

Supplementary Material: Universality and Predictability of Technology Diffusion

Benjamin P. Wagenvoort^{1, 2}, François Lafond^{1, 2}, Joel Dyer^{1, 3}, and J. Doyne
Farmer^{1, 2, 4, 5}

¹Institute for New Economic Thinking at the Oxford Martin School, University of
Oxford

²Smith School of Environment and Enterprise at the School of Geography and the
Environment, University of Oxford

³Department of Computer Science, University of Oxford

⁴Santa Fe Institute

⁵Macrocosm Inc.

August 6, 2026

Summary of supplementary material

Reference material. This supplement also includes a list of abbreviations, a glossary, and tables of notation used throughout.

Methods. Section 1.1 introduces S-curves and their relevance for technology diffusion. Section 1.2 details our Bayesian forecasting procedure, the universal cross-technology prior, and the adaptive widening transform (Section 1.2.5) we use to recalibrate prediction intervals. Section 1.3 defines the probability integral transform (PIT) and the Q-Q calibration plot used to assess distributional calibration, the method behind main paper Figure 3(d-f). Section 1.4 defines the *empirical surrogate datasets* used as a Bertalanffy-Richards-based null benchmark in subsequent in- and out-of-sample analyses (Sections 3.1, 3.2, 3.3, 4.1). Section 1.5 discusses limitations of the Bayesian inference and the noise specification, including the connection to the post-hoc widening of prediction intervals introduced in Section 1.2.5.

Data. Section 2 describes the data sources and preprocessing steps used in our analysis. Section 2.1 discusses limitations.

In-sample data analysis. Section 3.1 defines the technology-invariant parameters employed in the model. Section 3.2 examines empirical regularities in technology diffusion across the cleaned dataset. Section 3.3 compares the logistic, Gompertz, and Bertalanffy-Richards S-curve models.

Out-of-sample data analysis. Section 4.1 presents calibration results on real-world data, highlighting the model’s point-forecast accuracy and probabilistic calibration, and compares performance against the logistic and Gompertz models (main paper Figure 3), as well as the Bass model. We provide a prior-sensitivity analysis and a robustness check of our calibration on a less filtered dataset. We also introduce a post-hoc widening transform (Section 1.2.5) that calibrates the modestly overconfident base-case prediction intervals. Section 4.2 presents backtesting results on adversarially chosen, historically challenging energy technologies (gas, hydropower, oil, nuclear), demonstrating how our model handles markets that grow at approximately 2% per year, and time-series subject to exogenous shocks (e.g., the 1979 oil shock and major nuclear accidents).

Application. Section 5.1 provides a backtesting analysis of solar and wind useful energy generation, shares, and capacity over a 15-year forecast horizon. Section 5.2 presents probabilistic forecasts for solar and wind through 2050 (some cases to 2060), comparing against the IEA World Energy Outlook scenarios, the IPCC AR6 database, and the NGFS scenarios, similar to those reported in the main paper (Figure 4(d-f)). We provide robustness analysis of IEA net-zero milestones for solar capacity. We show forecasts assuming asymptotic exogenous growth (2% per year). We also show forecasts under a pessimistic prior. Qualitative conclusions remain consistent.

Contents

List of Abbreviations	8
Glossary of Terms	10
Notation and Variables	12
1 Methods	16
1.1 Mathematical modeling of technology diffusion	16
1.1.1 S-curve models: The role of symmetry	16
1.1.2 Technology adoption	19
1.1.3 Stochastic noise model	19
1.1.4 Normalizing S-curves	20
1.1.5 Model limitations	20
1.2 Our Bayesian forecasting procedure	21
1.2.1 Introduction to Bayesian inference	21
1.2.2 Estimating parameters of our stochastic model	21
1.2.2.1 The Bertalanffy-Richards curve log-likelihood	22
1.2.3 Distributional forecasts	24
1.2.4 Implementation: forecasting procedure for a time-series	24
1.2.5 Adaptive widening of the predictive distribution	27
1.2.5.1 The widening transform	27
1.2.5.2 The diffusion-dependent exponent	27
1.2.5.3 Per-technology calibration procedure	28
1.2.5.4 Recommended practice	28
1.2.6 Software implementation: how to run the forecasting procedure	28
1.2.6.1 MCMC hyperparameters.	28
1.2.6.2 How to run the forecasting procedure.	29
1.3 Evaluating distributional forecast performance	31
1.4 Empirical surrogate datasets	32
1.5 Limitations	34
2 Data	36
2.0.1 Datasets	36
2.0.2 Shares and non-shares	37
2.0.3 Data cleaning	39
2.0.3.1 Computing parameters for cleaning	39
2.0.3.2 Step 1: Removing death phase data	40
2.0.3.3 Step 2: Enforcing minimum data requirement	41
2.0.3.4 Step 3: Requiring complete S-curve coverage	41
2.0.3.5 Step 4: Selecting unique technologies and removing poor-quality data	42
2.0.3.6 Step 5: Manual identification and removal of outliers	45
2.0.3.7 Step 6: Removing highly correlated time-series	47
2.0.4 Behavior of the clean data	50
2.1 Limitations	53
3 In-sample data analysis	55

3.1	Defining technology-invariant parameters	55
3.1.1	Asymmetry parameter	55
3.1.1.1	In-sample fit and choice of asymmetry parameter	55
3.1.1.2	Growth replication	58
3.1.2	Moving average parameter	60
3.1.3	Terminal residual	62
3.1.4	Summary and limitations	62
3.2	Testing for empirical regularities	63
3.2.1	Do technologies grow exponentially?	63
3.2.2	Replicating observed growth behavior	65
3.2.2.1	Collapse plot of technology adoption curves	67
3.2.3	Other regularities in technology diffusion from literature	70
3.3	Comparison of S-curve models	72
3.3.1	Theoretical differences between Gompertz and logistic curves	72
3.3.1.1	Goodness of fit	73
3.3.2	Ensemble growth behavior under different S-curve assumptions	73
3.3.2.1	Replicating ensemble growth behavior via simulation	74
3.3.3	Ensemble averaging effects	76
3.3.4	Comparing Gompertz, logistic, and B-R collapse plots	79
4	Out-of-sample data analysis	82
4.1	Testing forecast performance on empirical data	82
4.1.1	Methods underlying main paper Figure 3	82
4.1.1.1	Priors	82
4.1.1.2	Logistic and Gompertz models	84
4.1.1.3	Measuring point forecast errors	85
4.1.1.4	Distributional calibration at each diffusion target	85
4.1.2	Backtest results	86
4.1.2.1	Point-forecast accuracy	86
4.1.2.2	Calibration results	87
4.1.2.3	Interpreting the results	88
4.1.3	The Bass model	90
4.1.4	Calibrating the widening transform on backtesting data	90
4.1.5	Robustness to subjective data cleaning	93
4.1.6	Further robustness checks	95
4.1.7	The role of the prior: bounding forecast quality	96
4.2	Backtest on energy technologies	98
4.2.1	Historical energy demand growth	98
4.2.2	Comparing backtest performance of three approaches	99
4.2.2.1	Modeling market growth explicitly	100
4.2.2.2	Comparing the three approaches.	101
4.2.3	Exogenous disruptions: nuclear energy	103
5	Application to solar and wind	105
5.1	Backtesting solar and wind time-series	105
5.2	Forecasting solar and wind useful energy	108
5.2.1	IEA net-zero milestones for solar capacity	108
5.2.1.1	Conditional milestone-year analysis	108
5.2.1.2	Distributional forecast for solar PV capacity	110
5.2.2	From capacity to useful energy	111
5.2.3	Forecasts for useful energy	112

