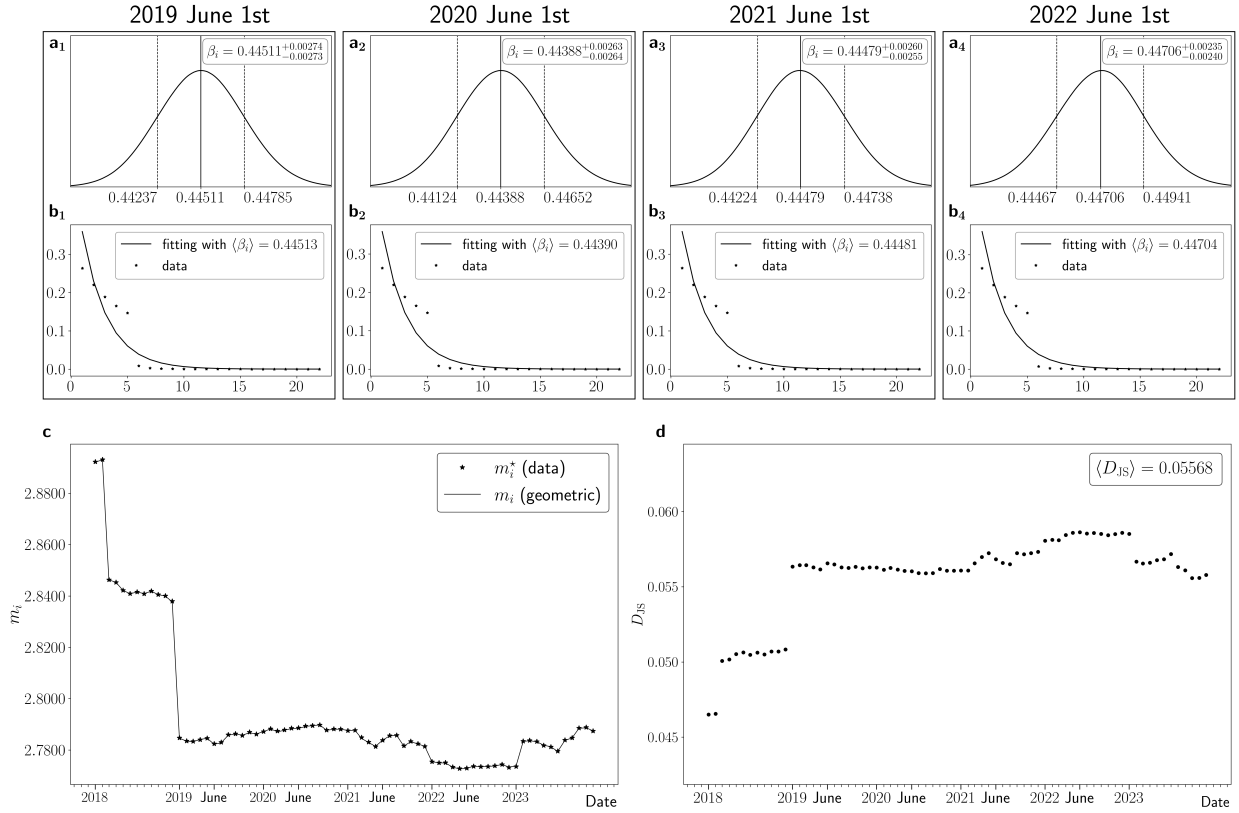

Figure SI 32: $i = 11000$

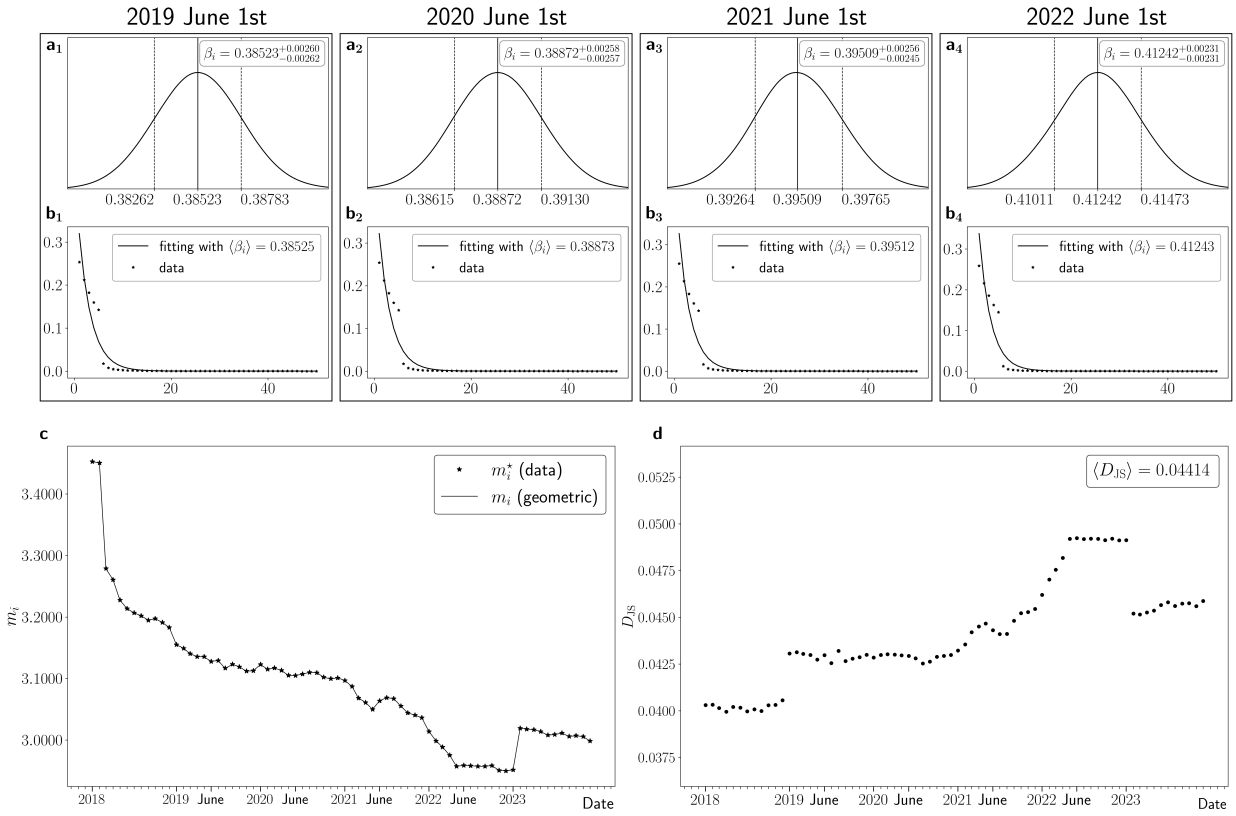

Figure SI 33: $i = 12000$

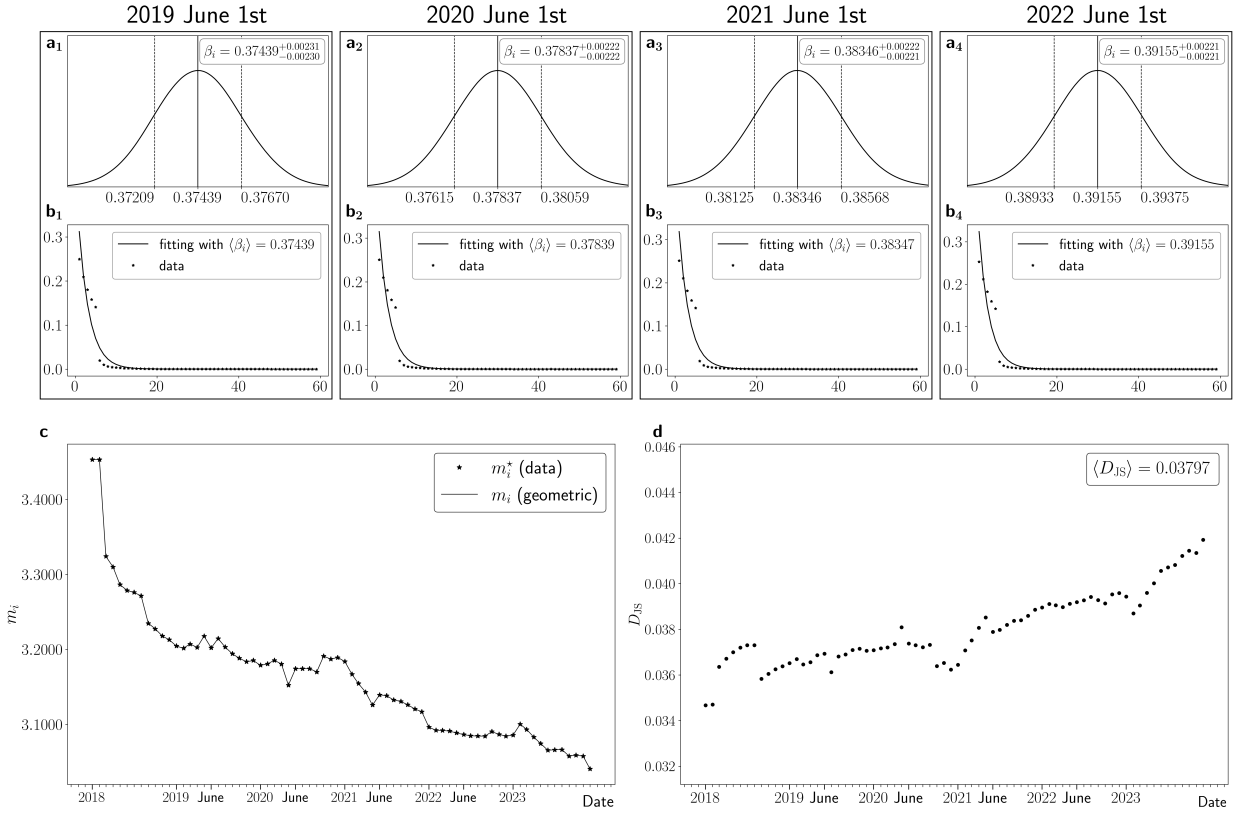

Figure SI 34: $i = 15000$

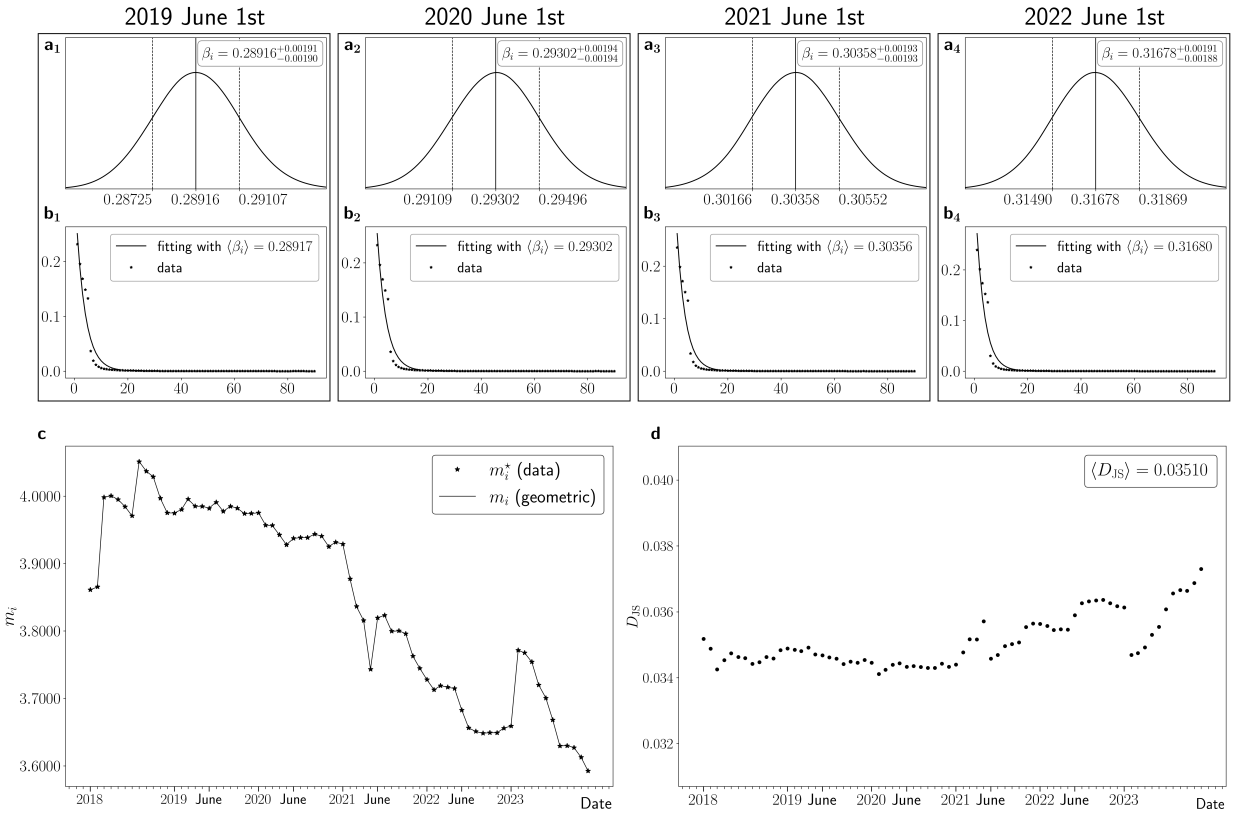

Figure SI 35: $i = 16000$

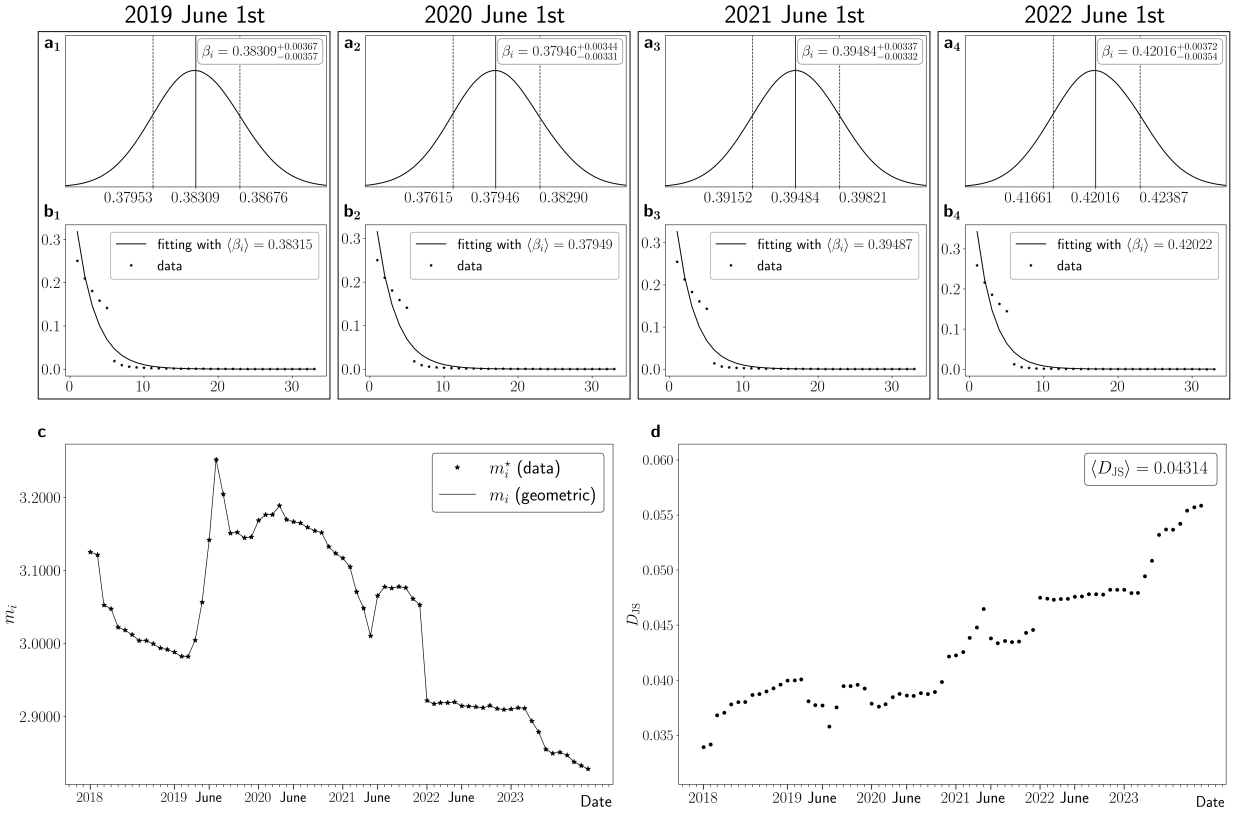

Figure SI 36: $i = 18000$

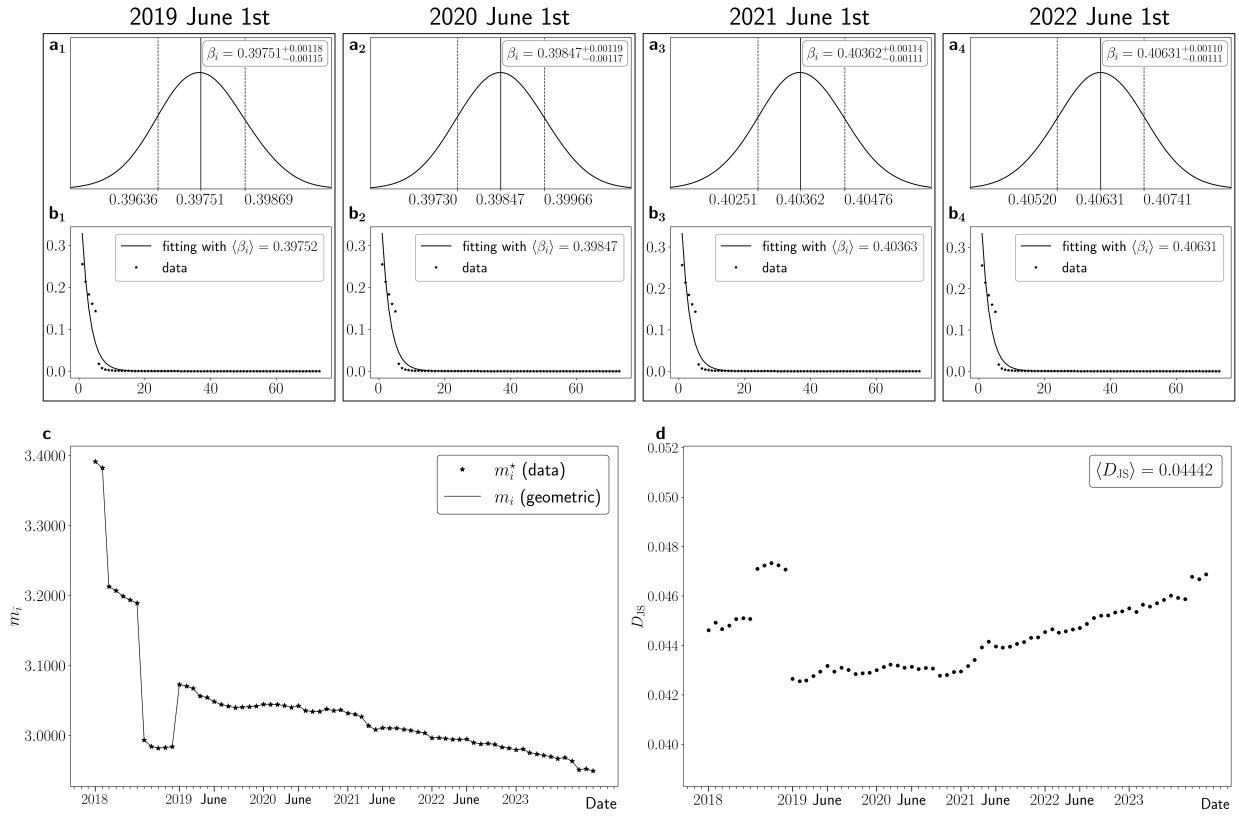

Figure SI 37: $i = 20000$

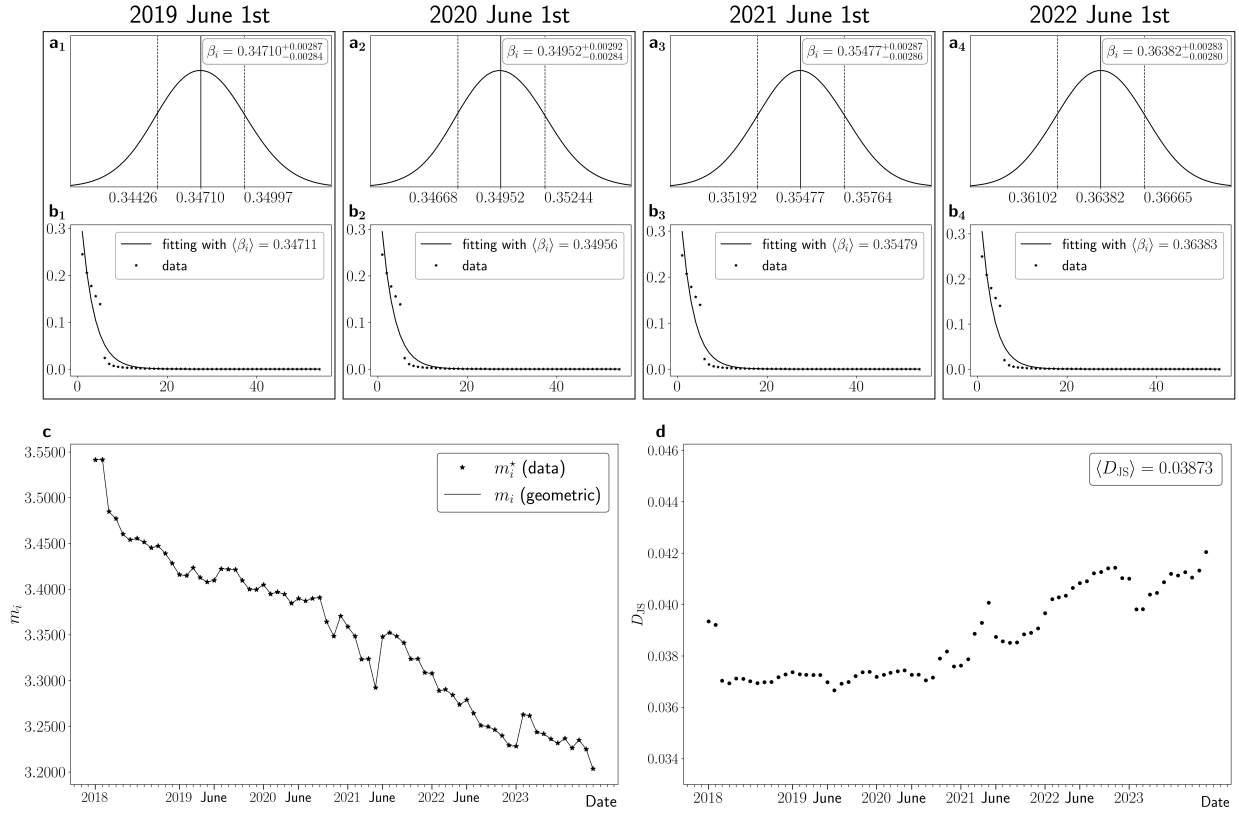

Figure SI 38: $i = 24000$