5.2.3.1 Setup	112
5.2.3.2 Analyzing the growth of solar and wind	113
5.2.3.3 Comparing forecasts to scenarios	114
5.2.4 A pessimistic prior	121
References	123

List of Figures

1	Comparison of the B-R, logistic, and Gompertz curves	18
2	Adjusted R^2 distribution for raw data	39
3	Normalized residuals after cleaning Steps 2 and 3	42
4	Examples of removed time-series	44
5	Comparison of adjusted R^2 and normalized residuals before and after Step 4(b) . .	45
6	Technologies with outliers	46
7	Cane sugar consumption in Japan with outliers removed	47
8	Sector representation in raw and cleaned dataset	51
9	Adjusted R^2 and normalized residuals of cleaned dataset	52
10	R^2 comparison across models	56
11	Replication of estimated β_i distribution	58
12	Testing technology-invariant β values for growth replication	59
13	Replication of estimated MA(1) parameter distribution	62
14	Testing for exponential growth at early diffusion stages	64
15	log-differences of early diffusion for technologies with a significant slope	65
16	Collapse plot of log-differences on technology time series	67
17	Collapse plots of the data	68
18	Collapse plot of ensemble and surrogate data	69
19	Growth rate vs. time	70
20	Growth rate vs. uncertainty	72
21	Comparison of R^2 values across Gompertz, logistic, and Bertalanffy-Richards curves	73
22	Collapse plots: Gompertz versus logistic growth	76
23	Comparison of shifted vs non-shifted exponential distributions	77
24	Ensemble averaging effects	79
25	(G)FP-transform collapse plots: three S-curve models	80
26	Priors used in the main text (universal prior)	83
27	MedALE by forecast horizon	86
28	Baseline and widened forecast calibration	88
29	Per-model widened calibration across forecast origins	92
30	Forecast accuracy on the 154-technology dataset	94
31	Calibration on the 154-technology dataset	95
32	Historical energy consumption	99
33	Historical useful energy generation for selected technologies	99
34	Backtest comparison examples for three forecasting methods on energy technologies	102
35	Backtest for oil	103
36	Backtest for nuclear energy	104
37	Useful energy share of nuclear electricity	104
38	15-year backtest results for solar and wind	106
39	15-year backtest results for solar and wind using asymptotic growth S-curve	106
40	15-year backtest results for solar and wind electricity shares	107
41	Conditional milestone-year analysis for solar PV capacity	109
42	Distributional forecast for solar PV installed capacity	111
43	Comparison of solar useful-energy forecasts from two modelling routes	112
44	Historical useful energy generation for solar PV and wind energy	114
45	Forecast for solar PV useful energy generation and share	115
46	Unwidened distributional forecast for solar PV and wind useful energy generation .	116
47	Comparison with the NGFS Phase 5 scenario ensemble	117
48	Posterior distribution of current diffusion for solar PV and wind	118
49	Exogenous growth forecast for solar PV and wind useful energy generation	120
50	Forecast for solar PV and wind useful energy generation under a pessimistic/optimistic prior	121
51	Comparing prior and posterior predictive distributions under a pessimistic prior . .	122

List of Tables

1	List of abbreviations	9
2	Glossary of terms	11
3	Technology diffusion variables	12
4	Noise and stochastic-process parameters	12
5	Transformations and normalised variables	12
6	Statistical estimators, vectors, and matrices	13
7	Distributional and Bayesian notation	13
8	Prior hyperparameters	13
9	Forecasting variables	14
10	Calibration and error metrics	14
11	Indices and counts	15
12	Subscript and superscript conventions	15
13	Key equations and transformations	15
14	Distribution of shares and non-shares by sector	38
15	Statistical tests comparing shares vs. non-shares	38
16	Summary of the data cleaning procedure.	39
17	List of 120 technology time-series	47
18	Sector representation in cleaned dataset	51
19	Statistics of normalized residuals at different cleaning stages	53
20	Comparison of estimated β_i distributions between empirical and surrogate data . .	57
21	Statistical tests comparing ICT vs non-ICT technologies	71
22	Comparison of three exponential models for the growth-rate distribution. Figure 23 plots the shifted exponential and the non-shifted (independent) models; the non-shifted (same rate) is a nested model included here only for the statistical comparison.	78
23	Cross-tech median ALE and pooled bias by model and forecast origin	87
24	Pointwise ALE by model: cross-tech median vs mean at five diffusion targets . . .	87
25	Mechanism check: cross-technology MedALE	89
26	Widening exponent w^* by forecast origin: asymmetric vs symmetric	91
27	Per-model widening across forecast origins	91
28	Leave-one-technology-out cross-validation of the widening exponent (120-tech) . .	91
29	Widening exponent w^* on the empirical surrogate	92
30	Widening exponent comparison between 120-technology and 154-technology back-testing datasets	93
31	Leave-one-technology-out cross-validation of the widening exponent (154-tech) . .	94
32	Diffusion-aware prior hyperparameters by diffusion band	97
33	Validation of conditional ML milestone-year method	109
34	Estimated year of reaching IEA NZE milestones for solar PV capacity	110

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ALE	Absolute Log Error, $ \ln(\hat{Y}/Y) $ (cross-technology median used in main paper Fig. 3a, b; cross-technology mean reported as robustness in Table 24)
AR6	IPCC Sixth Assessment Report
ARIMA	AutoRegressive Integrated Moving Average
ATM	Automated Teller Machine
B-R	Bertalanffy-Richards
BIC	Bayesian Information Criterion
CCS	Carbon Capture and Storage
CDF	Cumulative Distribution Function
CF	Capacity Factor
CHAT	Cross-country Historical Adoption of Technology dataset
CI	Confidence Interval
CV	Coefficient of Variation
DGP	Data-Generating Process
FP	Fisher–Pry
GDP	Gross Domestic Product
GFP	Generalized Fisher–Pry
HATCH	Historical Adoption of Technology dataset
ICT	Information and Communications Technology
IEA	International Energy Agency
IPCC	Intergovernmental Panel on Climate Change
IQR	Interquartile Range
IRENA	International Renewable Energy Agency
KDE	Kernel Density Estimate
K-S, KS	Kolmogorov–Smirnov
L_1, L_2	Absolute and squared error norms
MA(1)	Moving Average of order 1
MALE	Mean Absolute Log Error: cross-technology mean of the absolute log error (ALE) at a diffusion target; (SI Section 4.1)
MCMC	Markov Chain Monte Carlo
MedALE	Median Absolute Log Error: cross-technology median of the absolute log error (ALE) at a diffusion target, the primary error metric (main paper Fig. 3a, b; SI Section 4.1)
MLE	Maximum Likelihood Estimation
NLLS	Non-Linear Least Squares
NZE	Net-Zero Emissions (IEA scenario)
OLS	Ordinary Least Squares
OWID	Our World in Data
PDF	Probability Density Function
PIT	Probability Integral Transform
PV	Photovoltaic
PWh	Petawatt-hour (10^{15} Wh)
Q-Q	Quantile–Quantile (plot)
R^2	Coefficient of Determination
SD	Standard Deviation
SI	Supplementary Information
SSE	Sum of Squared Errors
SSR	Sum of Squared Residuals
TSS	Total Sum of Squares

Abbreviation	Definition
TW	Terawatt (10^{12} W)
TWh	Terawatt-hour (10^{12} Wh)

Table 1: List of abbreviations used throughout this document.

Glossary of Terms

Term	Meaning
<i>Curves and growth</i>	
S-curve	Sigmoidal $Y(t)$ for cumulative adoption (logistic, Gompertz, B-R).
B-R curve	Asymmetric S-curve with shape β
Logistic curve	Symmetric S-curve; B-R at $\beta=1$.
Gompertz curve	S-curve with exponentially decaying log-growth.
Inflection point	Point of maximum growth rate on the S-curve.
Asymptote (L)	Saturation level.
Diffusion path	Time-series $\{Y_{i,t}\}_t$ of a technology.
Median growth trajectory	Cross-tech median growth-rate sequence indexed by diffusion (not a smooth curve).
Growth rate	Log time-derivative $d \log Y / dt$; empirically $\Delta \log Y_{i,t}$.
Intrinsic growth rate (k)	Growth rate as $t \rightarrow -\infty$.
Death phase	Post-saturation decline; excluded in cleaning.
Stocks vs. flows	Cumulative adoption vs. per-period production.
Shares vs. non-shares	Diffusion as $[0, 1]$ market fraction vs. absolute level.
<i>Transformations and noise</i>	
GFP transform	Generalised Fisher–Pry; linearises the B-R curve.
FP transform	Fisher–Pry logit; GFP at $\beta=1$.
Shock ($\epsilon_{i,t}$)	Standardised Gaussian innovation in GFP space.
Shifted exponential distribution	$\xi + \text{Exp}(s)$; location ξ , rate s . Used for the cross-tech distribution of k_i .
Terminal residual ($\epsilon_{i,T}$)	Last in-sample shock; fixed at 0 universally.
Saturation-noise misspecification	GFP-additive noise vanishes as $d \rightarrow 1$ while empirical noise persists.
Comin effect	Faster-than-exponential early-diffusion growth at $d \lesssim 0.5\%$ [1].
<i>Models, parameters, surrogates</i>	
Technology-invariant parameter	Held equal across techs ($\beta=2/3$, $\rho=0.80$, $\epsilon_T=0$).
Asymmetry parameter (β)	Controls S-curve skewness.
Empirical surrogate dataset	Synthetic ensemble from B-R DGP at empirical-fit parameters; null benchmark.
Collapse plot	Pools normalised technologies onto a universal curve.
Ensemble averaging artifact	Apparent Gompertz decay from pooling heterogeneous B-R series.
Fixed-asymptote model	B-R with constant L .
Growing-market model	B-R with $L_i e^{gt}$ (exogenous market growth).
Shares model	B-R applied to market shares.
Adversarial energy backtest	Forecast of oil, gas, hydropower, and nuclear, chosen to expose B-R failure modes.
<i>Forecasting and calibration</i>	
Backtesting	Train to a chosen origin; compare forecasts to held-out data.
Forecast origin	Training end-point (diffusion or year).
Forecast horizon (τ, h)	Steps ahead being forecast.
Shifted forecast	Post-hoc GFP shift forcing zero last-training residual.
Nonshifted forecast	Direct posterior trajectory; no shift.
Distributional forecast	Predictive distribution from posterior sampling.
Posterior predictive distribution	Distribution of future observables marginalised over posterior parameter uncertainty.
Posterior median	50th percentile; L_1 -optimal point estimate.
Diffusion bin	Discrete diffusion-level interval (default 10%-wide) used for averaging ALE.

Term	Meaning
Probabilistic calibration	PIT values uniform on $[0, 1]$.
Underdispersion / overconfidence	Intervals too narrow; PIT mass in tails.
Overdispersion / underconfidence	Intervals too wide; PIT mass near centre.
Pointwise floor	Expected ALE at a single d under correct parameters.
Diffusion-averaged floor	Pointwise floor averaged over $d \in [0.05, 0.95]$.
<i>Priors</i>	
Universal prior	Single hyperparameter set, origin-independent (main-text).
Wide prior	Near-uninformative; floor benchmark.
Diffusion-aware prior	Hyperparameters vary by diffusion band; ceiling benchmark.
Empirical Bayes prior	Hyperparameters set from data.
Pessimistic prior	Favours low L / early saturation; sceptical scenario.
Well-specified prior	Prior equals the true distribution.
Weakly informative prior	Wider than the truth.
Strongly informative prior	Narrower than the truth.
<i>Calibration corrections</i>	
Widening transform	Post-hoc median-preserving power transform that widens intervals.
Sqrt-lower widening	Exponent w on upper tail, $\sqrt{w}$ on lower tail.
Adaptive widening	Diffusion-dependent w^* interpolation.
Oracle widening	Per- d optimal w^* ; calibration upper bound.
<i>Energy applications</i>	
Useful energy	Service-level energy after end-use losses.
Final energy	Energy delivered to end-users.
Secondary energy	Energy after primary-to-electricity conversion.
Capacity factor (CF)	Actual / max-potential output of capacity.
IEA NZE milestones	IEA Net-Zero-by-2050 targets at 2030/35/40/50.
Conditional milestone-year analysis	Invert B-R given L to find year of milestone.
<i>Other</i>	
Empirical regularity	Cross-tech pattern reproduced by the DGP.
Selection bias (in dataset)	Over-representation of completed S-curves.