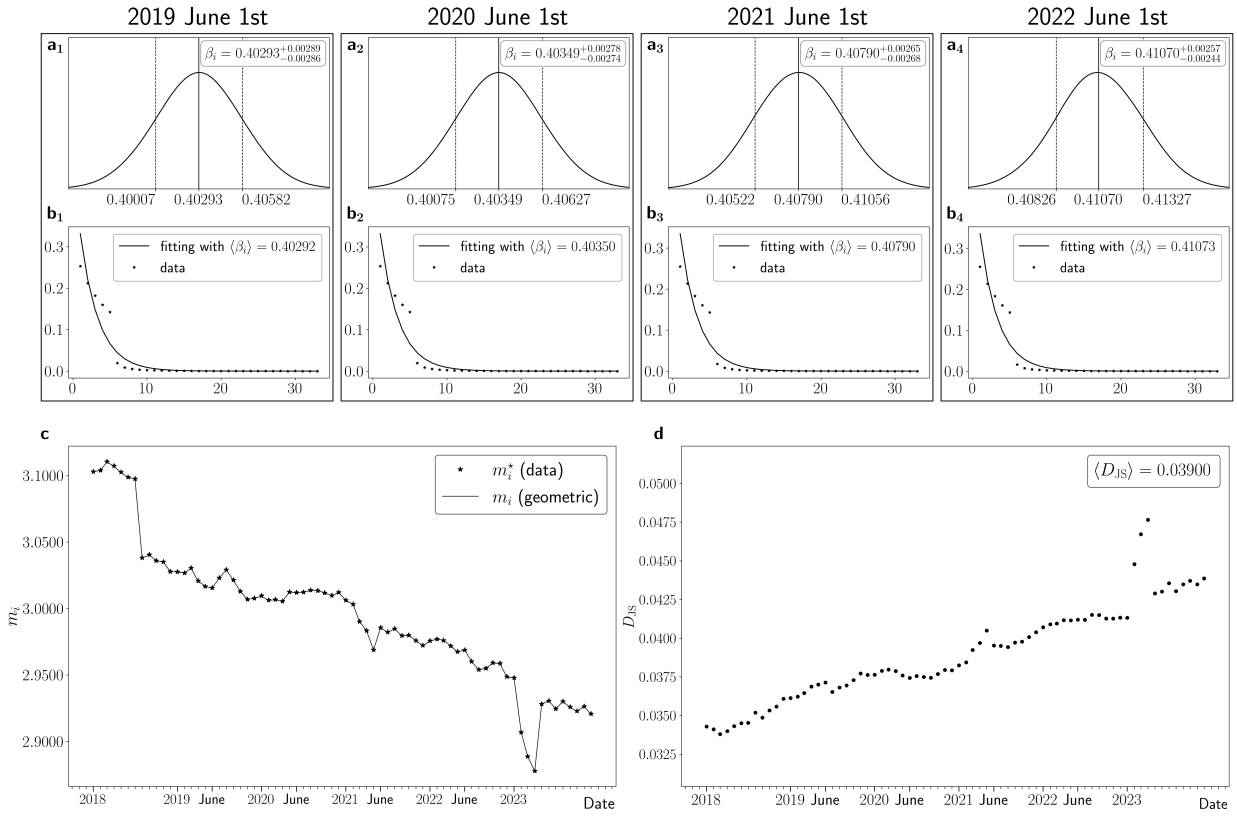

Figure SI 39: $i = 25000$

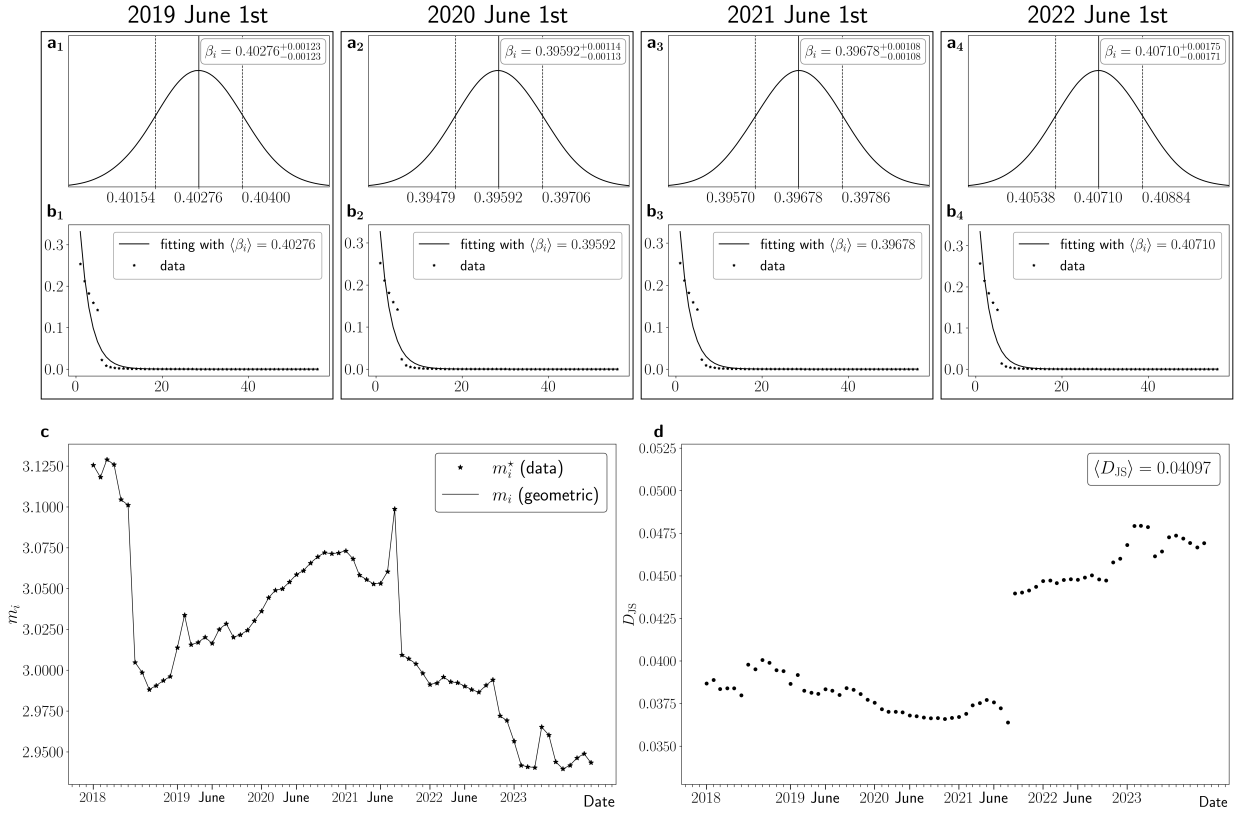

Figure SI 40: $i = 30000$

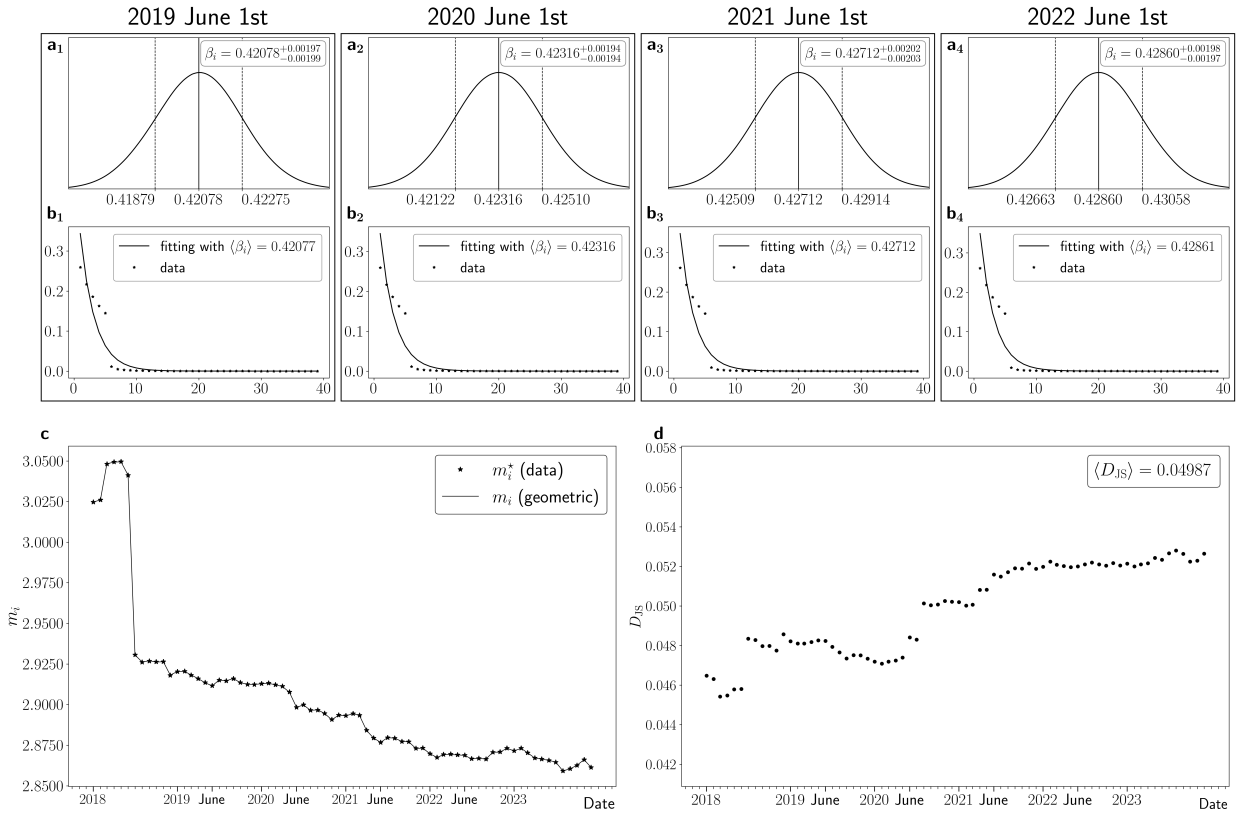

Figure SI 41: $i = 40000$

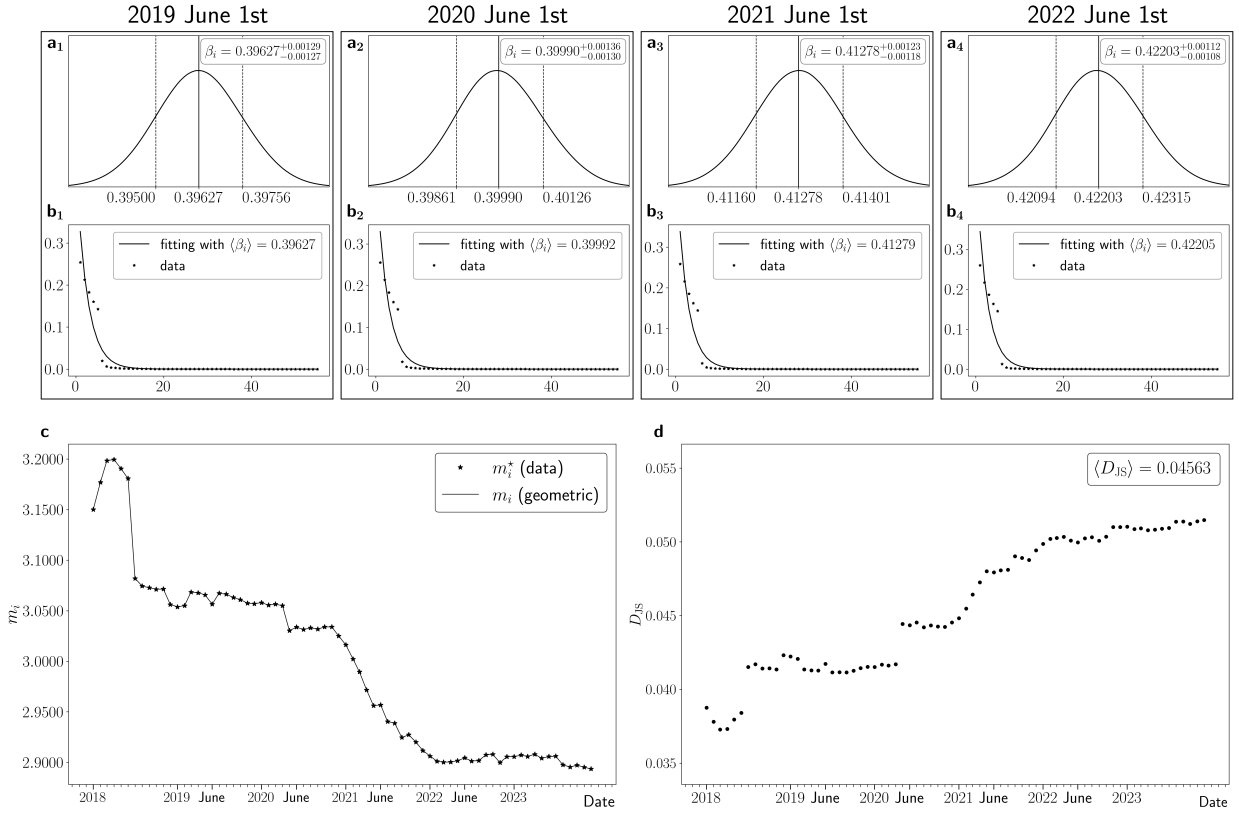

Figure SI 42: $i = 50000$

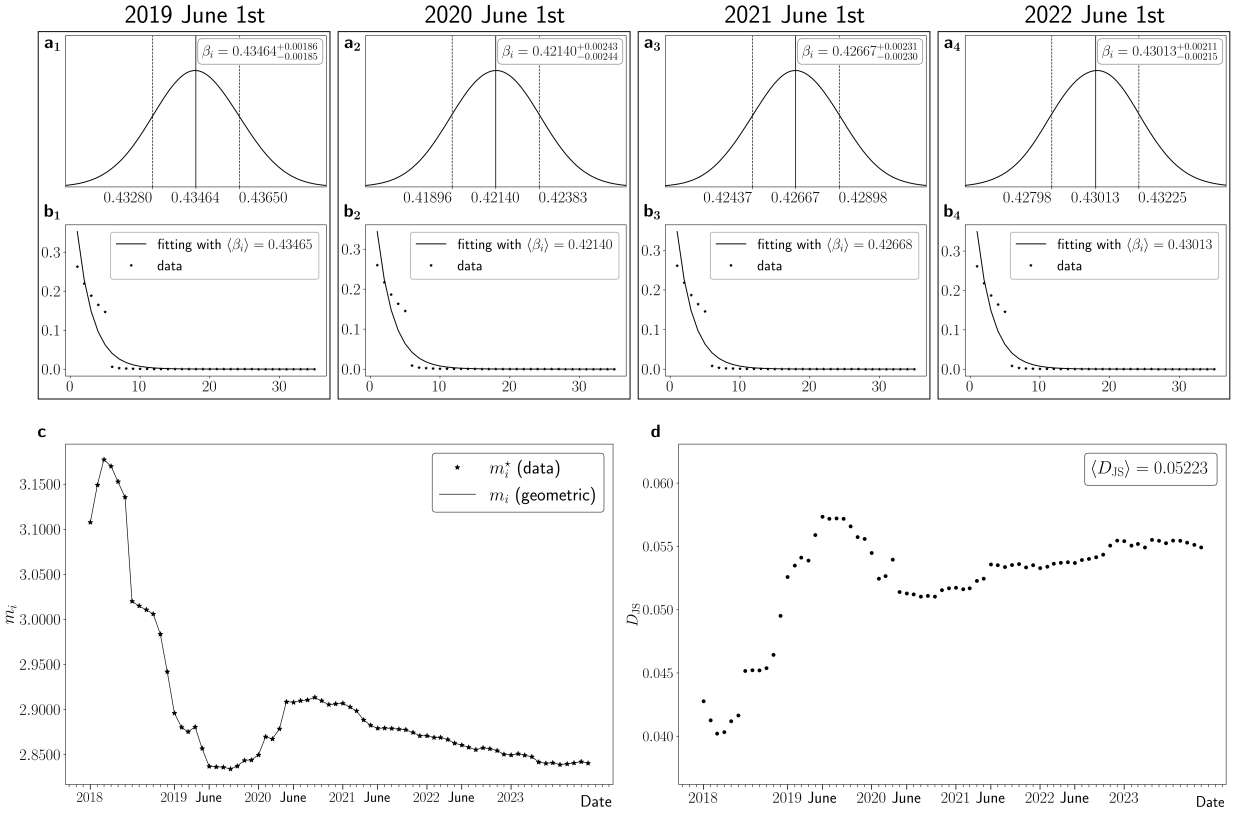

Figure SI 43: $i = 60000$

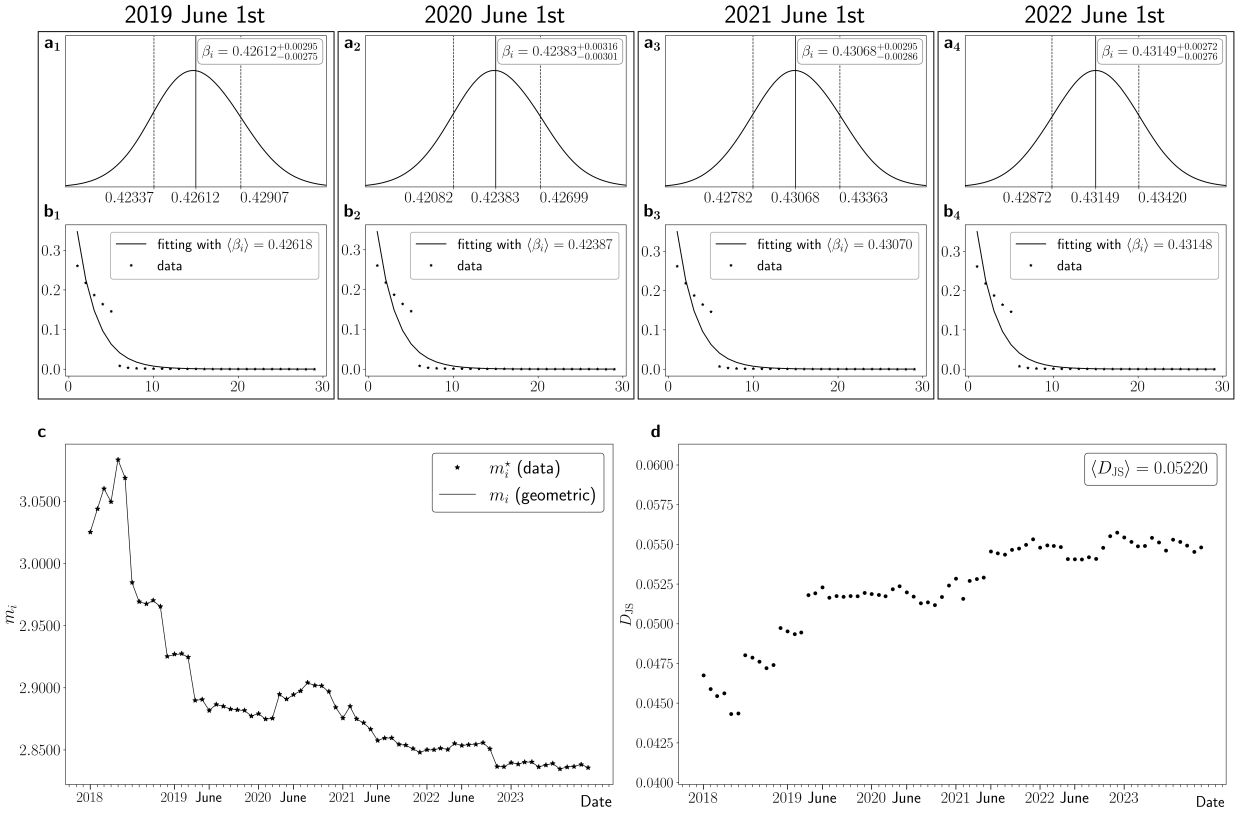

Figure SI 44: $i = 70000$

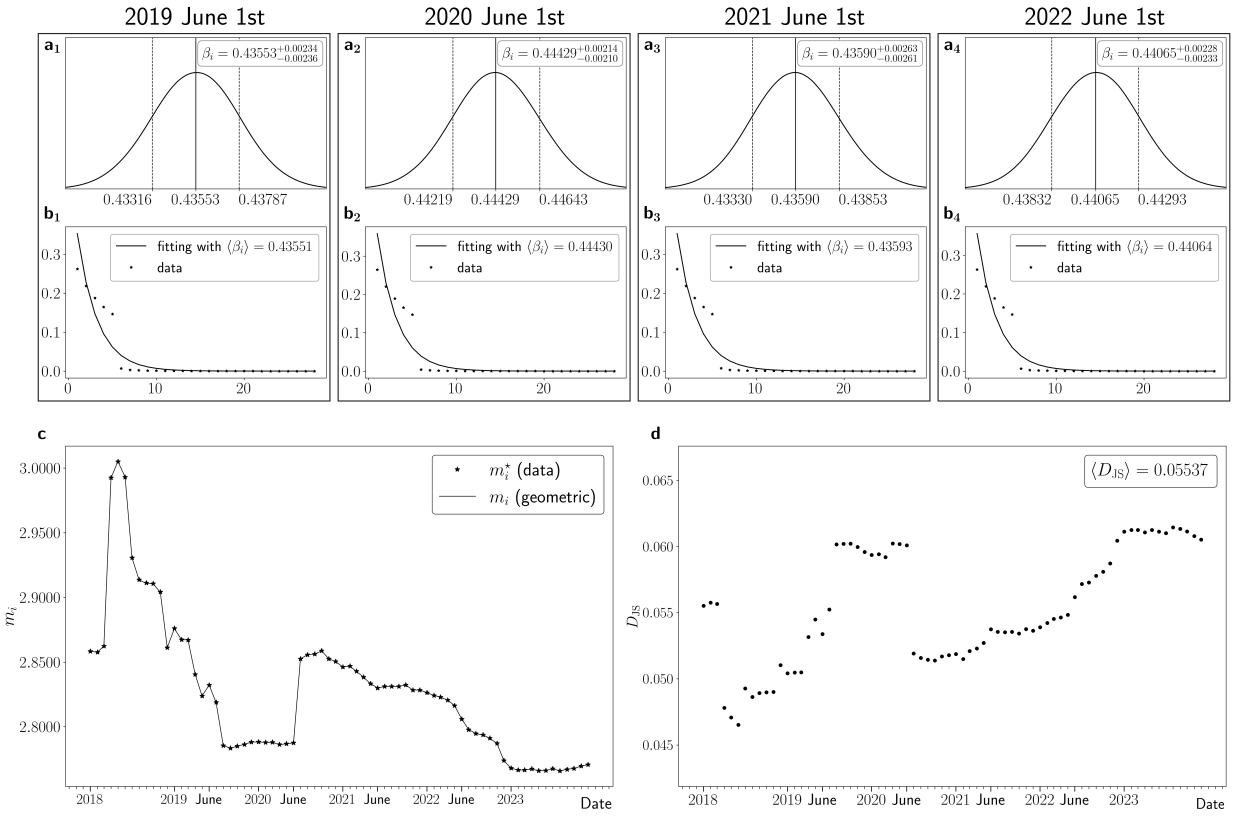

Figure SI 45: $i = 80000$

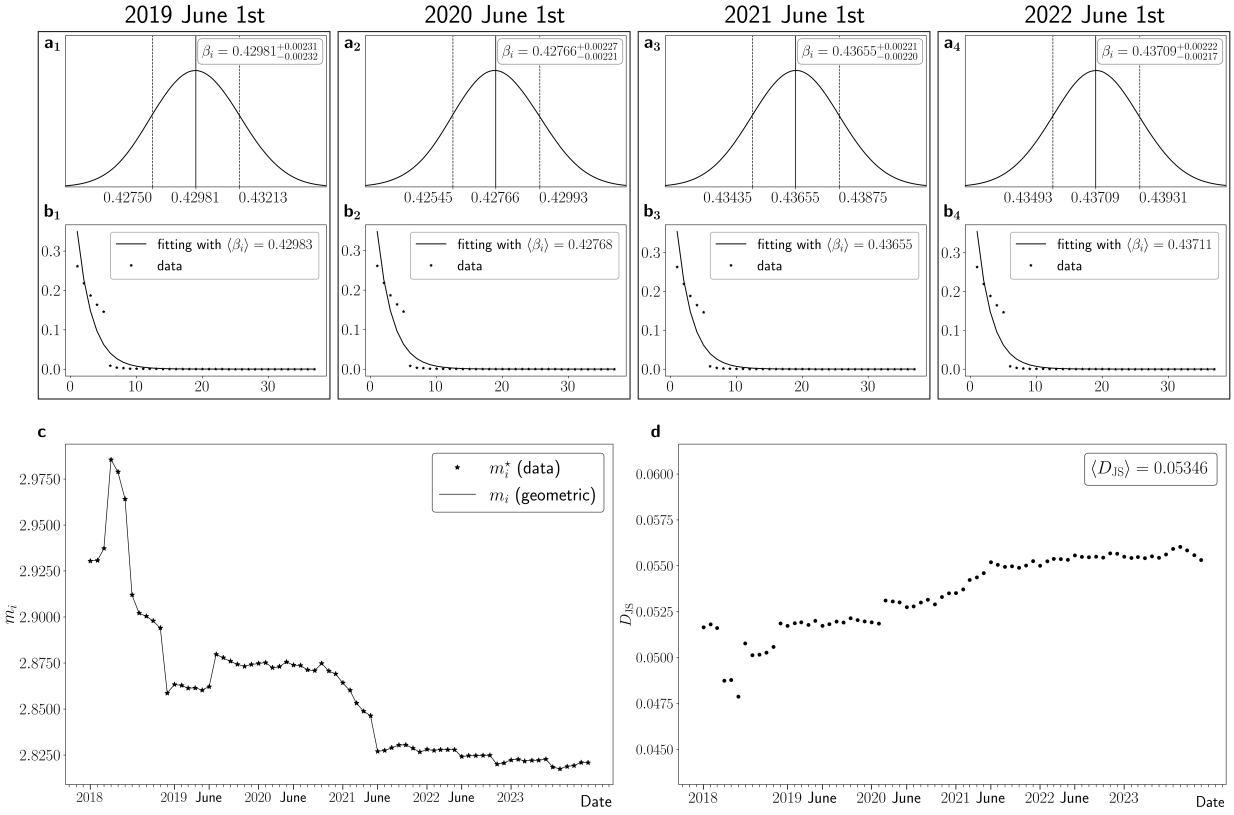