Table 2: Glossary of terms used throughout this document. Mathematical symbols are defined in the Notation section below.

Notation and Variables

Technology Diffusion Variables

Symbol	Description	Type
$Y_{i,t}$	Observed diffusion of technology i at time t	Variable
$Y_{i,1:T}$	Time-series vector $(Y_{i,1}, \dots, Y_{i,T})^\top$	Vector
$Y_i^{(n)}(t), \tilde{Y}_{i,t}$	Normalised diffusion	Variable
L_i	Asymptotic diffusion (saturation)	Parameter
$\tilde{L}_i$	Normalised asymptote $L_i/y_{i,T_{\max}}$	Parameter
$d_i(t)$	Empirical diffusion share $Y_{i,t}/\hat{L}_i$	Variable
$\hat{d}_i$	Estimated diffusion stage at training end, $1/\tilde{L}_i^{\text{med}}$	Estimator
t	Time (years)	Variable
$t_{0,i}$	Inflection-related time offset	Parameter
k_i	Intrinsic growth rate	Parameter
β	Asymmetry parameter	2/3 univ.
β_i	Per-tech asymmetry parameter	Parameter
$\Delta \log Y_{i,t}$	Log-difference $\log Y_{i,t} - \log Y_{i,t-1}$	Variable
h	Time step; also forecast horizon (context-dependent)	Typically 1 yr

Table 3: Technology diffusion variables. Continuous-time form $Y(t)$ in §1, discrete form $Y_{i,t}$ thereafter.

Noise and Stochastic Process

Symbol	Description	Type
$\epsilon_{i,t}, \epsilon_{i,t}$	Standardised Gaussian shock	$\mathcal{N}(0, 1)$
σ_i	Per-tech shock SD in GFP space	Parameter
$\sigma_{\epsilon,i}$	MA(1) marginal SD, $\sigma_i \sqrt{1 + \rho^2}$	Derived
ρ	MA(1) lag-1 autocorrelation	0.80 univ.
ρ_i	Per-tech autocorrelation	Parameter
$\epsilon_{i,T}$	Terminal residual; fixed at 0	0 univ.
$u_{i,t}$	White-noise innovation	$\mathcal{N}(0, 1)$

Table 4: Noise and stochastic-process parameters.

Transformations and Normalised Variables

Symbol	Description	Type
$f(Y_{i,t}; L_i, \beta)$	GFP transformation $\beta^{-1} \log[(L_i/Y_{i,t})^\beta - 1]$	Function
$x_{i,t}$	Centred latent, $f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i})$	Variable
$y_{i,t}$	Normalised diffusion ($y_{i,T_{\max}} = 1$)	Variable
$R_{i,t}$	Raw residuals	Variable
$e_{i,t}$	Normalised residuals	Variable
$s_i(R)$	RMS of raw residuals (with d.o.f. correction)	Statistic
r_{ij}	Pearson correlation between technologies	Statistic
$t^{(n)}, \tau_i$	Normalised (dimensionless) time, e.g. $\hat{k}_i(t - \hat{t}_{0,i})$	Variable

Table 5: Transformations and normalised variables.

Statistical Estimators and Vectors

Symbol	Description	Type
θ_i	Parameter set $\{L_i, k_i, t_{0,i}, \sigma_i\}$	Parameter
$\hat{\theta}_i$	Per-tech parameter estimates	Estimator
$\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i$	Individual estimates	Estimator
$\hat{\sigma}_i^{a-b}$	$\hat{\sigma}_i$ on GFP residuals restricted to diffusion window a – $b\%$ (primary window 1–95)	Estimator
$\hat{\beta}_i$	Estimated asymmetry parameter	Estimator
$\hat{k}_{\text{ssr},i}$	SSR-based growth-rate estimator	Estimator
$\hat{\epsilon}_{i,t}$	Estimated GFP-space residual	Variable
$\bar{R}_i$	Sample mean of raw residuals	Statistic
X_i	GFP latent vector $(x_{i,1}, \dots, x_{i,T})^\top$	Vector
$\tilde{X}_i$	Modified vector when $\epsilon_{i,T}$ is parameterised	Vector
V_i	MA(1) covariance matrix	Matrix
V_i^*	Modified covariance with $\epsilon_{i,T}$ extension	Matrix

Table 6: Statistical estimators, vectors, and matrices.

Distributional and Bayesian Notation

Symbol	Description	Type
$\pi(\theta)$	Prior distribution	Distribution
$\pi(\theta \mid Y_{1:T})$	Posterior distribution	Distribution
$p(Y \mid \theta)$	Likelihood	Distribution
$\mathcal{N}(\mu, \sigma^2)$	Normal	Distribution
$\Gamma(\alpha, \ell, s)$	Three-parameter Gamma: shape α , lower bound ℓ , scale s	Distribution
$\text{Exp}(\lambda)$	Exponential, rate λ	Distribution
ξ, s, s'	Location and rate parameters of (shifted) exponential fit on k_i	Parameters
$\text{Uniform}(a, b)$	Uniform on $[a, b]$	Distribution
$\mathbb{E}[\cdot]$	Expectation	Operator
$\mathbb{P}_q[\cdot]$	q -th percentile (e.g. $\mathbb{P}_{50}$ = median)	Operator
$\mathbf{P}(\cdot)$	Generic data-generating distribution	Distribution
F_i	Predictive (posterior) CDF of technology i	Function
$u_i(d^*)$	PIT value $F_i(Y_i(t_i^*))$ at diffusion target d^*	Variable

Table 7: Distributional and Bayesian notation. Γ uses (α, ℓ, s) throughout.

Prior Hyperparameters

Symbol	Description
α_L, ℓ_L, s_L	Hyperparameters of Γ prior on $\tilde{L}_i$
$\alpha_{t_0}, \ell_{t_0}, s_{t_0}$	Hyperparameters of Γ prior on $\tilde{t}_{0,i}$
$\tilde{t}_T$	Index of last observed time

Table 8: Prior hyperparameters. Numerical values for the universal, wide, and diffusion-aware priors are in §4.1.

Forecasting Variables

Symbol	Description	Type
τ, h	Forecast horizon (steps ahead)	Variable
$\boldsymbol{\tau}$	Forecast time vector $(T + 1, \dots, T + \tau)^\top$	Vector
$\boldsymbol{\epsilon}_{T+1:T+\tau}$	Forecast-period MA(1) noise	Vector
$\mathbf{Y}_m(\tau)$	Forecast trajectory for posterior sample m	Vector
$Y_{i,T+\tau}$	Forecast at horizon $T + \tau$	Variable
g	Exogenous market-growth rate (growing-market model)	Parameter
M_{t^*}	IEA milestone capacity at year t^* (§5.2)	Variable
t^*	Year at which diffusion reaches a milestone	Variable
η_{t^*}, σ_η	Time-domain noise term and its SD; $\sigma_\eta = \hat{\sigma}\sqrt{1 + \rho^2}$	Variable / Parameter
$\hat{\phi}$	Lag-1 autocorrelation in the Farmer–Lafond exponential-deviation forecast variance (distinct from ρ)	Estimator

Table 9: Forecasting variables and vectors.

Calibration and Error Metrics

Symbol	Description
d	Diffusion level (forecast origin or evaluation point)
b	Diffusion-bin index
$\mathcal{T}_{i,b}, n_{i,b}$	Time-index set and count of tech i observations in bin b
$\ell_{i,t}$	Absolute log-error $\log(\hat{Y}_{i,t}/Y_{i,t})$
$\overline{\text{ALE}}_{i,b}$	Mean absolute log-error for tech i in bin b
MALE_b	Median across techs of $\overline{\text{ALE}}_{i,b}$
$\delta L_i, \delta k_i, \delta t_{0,i}$	Parameter biases (estimate minus true) used in first-order ALE decomposition
$L_0, U(h), S(\sigma)$	First-order error decomposition: level $ \delta L_i/L_i $, curvature $q(d) \delta k_i h - k_i \delta t_{0,i} $, shock $q(d)\sigma_{\epsilon,i}\sqrt{2/\pi}$
$\mathcal{E}(h)$	Quadrature scale $\sqrt{L_0^2 + U(h)^2 + S(\sigma)^2}$
$q(d)$	Diffusion prefactor $1 - d^\beta$
D_{KS}	Kolmogorov–Smirnov statistic
$D_{KS}(\beta), D_{KS}(\rho)$	KS distance varying β or ρ
w, w^*	Widening exponent and calibration optimum
$\hat{d}_i$	Estimated diffusion stage for w^* interpolation
$\tilde{L}_i^{\text{med}}$	Posterior median of $\tilde{L}_i$, $\mathbb{P}_{50}[\pi(\tilde{L}_i Y_{i,1:T})]$

Table 10: Calibration and error metrics.

Indices and Counts

Symbol	Description	Value
i	Technology index ($i = 1, \dots, N$)	Index
N	Number of technologies in ensemble	120 (or 154)
$T, T_i, T_{\max}$	Final / per-tech / max time indices	Variable
M	Surrogate datasets per analysis	100–1000
N_{samples}	MCMC posterior samples per technology	10 000
s	Surrogate index ($s = 1, \dots, M$)	Index

Symbol	Description	Value
m	Posterior-sample index	Index
b	Diffusion-bin index	Index

Table 11: Indices and counts. M (surrogate count) and N_{samples} (MCMC count) are distinct.

Subscript and Superscript Conventions

Notation	Meaning
$\cdot_i$	Technology-specific (e.g. k_i, L_i)
$\cdot_{i,t}$	Tech i at discrete time t
$\cdot_{i,1:T}$	Time-series vector
$\cdot(t)$	Continuous-time generic notation (used in §1)
$\hat{\cdot}$	Estimated quantity
$\tilde{\cdot}$	Normalised or modified quantity
$\bar{\cdot}$	Sample mean
$\cdot^{(n)}$	Normalised
$\cdot^{(l)}, \cdot^{(g)}$	Logistic / Gompertz model variant
$\cdot^{(s)}$	Surrogate-simulation index
$\cdot_m$	Posterior-sample index
$\cdot, \cdot$	Vector or matrix

Table 12: Subscript and superscript conventions.

Key Equations and Transformations

Name	Equation
B-R curve	$Y(t) = L [1 + \exp(-\beta k(t - t_0))]^{-1/\beta}$
Logistic curve	$Y(t) = L / [1 + \exp(-k(t - t_0))]$
Gompertz curve	$Y(t) = L \exp(-\exp(-k(t - t_0)))$
GFP transformation	$f(Y_{i,t}; L_i, \beta) = \beta^{-1} \log[(L_i/Y_{i,t})^\beta - 1]$
FP transformation	$f(Y_{i,t}; L_i) = \log[(L_i - Y_{i,t})/Y_{i,t}] \quad (\text{GFP}, \beta=1)$
MA(1) noise	$\sigma_i \epsilon_{i,t} + \rho \sigma_i \epsilon_{i,t-1}$
Centred latent	$x_{i,t} \equiv f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i})$
Log-difference	$\Delta \log Y_{i,t} = \log Y_{i,t} - \log Y_{i,t-1}$
Analytical ALE floor	$\mathbb{E}[\text{ALE} \mid d] = (1 - d^\beta) \sigma_{\epsilon,i} \sqrt{2/\pi}$

Table 13: Key equations and transformations.

1

Methods

1.1 Mathematical modeling of technology diffusion

This section introduces and motivates the use of S-curves, providing an overview of their applications and the model we use to make well-calibrated forecasts.

Background The energy transition is a technological transition [2, 3], characterized by phasing out fossil fuels and adopting renewables. Technological adoption is often modelled as an *S-curve*. Throughout this SI, we use the term for any sigmoidal $Y(t)$ describing cumulative adoption — for example, the logistic, Gompertz, and Bertalanffy-Richards forms introduced below (Eqs. (2)–(3)). Technological shifts occur throughout history due to necessity or innovation [4]. Examples include shifts from horses to tractors in agriculture, diesel locomotives replacing steam engines, and transitions from wood to coal to oil to renewables for energy production. Understanding historical diffusion dynamics across industries helps us model and forecast renewable energy adoption.

Technology diffusion can be measured in stocks or flows. The flow measures production or sales of a technology per unit time—for example, the number of mobile phones sold in a given year. Technology stocks measure the *cumulative* production of a technology, minus retirements—for example, the number of mobile phones currently in operation. Stocks are typically the preferred measure for modeling technological diffusion [5, 6].

Griliches introduced S-curves to model technology stocks in 1957 [5]. Researchers have since used S-curves extensively [7–10] and developed many S-curve models with increasing complexity, often introducing additional parameters to improve fits [11–16].