Figure SI 46: $i = 90000$

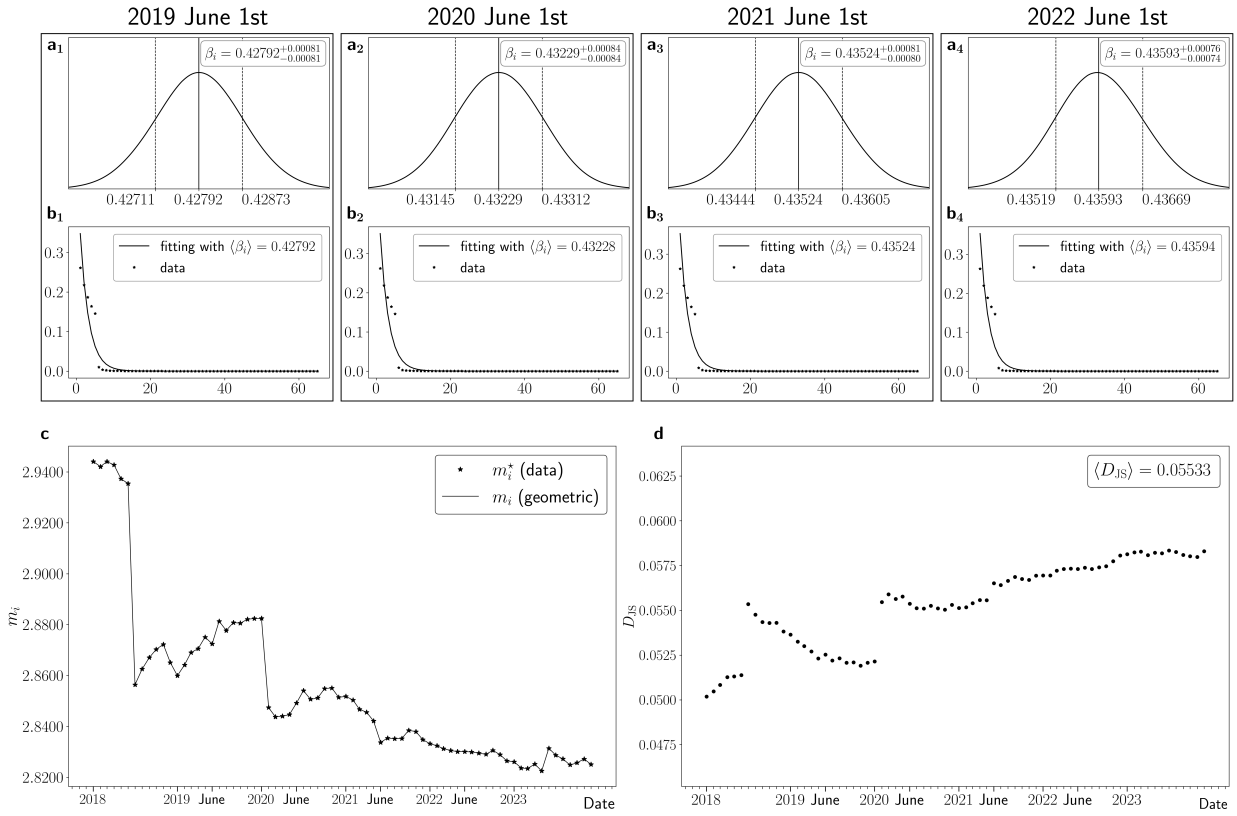

Figure SI 47: $i = 100000$

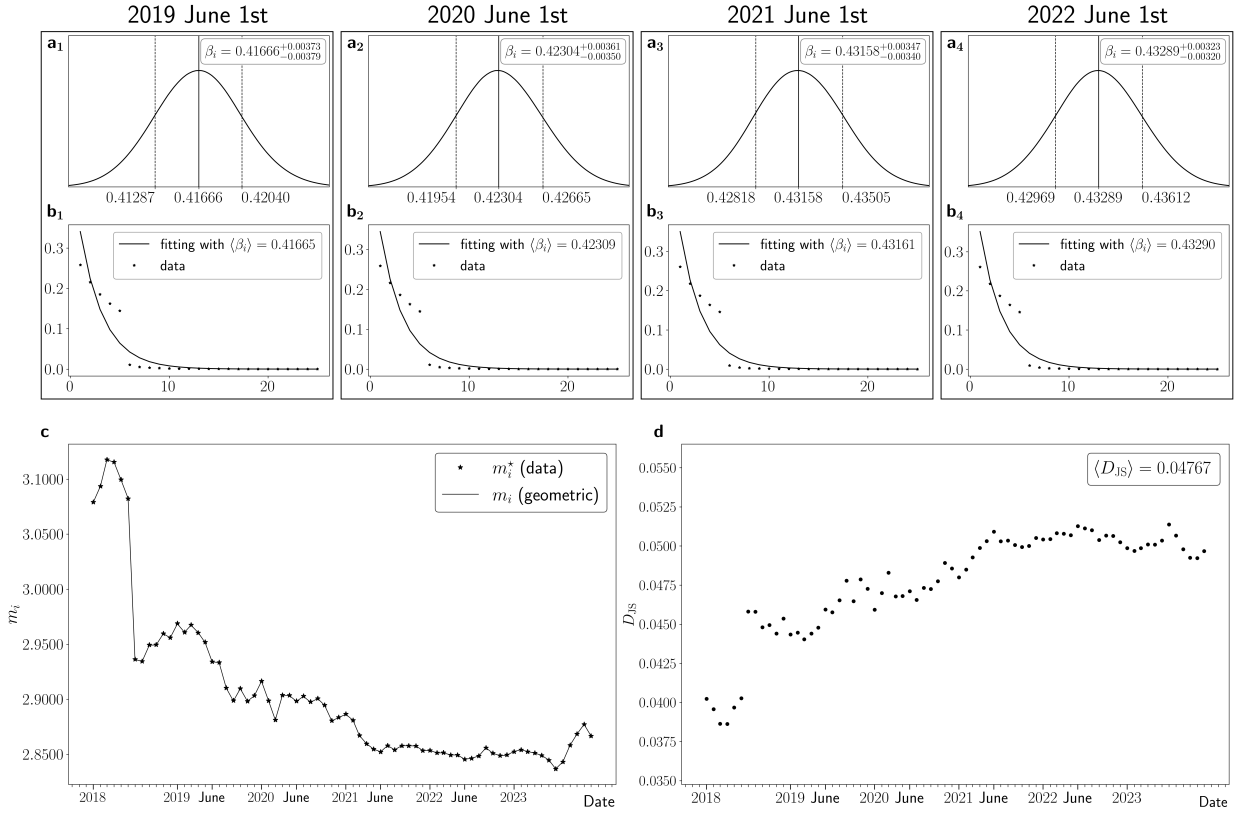

Figure SI 48: $i = 120000$

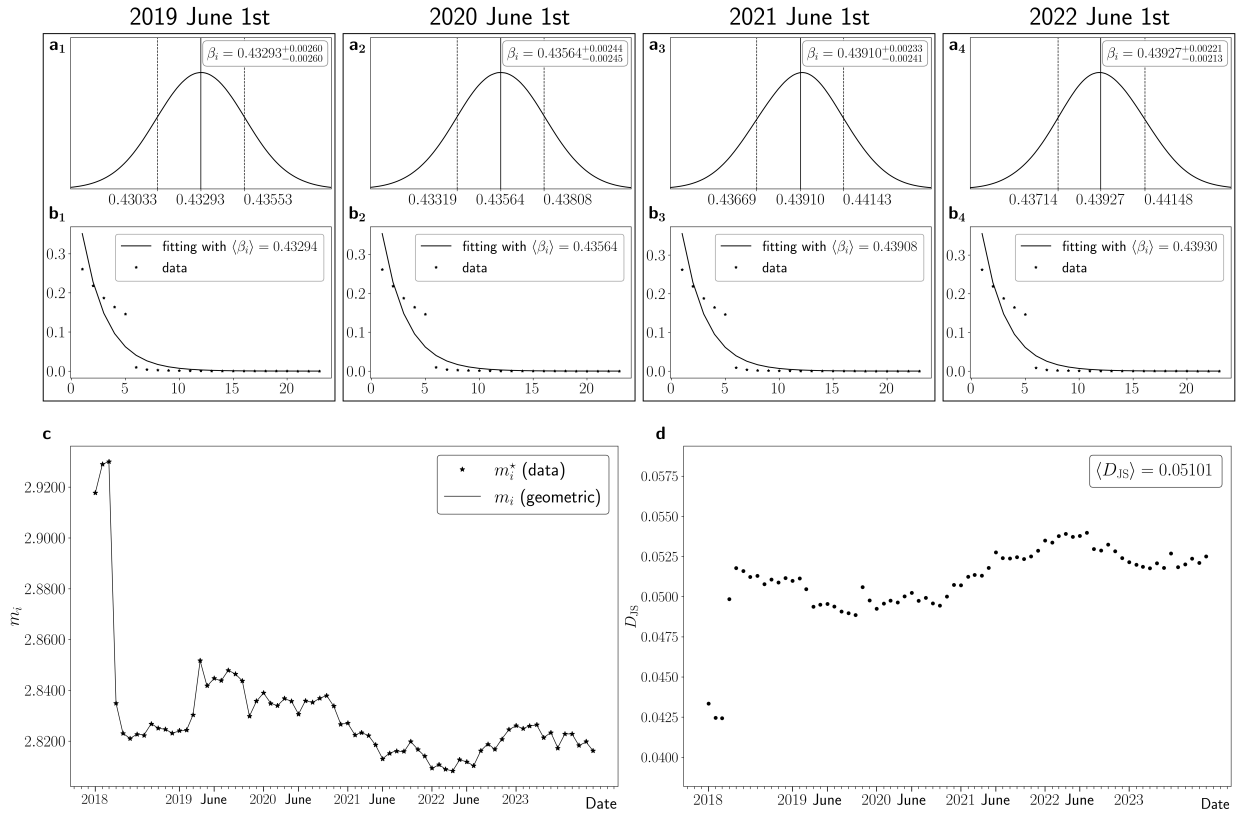

Figure SI 49: $i = 150000$

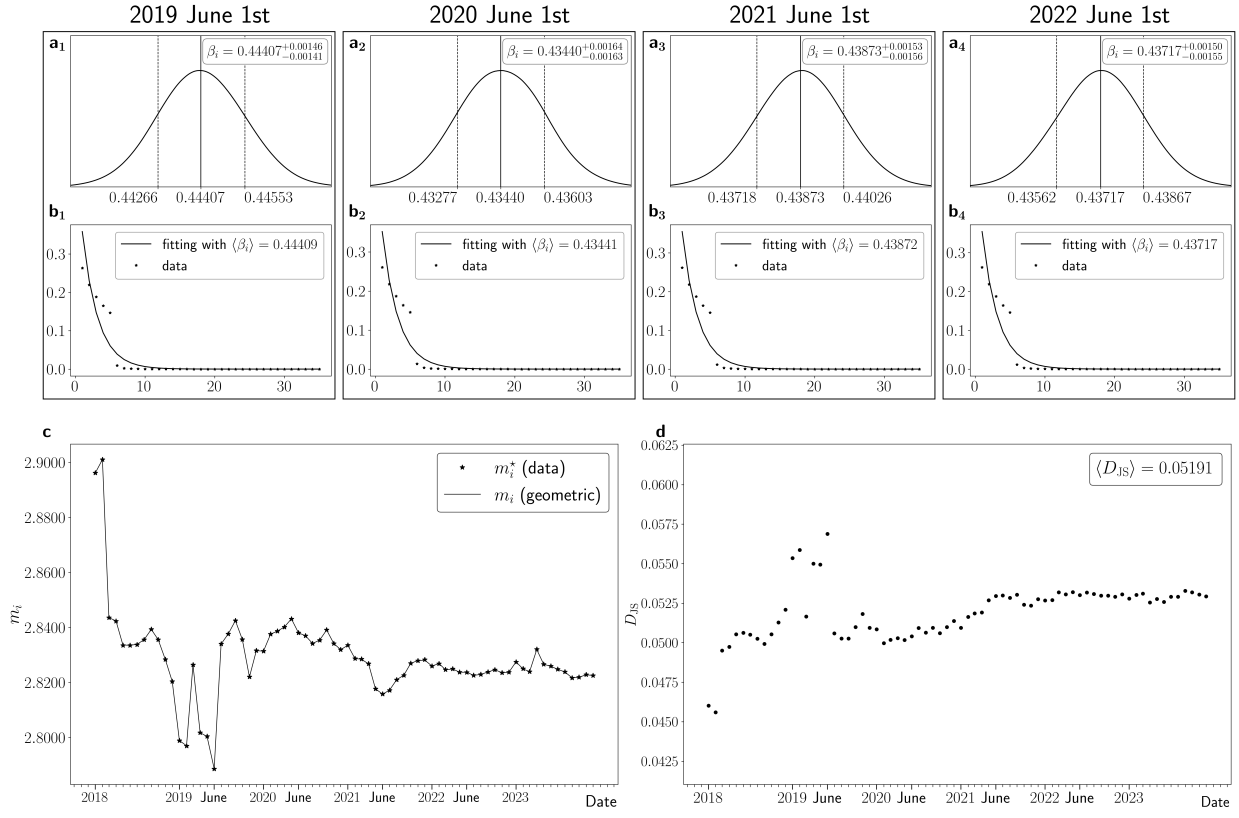

Figure SI 50: $i = 200000$

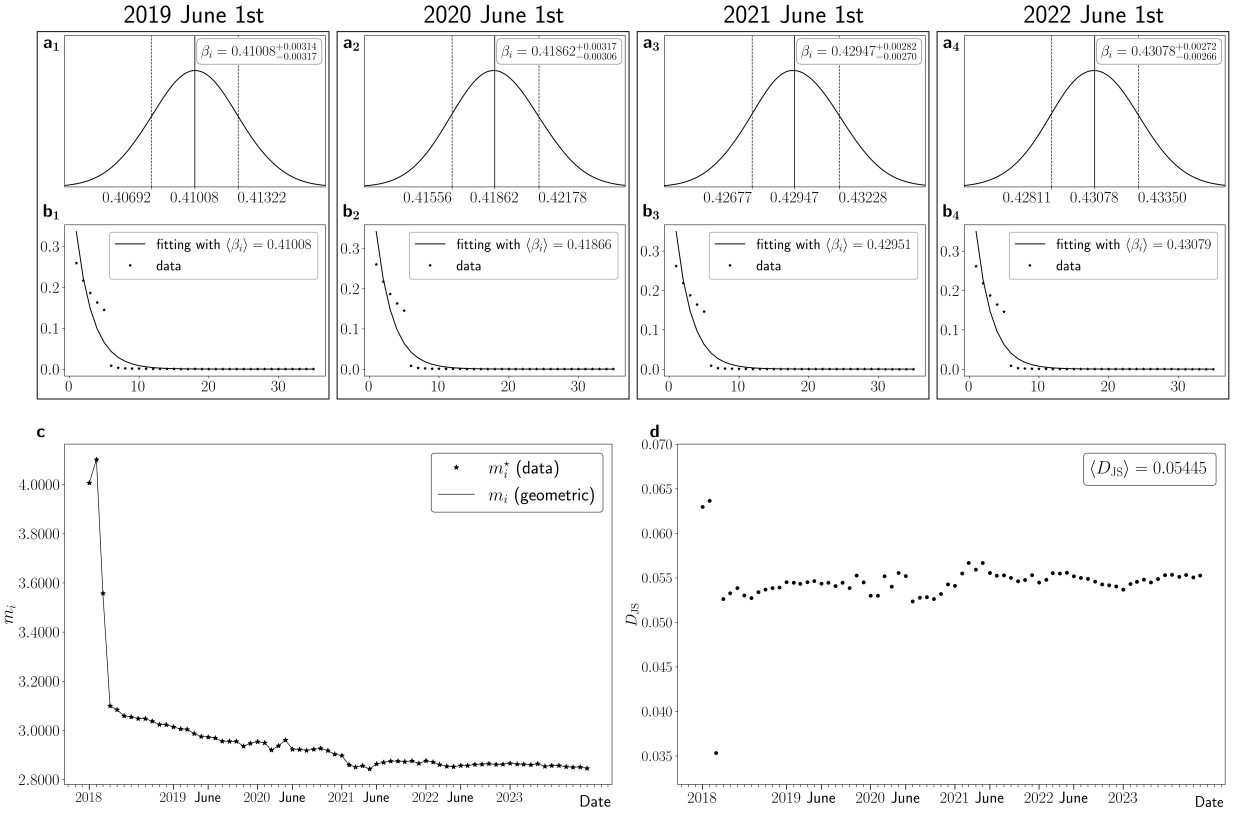

Figure SI 51: $i = 250000$

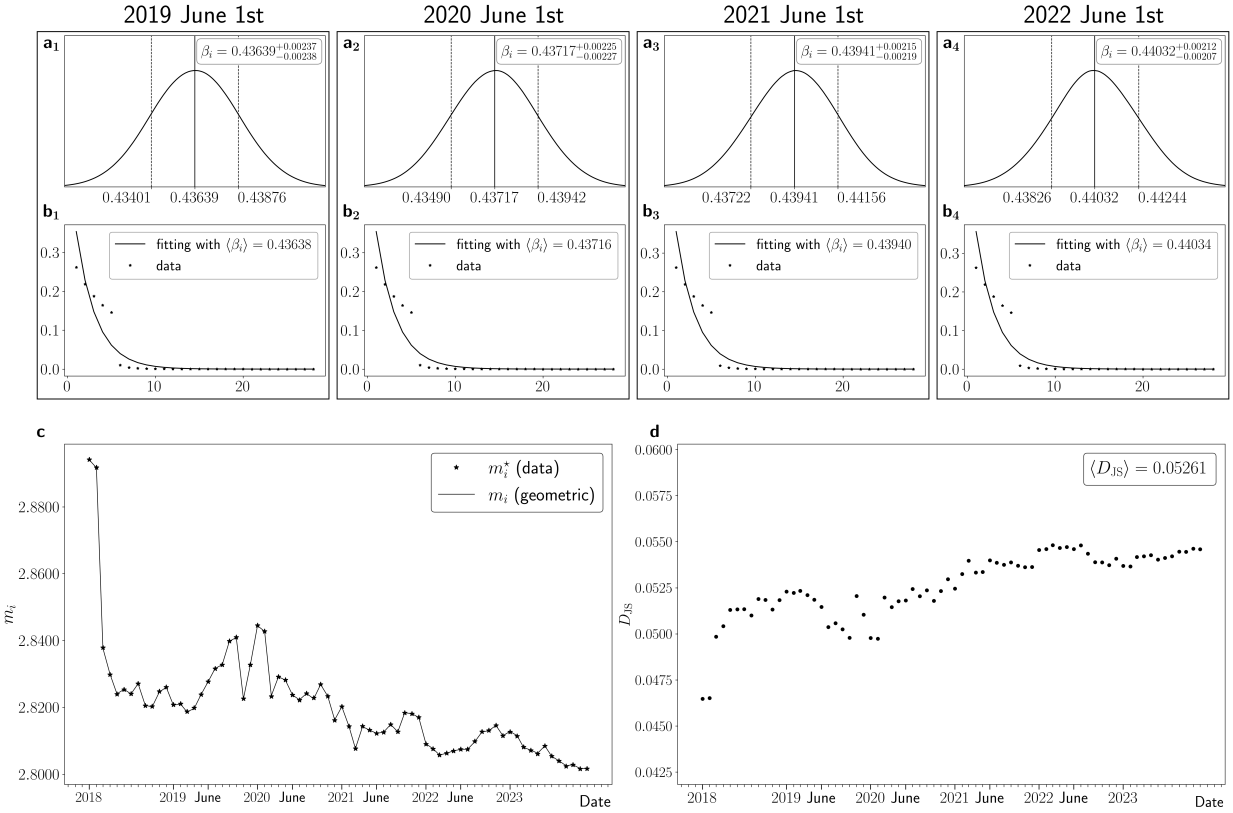

Figure SI 52: $i = 300000$

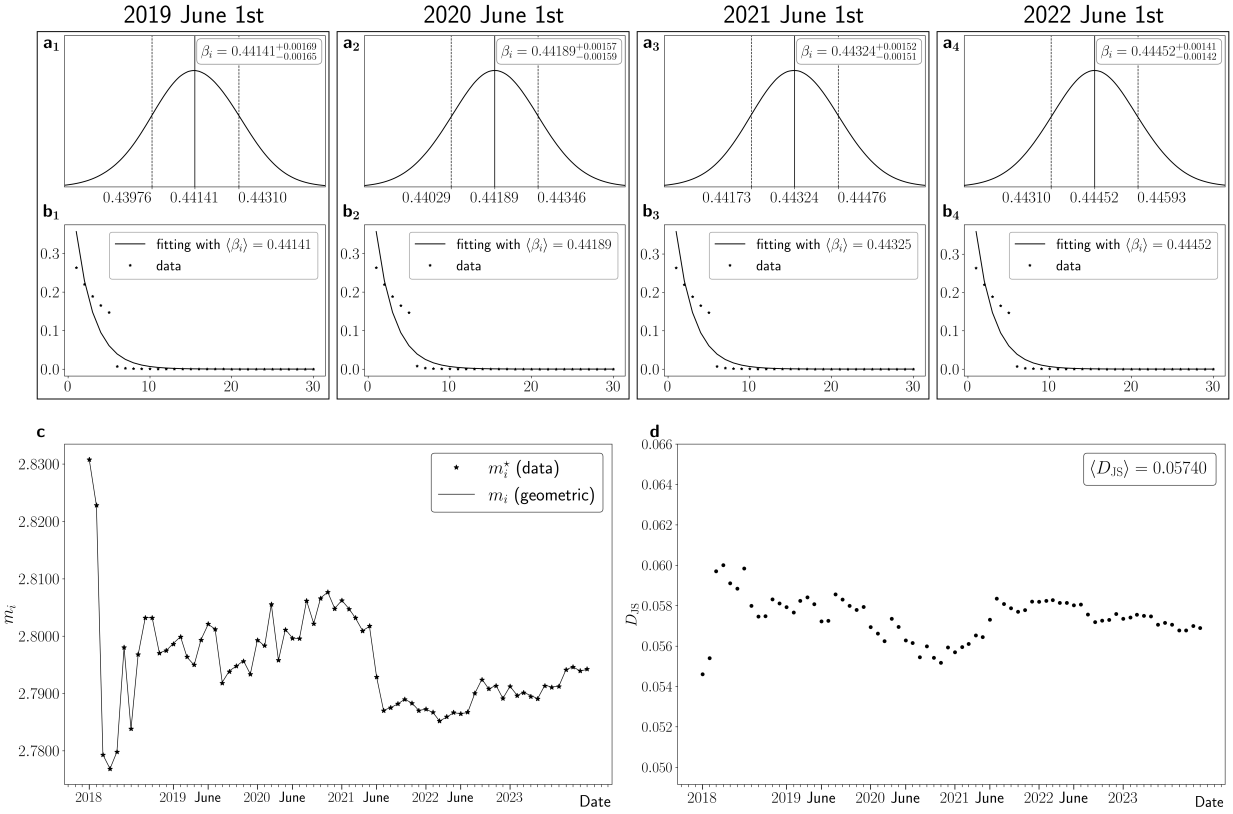