1.1.1 S-curve models: The role of symmetry

Consider the generalized logistic differential equation.

$$\frac{dY(t)}{dt} = kY(t)^\alpha \left[1 - \left(\frac{Y(t)}{L} \right)^\beta \right]^\gamma, \quad (1)$$

where $\frac{dY(t)}{dt}$ describes the technology flow, $Y(t)$ is the stock at time t , L is the asymptotic stock as $t \rightarrow \infty$, and k is the *intrinsic growth rate*. Throughout this SI, *growth rate* refers to the log time-derivative $d \log Y / dt$, with empirical analogue $\Delta \log Y_{i,t} = \log Y_{i,t} - \log Y_{i,t-1}$ (see Section 3.2.1); k is the asymptotic value of this rate as $t \rightarrow -\infty$. The parameters α , β , and γ determine the shape of the S-curve. In particular, β controls the steepness and affects the symmetry of the function. When $\alpha = \gamma = 1$, Eq. (1) reduces to the Bertalanffy-Richards (B-R) curve [17–19], described as

$$\frac{dY(t)}{dt} = kY(t) \left[1 - \left(\frac{Y(t)}{L} \right)^\beta \right]. \quad (2)$$

When $\beta > 1$, the inflection point occurs after 50% adoption, and when $\beta < 1$, it occurs before 50% adoption. The analytical solution to Eq. (2) is

$$Y(t) = \frac{L}{(1 + \exp(-\beta k(t - t_0)))^{1/\beta}}. \quad (3)$$

The logistic ($\beta = 1$) and Gompertz (divide by β and take $\beta \rightarrow 0$ limit) curves are special cases [20]. Throughout, t_0 is a *location parameter* (typically in years) that shifts the curve along the time axis. For the symmetric logistic and the Gompertz limit, t_0 coincides with the inflection point: the time of maximum growth, or for technology applications, the year of peak annual addition. For the asymmetric B-R curve with $\beta \neq 1$, the inflection shifts to $t_0 - \log(\beta)/(\beta k)$ (Section 3.1), so t_0 remains a curve offset but no longer marks peak growth.

The logistic curve. When $\beta = 1$, Eq. (2) reduces to the symmetric logistic curve

$$\frac{dY(t)}{dt} = kY(t) \left[1 - \frac{Y(t)}{L} \right], \quad (4)$$

first introduced by Verhulst [21] in 1838 to model population growth. Integrating by separation of variables, we obtain the analytical solution

$$Y(t) = \frac{L}{1 + A \exp(-kt)}, \quad (5)$$

where A is a constant depending on the initial condition. The parameter A defines a shift in the curve relative to the point of maximum growth. Let A take the form

$$A = \exp(kt_0), \quad (6)$$

so that maximum growth occurs exactly when $t = t_0$, at $Y(t_0) = L/2$, recovering the location-parameter interpretation introduced above for the logistic case.

Substituting Eq. (6) into Eq. (5), we derive

$$Y(t) = \frac{L}{1 + \exp(-k(t - t_0))}. \quad (7)$$

We refer to this as the logistic curve. Applications include disease modeling [22, 23], predicting gender from weight and height [24], and ecology and cancer modeling [6, 25–27].

The Gompertz curve. Starting from Eq. (2), dividing by β and taking the limit as $\beta \rightarrow 0$, we obtain the Gompertz curve,

$$Y(t) = L \exp(-\exp(-k(t - t_0))). \quad (8)$$

Gompertz introduced this curve in 1825 [28] to model population dynamics. Researchers have since applied the model extensively to plant growth [29, 30], fish growth [31], tumor growth [32], bacterial growth [33], and other applications [34–39].

Comparing the three S-curves. S-curves can overfit, yielding inconsistent out-of-sample performance [40]. We therefore fix $\beta = 2/3$ as a technology-invariant value, balancing the early-exponential property of the logistic with the asymmetry of the Gompertz without adding free parameters.

To characterize growth behavior, we compute log-growth rates $\frac{d}{dt} \log Y(t)$. For the B-R curve (Eq. (3)),

$$\frac{d}{dt} \log Y(t) = \frac{k}{1 + \exp(\beta k(t - t_0))}, \quad (9)$$

where setting $\beta = 1$ yields the logistic case and $\beta = 2/3$ yields our B-R case, with limiting behavior

$$\lim_{t \rightarrow -\infty} \frac{d}{dt} \log Y(t) = k, \quad \lim_{t \rightarrow \infty} \frac{d}{dt} \log Y(t) \sim k \exp(-\beta k(t - t_0)). \quad (10)$$

For the Gompertz curve, the log-growth rate takes a different form:

$$\frac{d}{dt} \log Y^{(g)}(t) = k \exp(-k(t - t_0)), \quad (11)$$

with limiting behaviour

$$\lim_{t \rightarrow -\infty} \frac{d}{dt} \log Y^{(g)}(t) = \infty, \quad \lim_{t \rightarrow \infty} \frac{d}{dt} \log Y^{(g)}(t) \sim k \exp(-k(t - t_0)). \quad (12)$$

The B-R and logistic curves exhibit exponential growth at early diffusion (constant log-growth rate k), while the Gompertz curve has no constant early-growth regime. At late diffusion, all three decay exponentially, at rate βk for the B-R curve (rate k when $\beta = 1$) and rate k for the Gompertz curve. In particular, the logistic ($\beta = 1$) and Gompertz have identical late-stage decay.

Figure 1 accompanies main paper Figure 1. It shows the B-R curve ($\beta = \frac{2}{3}$), logistic ($\beta = 1$), and Gompertz curves in levels (Figure 1a), log-differences against time (Figure 1b), and log-differences against percentage diffusion (Figure 1c). All curves use $L = 100$ and $t_0 = 0$. We set $k = 1$ for the B-R curve and adjust the logistic ($k = 0.749$) and Gompertz ($k = 0.526$) growth rates so that all three curves spend the same time traversing from 10% to 90% diffusion.

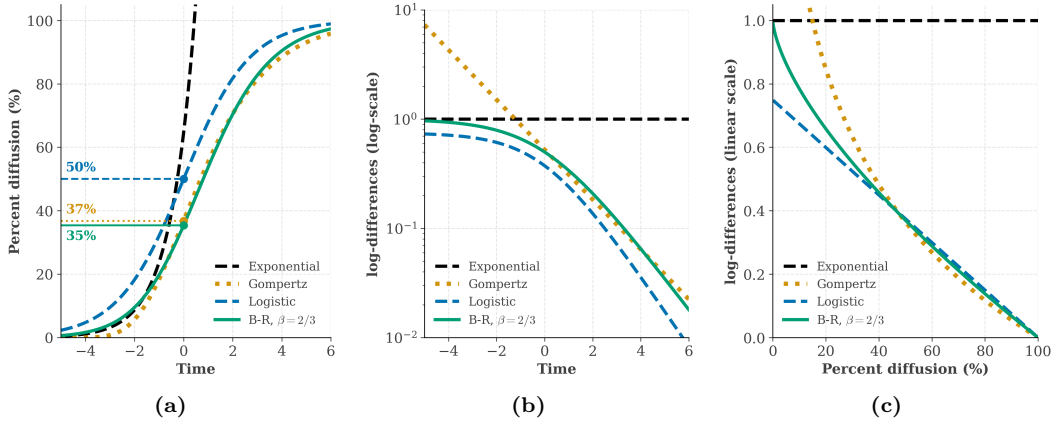


Figure 1: Comparison of the B-R curve with $\beta = \frac{2}{3}$ (green) to the logistic ($\beta = 1$, blue) and Gompertz (red) curves. Exponential growth (black) is plotted for reference. All curves use $L = 100$ and $t_0 = 0$; growth rates are $k = 1$ (B-R), $k = 0.749$ (logistic), and $k = 0.526$ (Gompertz), chosen so that time between 10–90% diffusion is the same across curves. (a) S-curves in levels; percent diffusion at $t = 0$ is highlighted. (b) Log-differences on log-scale against time. (c) Log-differences on linear scale against percent diffusion.

The three curves differ in two key respects. (1) *Inflection point*: for the logistic and Gompertz curves, $t = t_0$ defines the inflection point (point of maximum growth), occurring at exactly 50% and approximately 37% diffusion, respectively. For any asymmetric B-R curve ($\beta \neq 1$), the inflection point shifts to $t = t_0 - \frac{\log \beta}{\beta k}$. At $t = 0$, the B-R curve with $\beta = 2/3$ reaches approximately 35% adoption and is still accelerating, while the logistic and Gompertz curves begin to decelerate. For $k = 1$, $\beta = 2/3$, the B-R curve continues accelerating until $t \approx 0.6$, corresponding to approximately 46% adoption (Figure 1a). (2) *Early growth*: under the log transform, the logistic curve grows at an approximately constant rate k for $t \ll t_0$ (approximate exponential growth) and then decays to zero, while the Gompertz curve’s log-growth decays exponentially throughout. Figure 1b plots exponential growth at rate $k = 1$ (black) for reference. The logistic curve runs parallel to, but below, the exponential line because its adjusted growth rate is $k = 0.749 \neq 1$; this parallel structure confirms approximate exponential growth at early diffusion levels. The Gompertz curve, by contrast, decreases linearly with time on the log-scale. Figure 1c shows that the logistic curve’s growth decreases linearly with percentage diffusion, while the B-R curve’s growth decreases more gradually, lying between the logistic and Gompertz. All three curves show similar exponentially decaying growth rates in later stages, evident from the overlapping lines at high diffusion in Figure 1c.

Empirically, neither the logistic nor Gompertz model dominates. Comparative studies find results vary by technology, time period, and domain [41–48]. Comin et al. [1] find the logistic

curve systematically fails to capture early diffusion dynamics; we later show this also applies to the Gompertz curve under ensemble averaging. Each model captures one desirable property but not the other: many technologies exhibit approximate exponential growth at early diffusion levels [13, 40, 49–51] (logistic, not Gompertz), and many diffusion processes are asymmetric [6, 52, 53] (Gompertz, not logistic). Meade and Islam [53] compared 22 S-curve formulations and found curve symmetry matters more than functional form. The B-R curve combines both properties: it preserves exponential growth at early diffusion levels while allowing asymmetry [54, 55].

1.1.2 Technology adoption

The S-curve has three phases: an *initial* slow-growth phase as the technology enters the market, a *growth* phase of rapid expansion, and a *saturation* phase as adoption plateaus [6, 27, 56].

Phase-out of technologies. As technologies mature, better substitutes often emerge [6, 57]. The mature technology then declines, entering a fourth *death phase* as the substitute takes substantial market share. We do not model phase-out and therefore omit the death phase from our analysis. For the first three phases, the S-curve formulation suffices to describe and predict adoption.

1.1.3 Stochastic noise model

The Bertalanffy-Richards curve in Eq. (3) is deterministic. We extend it to a stochastic model below.

Transforming the S-curve. Fisher and Pry [58] introduced a transformation that linearises the logistic curve, later extended by Bhargava [59] with a time-dependent growth parameter. We generalise this transformation to the B-R curve and define the generalised Fisher–Pry (GFP) transformation and noise model:

$$f(Y_{i,t}; L_i, \beta) \equiv \frac{1}{\beta} \log \left(\frac{L_i^\beta}{Y_{i,t}^\beta} - 1 \right) = -k_i(t - t_{0,i}) + \sigma_i(\epsilon_{i,t} + \rho\epsilon_{i,t-1}), \quad (13)$$

where ρ is a technology-invariant MA(1) correlation parameter, σ_i is the technology-specific noise standard deviation (i.e. the standard deviation of the noise in the GFP transformed space), and $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$ are independent noise terms. The function $f(Y_{i,t}; L_i, \beta)$ depends on the observed stock $Y_{i,t}$, the asymptote L_i , and the (a)-symmetry parameter β .

Then, the technology stock $Y_{i,t}$ follows the stochastic relationship

$$Y_{i,t} = \frac{L_i}{(1 + \exp(-\beta k_i(t - t_{0,i}) + \beta \sigma_i(\rho\epsilon_{i,t-1} + \epsilon_{i,t})))^{1/\beta}} \quad (14)$$

$$\epsilon_{i,t} \sim \mathcal{N}(0, 1) \quad \forall t,$$

where $\{L_i, k_i, t_{0,i}, \sigma_i\}$ are technology-specific parameters and $\{\beta, \rho\}$ are technology-invariant for parsimony (justified in Section 3.1).