Figure SI 53: $i = 500000$

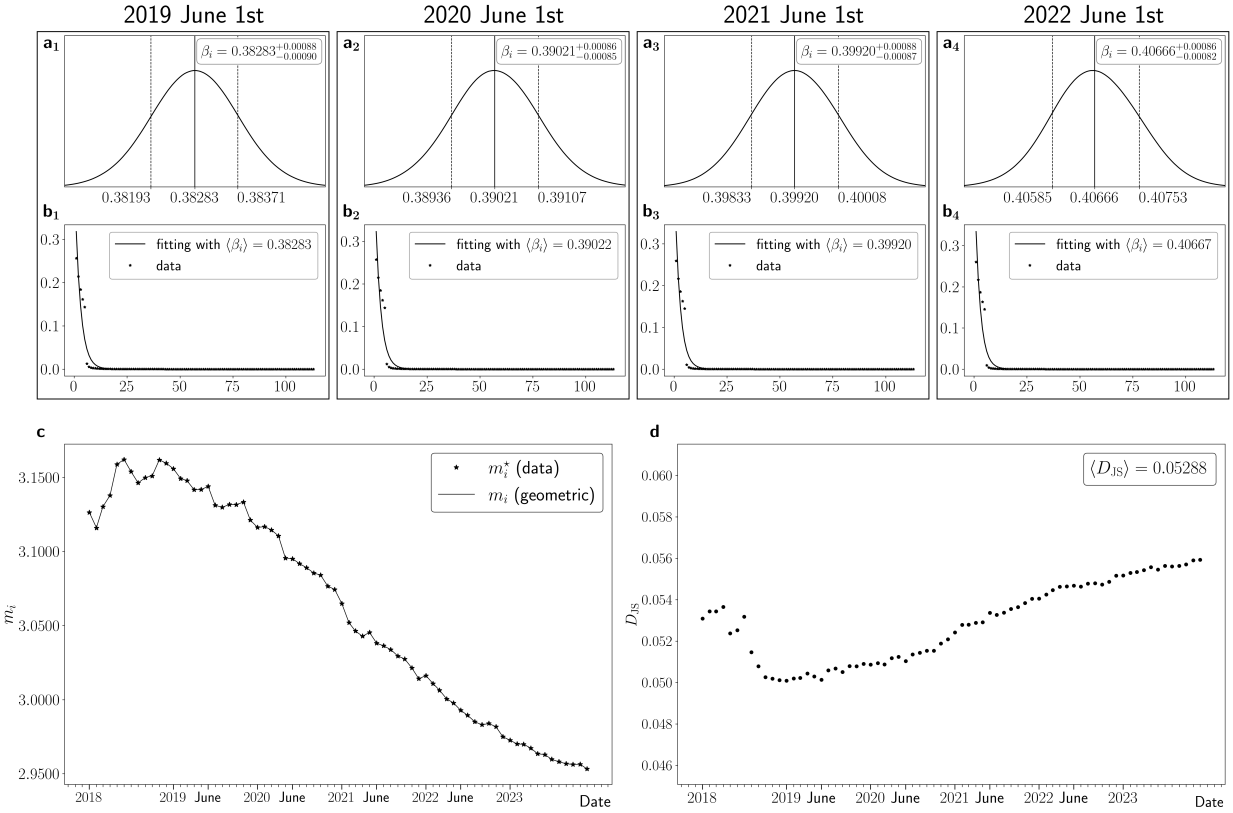

Figure SI 54: $i = 1000000$

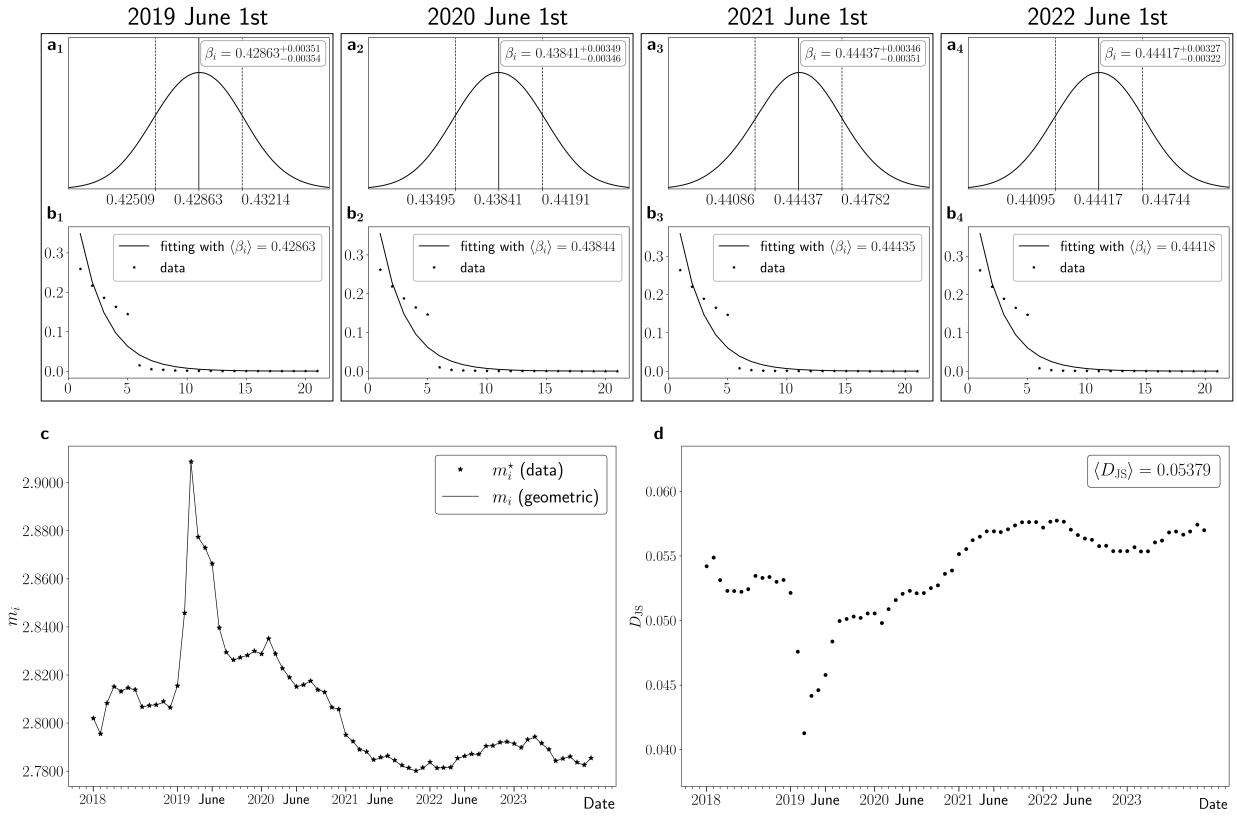

Figure SI 55: $i = 1500000$

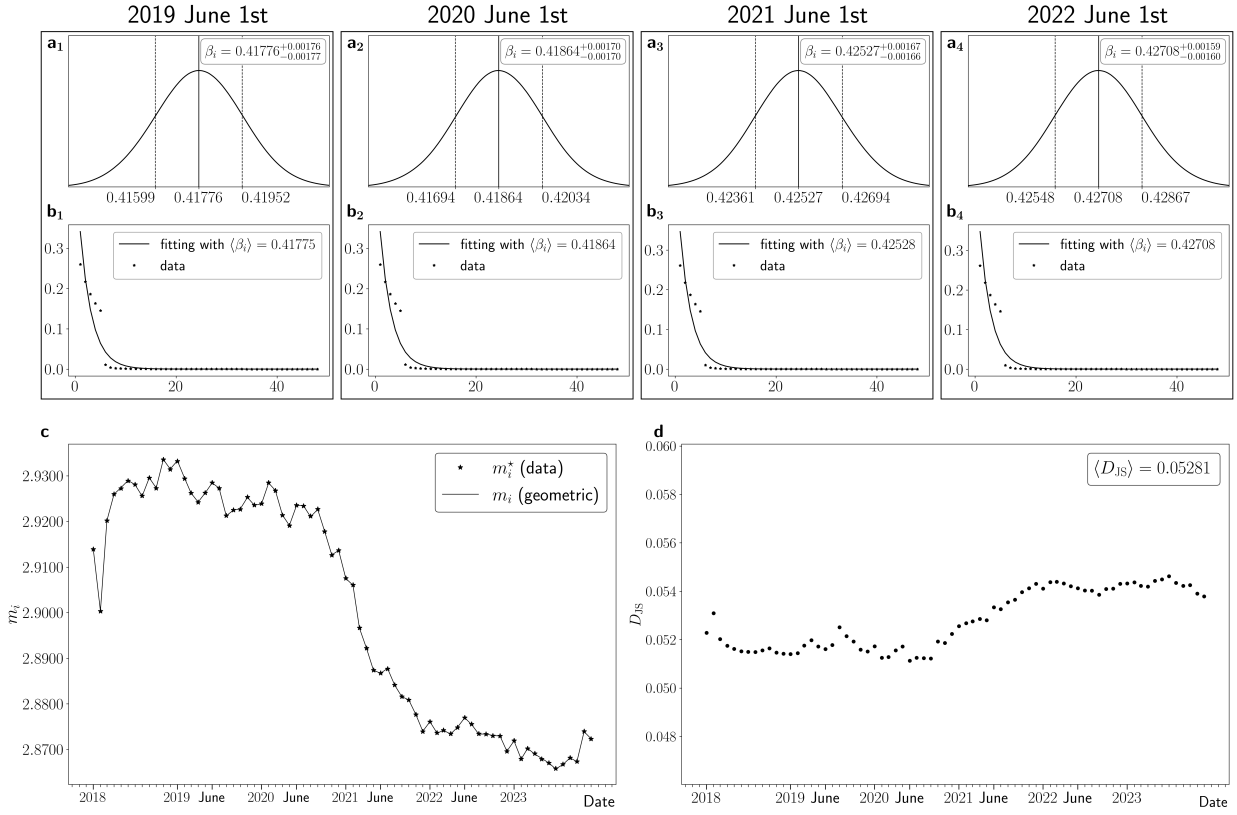

Figure SI 56: $i = 2000000$

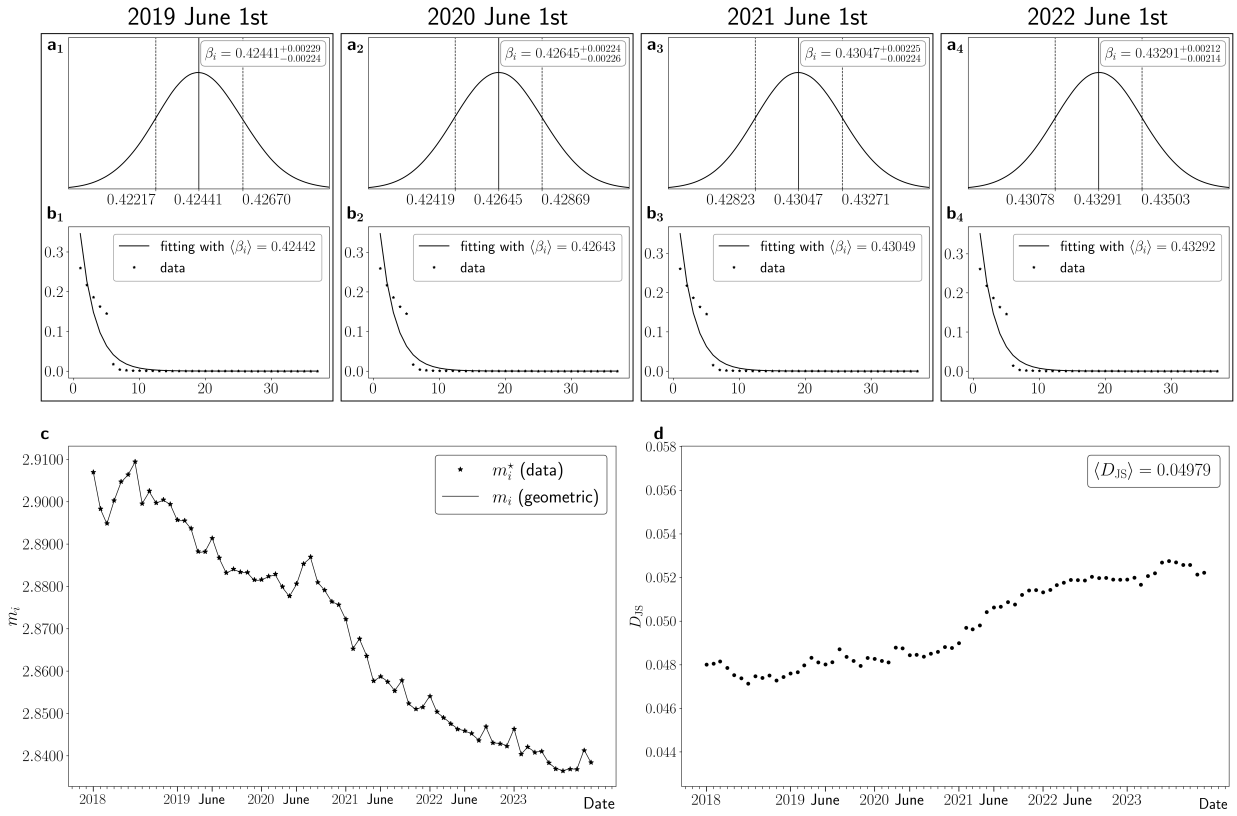

Figure SI 57: $i = 3000000$

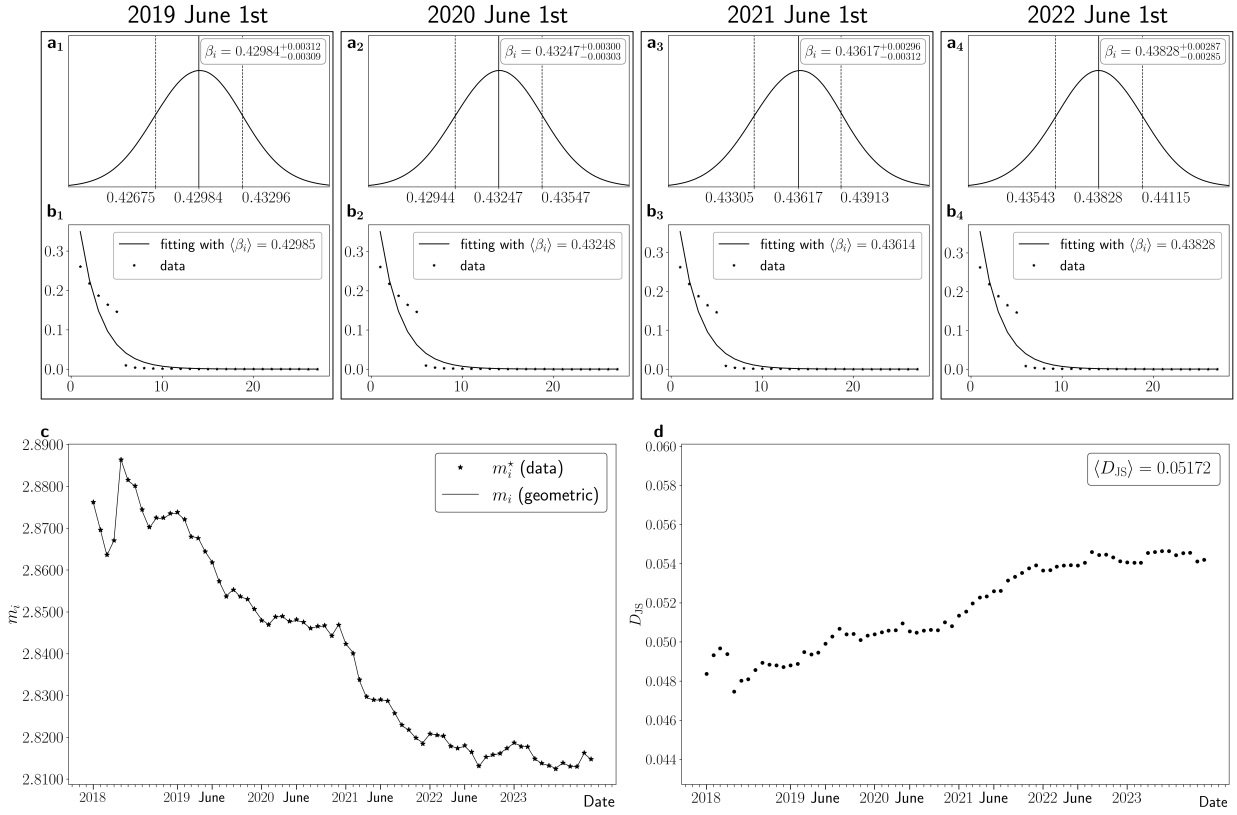

Figure SI 58: $i = 4000000$

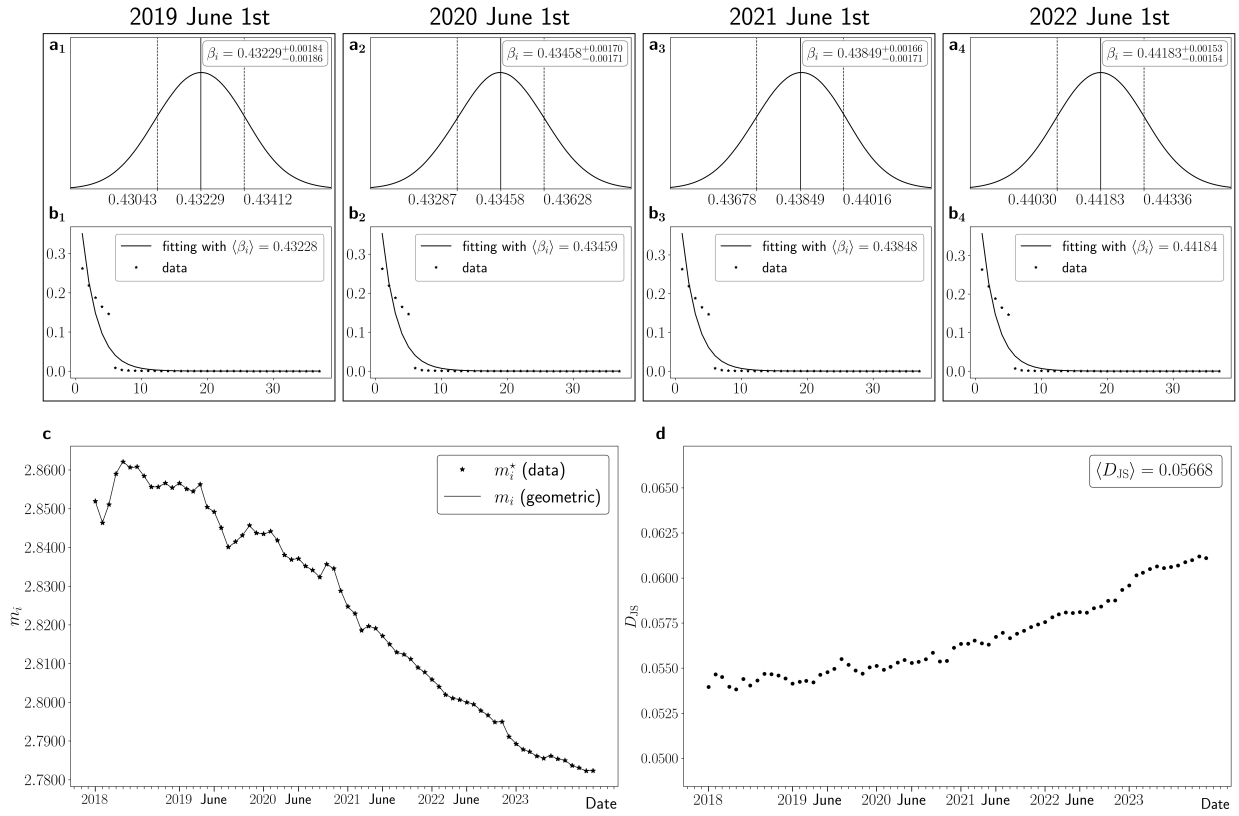

Figure SI 59: $i = 5000000$

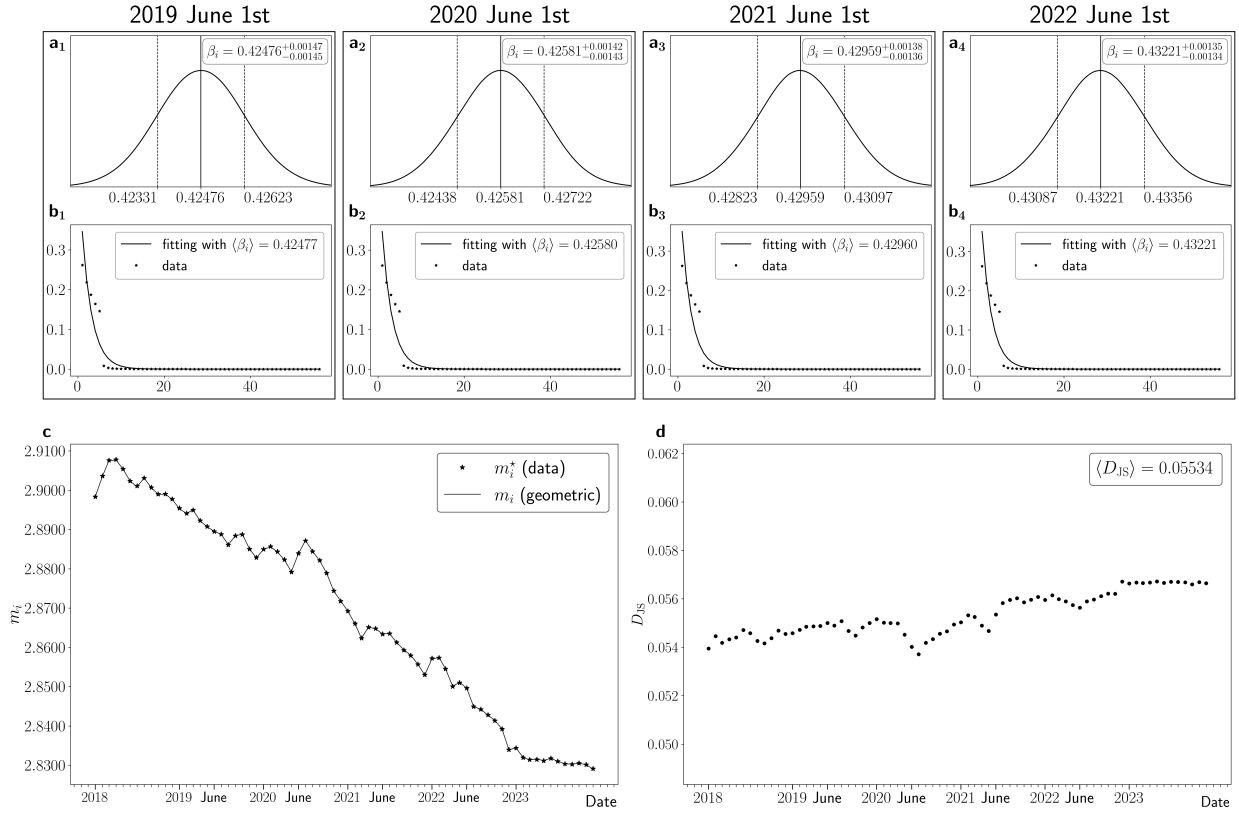

Figure SI 60: $i = 10000000$

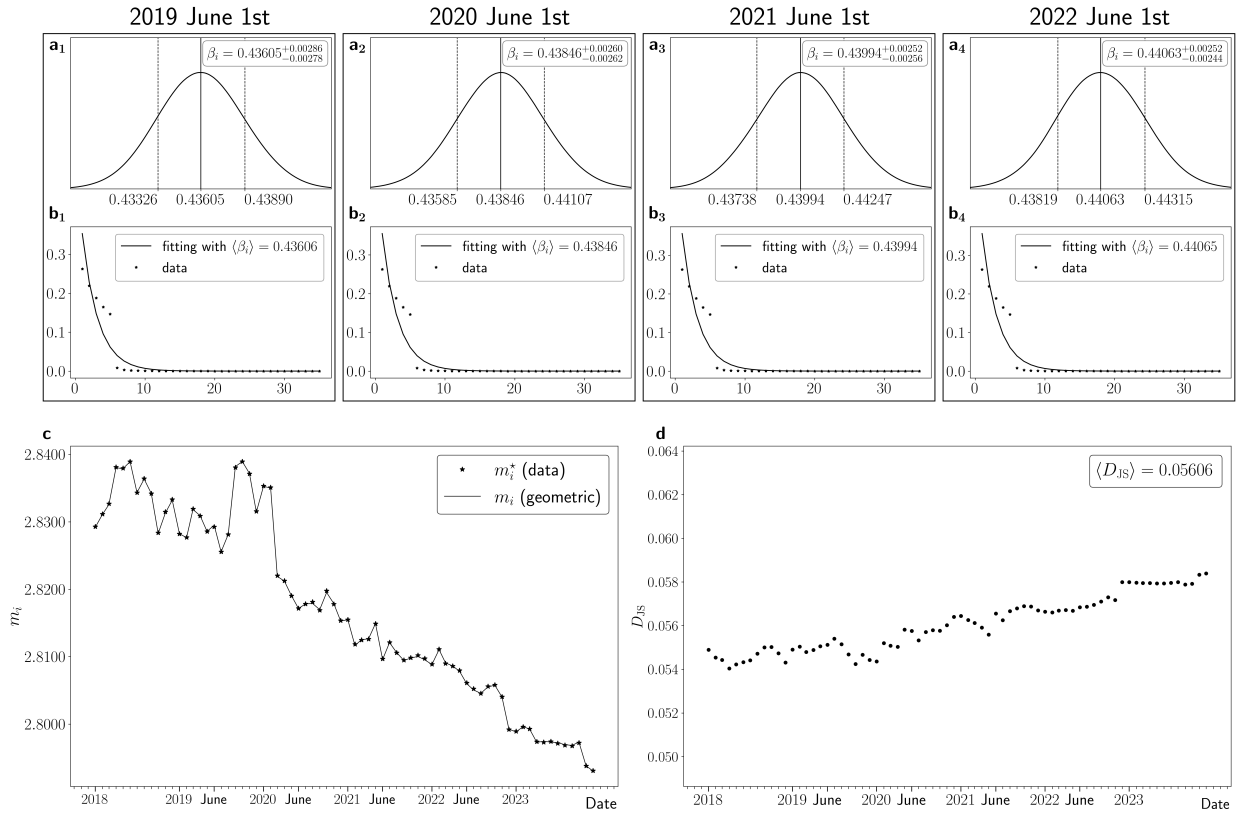

Figure SI 61: $i = 20000000$

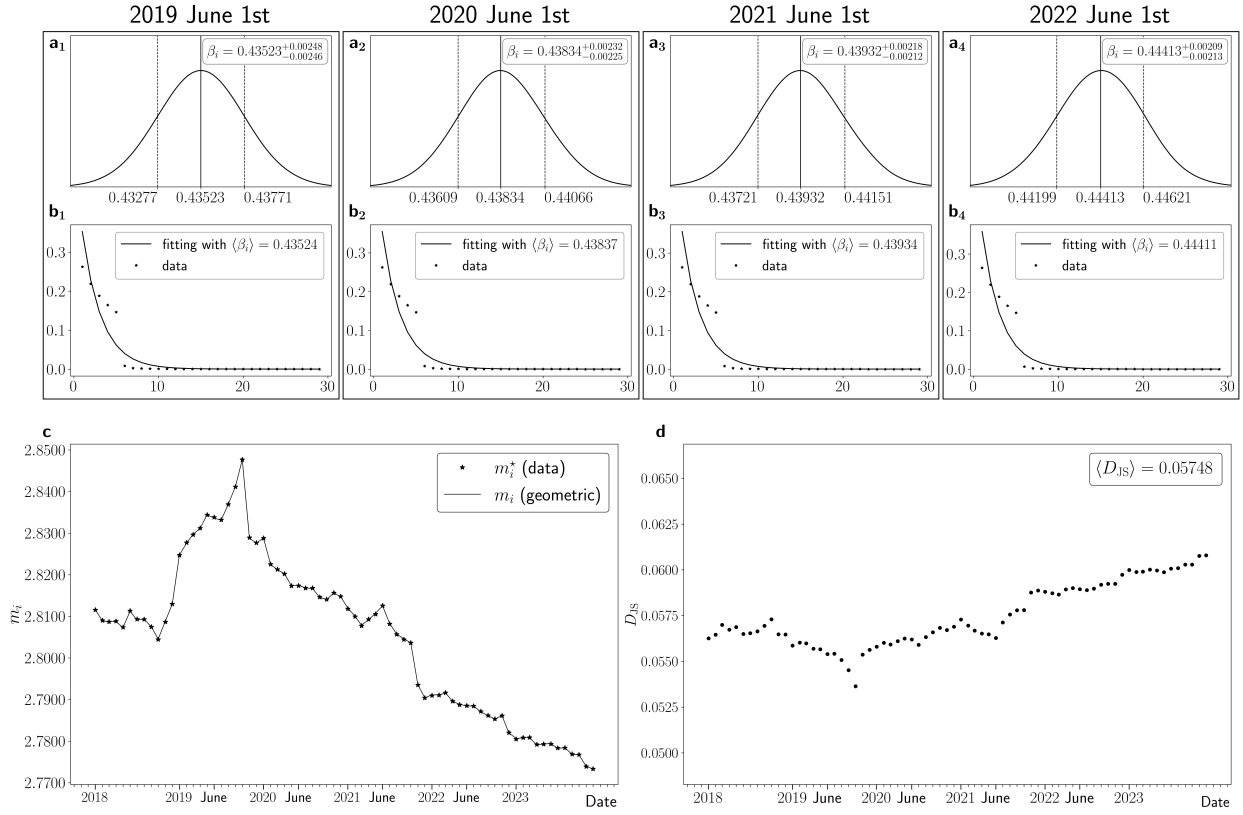

Figure SI 62: $i = 50000000$

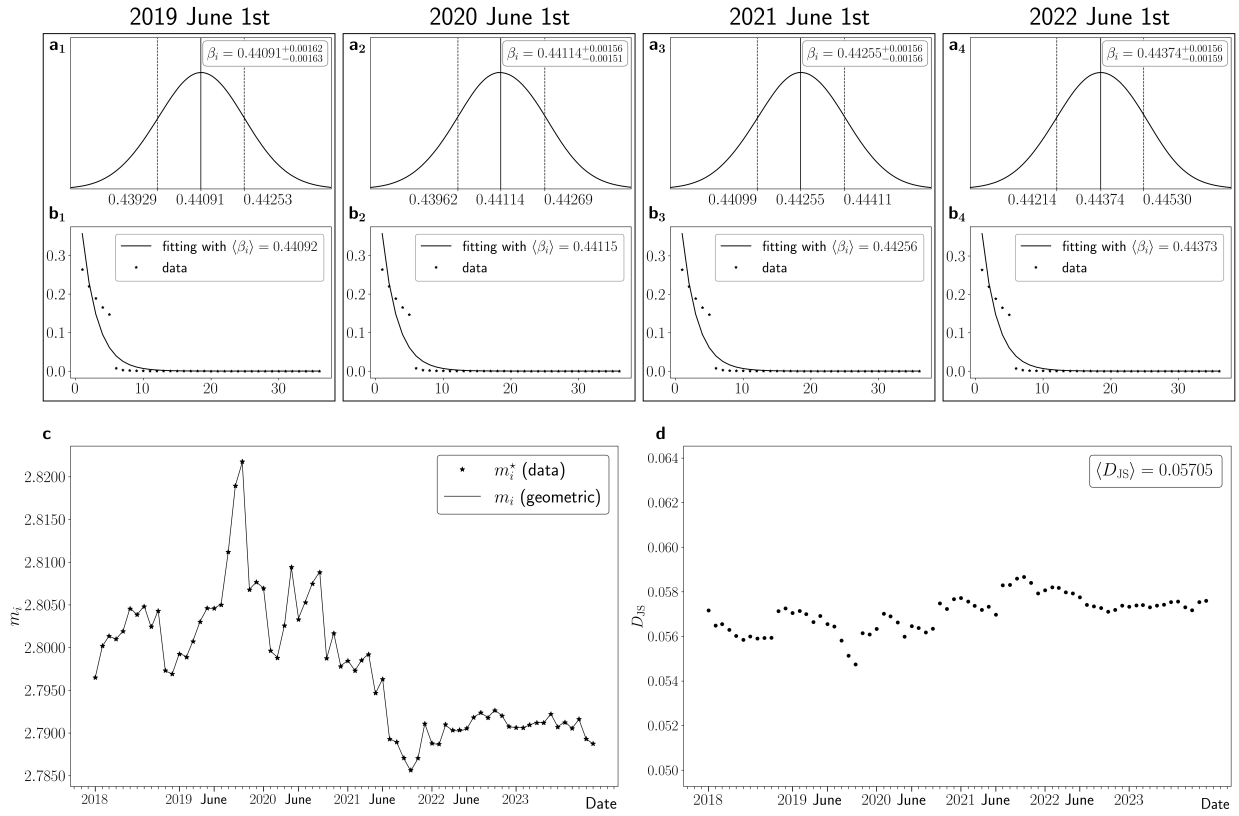

Figure SI 63: $i = 100000000$

Robustness checks supporting the geometric hypothesis

We summarise here four independent tests that complement the Jensen–Shannon divergence analysis of the main text. All tests are evaluated on the same 4,536 denomination–month samples used throughout.

(1) β–m self-consistency

The truncated geometric distribution is a one-parameter family in which the inverse temperature β and the mean m are mutually determined by the analytic relation given in the main text. Given the Bayesian-inferred β_i, the theoretical mean $m_i(\beta_i)$ must coincide with the empirical mean $m_i^\star$; equivalently, inverting the relation at $m_i^\star$ yields a second estimate $\beta_i^{(m)}$ that must match the Bayesian β_i. This is a nontrivial test that no two-parameter phenomenological model (log-normal, negative binomial) can satisfy by construction, because a single empirical mean does not fix two free parameters.

Figure SI 64 shows the comparison across all 4,536 samples. The Pearson correlation between β_i and $\beta_i^{(m)}$ is $r = 0.99999991$. The relative residual $|\beta_i - \beta_i^{(m)}|/\beta_i$ has mean 4.5×10^{-5} and maximum 4.8×10^{-4}: 89.2% of samples fall below 10^{-4}, and every sample falls below 10^{-3}.

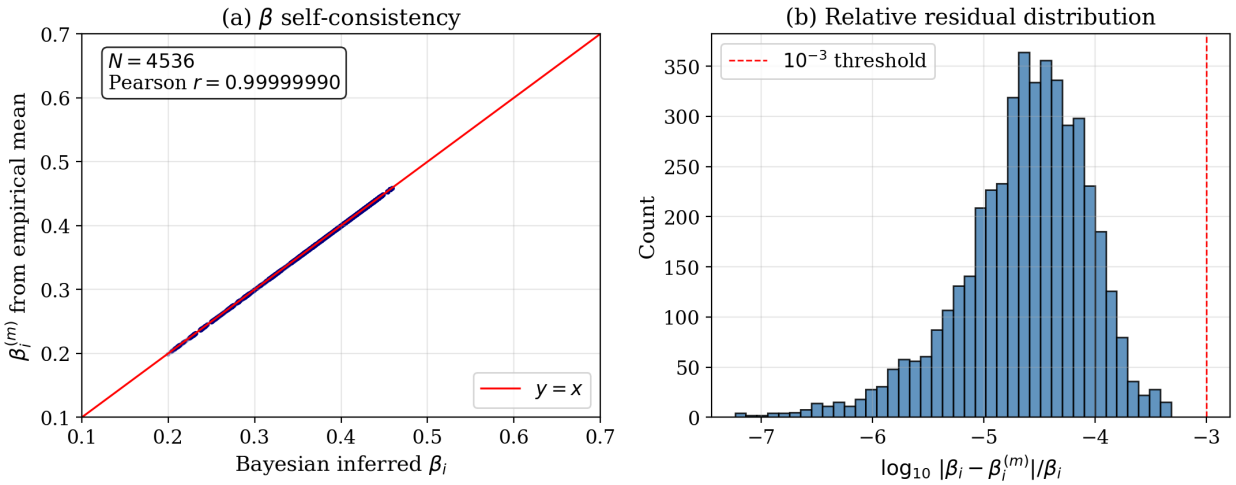

Figure SI 64: β–m **self-consistency of the truncated geometric model.** (a) Bayesian-inferred β_i versus $\beta_i^{(m)}$ obtained by inverting the analytic mean relation at the empirical mean $m_i^\star$, for all 4,536 denomination–month samples. Points lie essentially on the diagonal. (b) Distribution of $\log_{10}$ of the relative residual $|\beta_i - \beta_i^{(m)}|/\beta_i$. All samples satisfy the threshold 10^{-3} (red dashed line).

(2) Smoothing-window sensitivity

The main analysis uses a centred moving average of width $w = 9$ applied to the raw occupancy counts $\{n_k^\star\}$ prior to fitting. To confirm that this choice is not responsible for the observed goodness-of-fit, we repeated the geometric fit over $w \in \{1, 5, 9, 15, 21\}$, where $w = 1$ corresponds to no smoothing. For each w, the per-denomination common support $k_{\max,i}$ was redetermined from the corresponding smoothed counts, ensuring that the truncation matches

the smoothing scale; the geometric was then refitted by maximum-likelihood estimation on the smoothed counts (this substitutes the full Bayesian pipeline used in the main analysis for a fast, uniform protocol across all five window widths, and so yields slightly different percentages at $w = 9$ than the 99.74% of the main-text Table 1). Table 1 summarises the fraction of samples with $D_{\mathrm{JS}} < 0.08$ and < 0.10; Fig. SI 65 visualises the same data.

w	N	mean D_{JS}	$D_{\mathrm{JS}} < 0.08$ (%)	$D_{\mathrm{JS}} < 0.10$ (%)
1 (raw)	4,536	(tail-dominated)	76.6	85.0
5	4,536	0.048	94.7	97.5
9	4,536	0.048	98.4	98.4
15	4,536	0.052	98.4	99.2
21	4,536	0.055	98.4	99.7

Table 1: **Sensitivity of the geometric goodness-of-fit to the smoothing window width w.** All window choices $w \in \{5, 9, 15, 21\}$ yield $> 94\%$ of samples below $D_{\mathrm{JS}} = 0.08$; the raw case $w = 1$ is dominated by shot noise in the sparse tail rather than by any distributional mismatch.

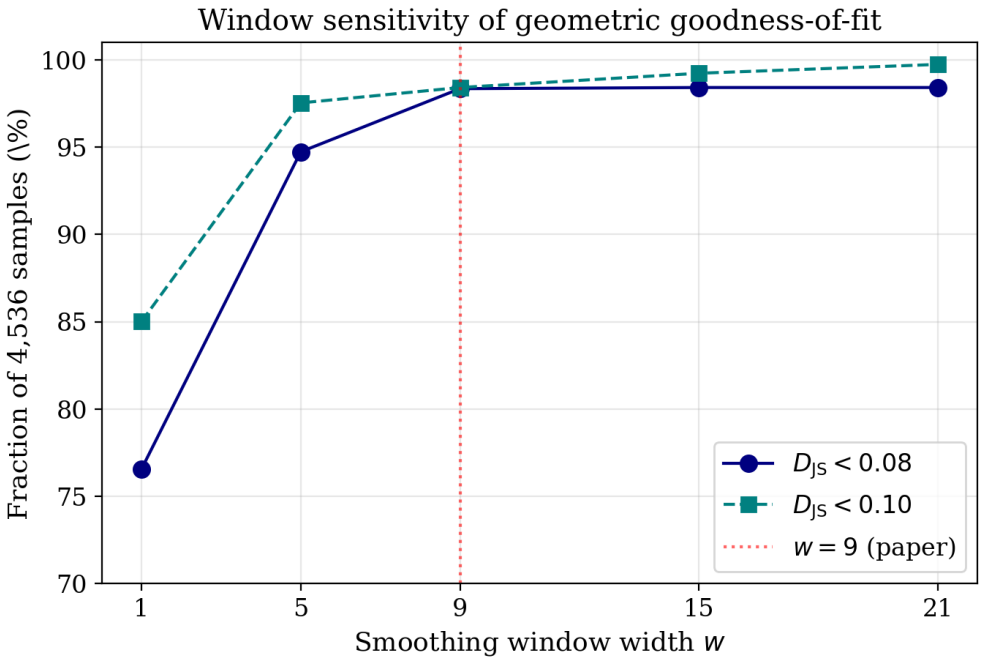

Figure SI 65: **Window-size sensitivity.** Fraction of 4,536 samples with D_{JS} below 0.08 (navy) and 0.10 (teal) as a function of the smoothing window width w. The main-text choice $w = 9$ (red dotted line) is indicated.

(3) Gini coefficient and Kolmogorov–Smirnov test

As two further parameter-free checks, we compare (a) the empirical Gini coefficient of each sample against the theoretical Gini coefficient implied by the fitted β_i and the truncated geometric form, and (b) compute the discrete Kolmogorov–Smirnov (K–S) statistic $D =$

$\max_k |\bar{F}_\star(k) - F_\beta(k)|$ between the smoothed empirical and theoretical cumulative distribution functions, where $\bar{F}_\star$ denotes the cumulative distribution derived from the smoothed $\bar{P}_\star(k)$.

The Gini coefficients track the theoretical prediction with Pearson correlation $r = 0.986$ (Fig. SI 66a), and the K–S statistic lies below 0.15 in 97.6% of samples (mean 0.099, median 0.097; Fig. SI 66b). Both tests independently support the geometric hypothesis.

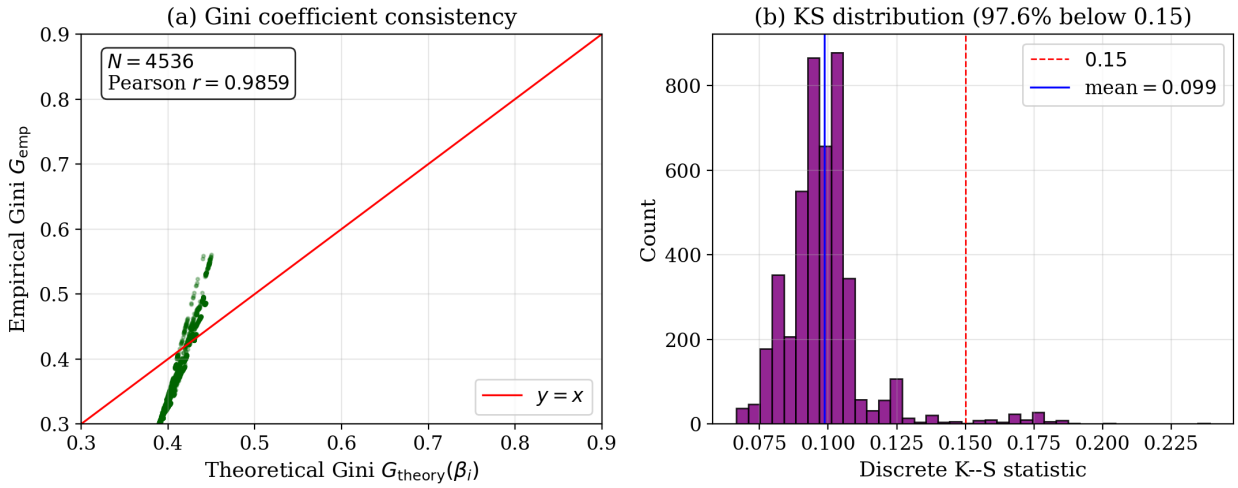

Figure SI 66: **Independent distributional checks.** (a) Scatter of empirical Gini coefficients versus theoretical values computed from the Bayesian-inferred β_i. (b) Histogram of the discrete Kolmogorov–Smirnov statistic across all samples; the 0.15 threshold is indicated in red, and the sample mean in blue.

(4) Time-stability of β_i per denomination

The Bayesian-inferred inverse-temperature parameter β_i is determined independently for each of the $63 \times 72 = 4{,}536$ denomination–month samples. To quantify the degree to which β_i is a stable structural property of denomination i, we compute the per-denomination coefficient of variation $\mathrm{CV}_i = \sigma(\beta_i)/\langle\beta_i\rangle$ across the 72 monthly snapshots. Figure SI 67(a) shows CV_i against the denomination i; panel (b) shows its distribution across the 63 denominations. The median CV_i is 1.7%, with 98.4% of denominations satisfying $\mathrm{CV}_i < 10\%$ and 92.1% satisfying $\mathrm{CV}_i < 5\%$. This near-constancy across a six-year window—which spans multiple Bitcoin market cycles—indicates that β_i is a structural property of the denomination rather than a transient feature of any one snapshot.

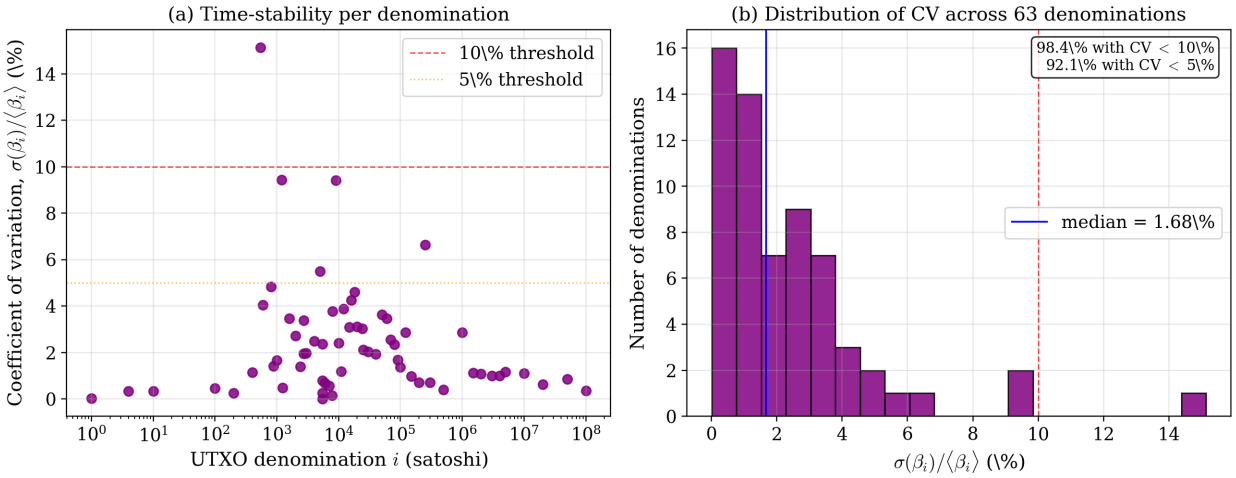

Figure SI 67: **Time-stability of β_i.** (a) Per-denomination coefficient of variation $\sigma(\beta_i)/\langle\beta_i\rangle$ as a function of denomination i. (b) Histogram of CV_i across all 63 denominations, with the median (blue) and the 10% threshold (red dashed) indicated.

Per-denomination $\langle\beta_i\rangle$ summary

Table 2 lists $\langle\beta_i\rangle$, $\sigma(\beta_i)$, and CV_i for each of the 63 denominations, averaged over the 72 monthly snapshots.

Table 2: **Per-denomination β_i statistics.** For each UTXO denomination i, the mean $\langle\beta_i\rangle$, standard deviation $\sigma(\beta_i)$, and coefficient of variation $\mathrm{CV}_i = \sigma(\beta_i)/\langle\beta_i\rangle$ are computed across the 72 monthly snapshots from January 2018 to December 2023.

i	$\langle\beta_i\rangle$	σ	CV (%)	i	$\langle\beta_i\rangle$	σ	CV (%)	i	$\langle\beta_i\rangle$	σ	CV (%)
1	0.4307	0.0001	0.03	2400	0.4043	0.0056	1.39	30000	0.4024	0.0082	2.03
4	0.3574	0.0012	0.34	2700	0.3543	0.0120	3.38	40000	0.4225	0.0082	1.93
10	0.2739	0.0009	0.34	2730	0.3192	0.0062	1.95	50000	0.4063	0.0147	3.62
100	0.2298	0.0010	0.46	3000	0.3917	0.0077	1.97	60000	0.4231	0.0146	3.45
200	0.2815	0.0007	0.26	4000	0.3688	0.0092	2.49	70000	0.4248	0.0108	2.54
400	0.2399	0.0027	1.13	5000	0.3782	0.0208	5.50	80000	0.4366	0.0103	2.35
546	0.3249	0.0492	15.14	5430	0.2898	0.0023	0.79	90000	0.4306	0.0072	1.68
600	0.3703	0.0149	4.03	5460	0.3955	0.0093	2.36	100000	0.4314	0.0059	1.37
800	0.2617	0.0126	4.83	5480	0.4536	0.0000	0.01	120000	0.4225	0.0121	2.86
880	0.3219	0.0045	1.41	5500	0.4462	0.0011	0.26	150000	0.4359	0.0042	0.97
1000	0.4066	0.0067	1.66	6000	0.4156	0.0029	0.69	200000	0.4362	0.0030	0.70
1200	0.2831	0.0267	9.42	7000	0.4398	0.0025	0.56	250000	0.4150	0.0276	6.64
1250	0.3687	0.0017	0.47	7800	0.4571	0.0007	0.14	300000	0.4378	0.0031	0.70
1600	0.3397	0.0118	3.47	8000	0.2165	0.0082	3.77	500000	0.4428	0.0018	0.40
2000	0.3756	0.0102	2.72	9000	0.3296	0.0310	9.42	1000000	0.3957	0.0114	2.87
				10000	0.4067	0.0097	2.40	1500000	0.4404	0.0050	1.13
				11000	0.4428	0.0053	1.19	2000000	0.4228	0.0045	1.07
				12000	0.3919	0.0152	3.87	3000000	0.4286	0.0043	1.00
				15000	0.3805	0.0117	3.08	4000000	0.4340	0.0043	1.00
				16000	0.3016	0.0128	4.26	5000000	0.4373	0.0051	1.17
				18000	0.4023	0.0185	4.61	10000000	0.4286	0.0047	1.10
				20000	0.4003	0.0124	3.11	20000000	0.4385	0.0027	0.62
				24000	0.3548	0.0107	3.03	50000000	0.4408	0.0037	0.85
				25000	0.4064	0.0086	2.12	100000000	0.4423	0.0015	0.34

Summary

The five tests described above—Jensen–Shannon divergence (main text), β–m self-consistency (Fig. SI 64), smoothing-window insensitivity (Table 1, Fig. SI 65), the Gini–Kolmogorov–Smirnov pair (Fig. SI 66), and the time-stability of β_i (Fig. SI 67, Table 2)—are methodologically distinct and probe complementary aspects of the fit, and all converge on the same conclusion: across all 4,536 denomination–month samples, Bitcoin UTXO ownership statistics are described by a single-parameter truncated geometric distribution to a precision that no multi-parameter phenomenological alternative matches in the self-consistency sense.