Noise structure in logit space. The Fisher–Pry noise model is additive in transformed (logit) space. Since we observe stocks in levels, the variance of $Y_{i,t}$ depends on how GFP-space perturbations propagate to the level scale. Consider the logistic case ($\beta = 1$) with i.i.d. additive noise in logit space

$$Y(t) = \frac{L}{1 + \exp(-k(t - t_0) - \sigma\epsilon_t)},$$

where the GFP value is $f \equiv -\log(L/Y - 1) = k(t - t_0) + \sigma\epsilon_t$. The derivative is

$$\frac{\partial Y}{\partial f} = \frac{\partial}{\partial f} \left[\frac{L}{1 + e^{-f}} \right] = \frac{Le^{-f}}{(1 + e^{-f})^2} = \frac{Y(L - Y)}{L}. \quad (15)$$

Thus, the noise variance in levels is proportional to $Y^2(L - Y)^2$, which equals zero at $Y = 0$, reaches maximum at the inflection point $Y = L/2$, and decreases to zero as $Y \rightarrow L$. For the B-R curve

with $\beta = 2/3$, the derivative $\frac{\partial Y}{\partial f}$ has the same qualitative structure: zero at boundaries, maximum at the inflection point.

We assume that the noise in technology flows follows a moving-average MA(1) process. Under MA(1), if the stock at time t exceeds its deterministic value, the stock at $t + 1$ is also more likely to exceed the deterministic prediction.

1.1.4 Normalizing S-curves

To create robust models, we pool data from various technologies. Technologies have unique adoption trajectories with different growth rates and saturation levels. Despite this, we can normalize technology-specific S-curves to analyze cross-technological regularities.

Collapsing S-curves. Let $Y_i(t)$ be the time-series stock data of technology i , and assume different technologies follow the same law with different parameters. We define the normalized technology stock as

$$Y_i^{(n)}(t) \equiv \frac{Y_{i,t}}{L_i}, \quad (16)$$

and normalized time as

$$\tau_i \equiv k_i(t - t_{0,i}). \quad (17)$$

Next, we normalize the noise by dividing it by its standard deviation. That is, the normalized residuals are defined as

$$\frac{\log \left[Y_i^{(n)}(t)^{-\beta} - 1 \right] + \beta \tau_i}{\beta \sigma_i \sqrt{1 + \rho^2}} = u_{i,t} \quad (18)$$

Then the normalized Bertalanffy-Richards curve becomes

$$Y_i^{(n)}(t) = \frac{1}{\left(1 + \exp \left(-\beta \tau_i + \beta \sigma_i \sqrt{1 + \rho^2} u_{i,t} \right) \right)^{1/\beta}}, \quad (19)$$

where the curve still contains technology-specific noise terms. Crucially, $\frac{1}{\beta}$ remains in Eq. (19), allowing us to assess whether technologies collapse onto a similar asymmetric S-curve or exhibit unique asymmetries.¹

1.1.5 Model limitations

Our model assumes independent S-curve diffusion for each technology and excludes declining phases. In reality, technologies emerge, compete, and phase out dynamically [6, 57, 58], and markets expand. The framework here is therefore a univariate benchmark; multivariate distributional forecasts for competing technologies in evolving markets remain for future work. S-curves also describe *how* technologies diffuse, not *why*; mechanistic explanations are likewise outside our scope.

¹For model parsimony and to reduce overfitting risk, we fix $\beta = 2/3$ as a universal constant rather than estimating technology-specific β_i values. Section 3.1 discusses this. We show the universal parameter $\beta = \frac{2}{3}$ replicates the data well in Sections 3.2 and 4.1.

1.2 Our Bayesian forecasting procedure

Here we describe parameter estimation and forecasting for the stochastic model of Eq. 14, using Bayesian inference to generate distributional forecasts.

Notation conventions. We adopt standard Bayesian notation throughout. $\pi(\cdot)$ denotes probability distributions over parameters: $\pi(\theta)$ is the prior and $\pi(\theta | Y)$ is the posterior. $p(\cdot)$ denotes probability densities for observed data and noise: $p(Y | \theta)$ is the likelihood and $p(Y)$ is the marginal likelihood. Throughout this section, $Y_{i,t}$ denotes the observed adoption level for technology i at time t , with vector notation $Y_{i,1:T} = (Y_{i,1}, \dots, Y_{i,T})$ for the full time series. The transformation f maps raw data to transformed variables. We denote the first observation time by 1 to avoid confusion with the inflection-point parameter t_0 .

1.2.1 Introduction to Bayesian inference

Bayesian inference is based on Bayes's theorem, which states

$$\pi(\theta | Y_{1:T}) = \frac{p(Y_{1:T} | \theta)\pi(\theta)}{p(Y_{1:T})}, \quad (20)$$

where $\pi(\theta | Y_{1:T})$ is the posterior distribution of parameters θ given observed data $Y_{1:T}$, $p(Y_{1:T} | \theta)$ is the likelihood, that is, the density of the observed data $Y_{1:T}$ viewed as a function of θ , $\pi(\theta)$ is the prior distribution of θ , and $p(Y_{1:T})$ is the marginal likelihood (evidence). Since $p(Y_{1:T})$ serves as a normalizing constant, Eq. (20) can be rewritten as

$$\pi(\theta | Y_{1:T}) \propto p(Y_{1:T} | \theta)\pi(\theta). \quad (21)$$

Given the likelihood $p(Y_{1:T} | \theta)$ and prior $\pi(\theta)$, the posterior $\pi(\theta | Y_{1:T})$ is determined up to a normalising constant.

1.2.2 Estimating parameters of our stochastic model

We apply Bayesian inference to obtain posterior distributions for the parameter vector $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$ for each technology i .

Log-likelihood of the observed data. The B-R DGP of Eq. (13) specifies additive Gaussian noise in the GFP-transformed space. Define the centred latent variable

$$x_{i,t} \equiv f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i}),$$

which is Gaussian with MA(1) covariance under the DGP (Section 1.1); its closed form in terms of $Y_{i,t}$ appears in Eq. (23) below. By the change-of-variables formula [60], the log-likelihood of the observed technology stocks $Y_{i,1:T}$ is

$$\log \mathcal{L}(Y_{i,1:T} | \theta_i) = \log p_x(x_{i,1:T} | \theta_i) + \log \det \left(\frac{\partial f}{\partial Y_{i,1:T}} \right), \quad (22)$$

where p_x is the joint Gaussian density of $x_{i,1:T}$ under the MA(1) noise structure.² The Jacobian is taken with respect to f alone, since the deterministic trend $k_i(t - t_{0,i})$ has zero derivative in $Y_{i,t}$. Computing the log-likelihood therefore reduces to applying the GFP transform f to the observed data, evaluating the latent Gaussian density at $x_{i,1:T}$, and adding the closed-form Jacobian.

²The textbook form of the change-of-variables theorem reads $p_Y(y) = p_X(g^{-1}(y)) |dg^{-1}/dy|$, with $g : X \mapsto Y$ the level function (here the B-R formula). The formula above is identical: $g^{-1}(Y_{i,t}) = x_{i,t}$, and $|dg^{-1}/dY_{i,t}| = |df/dY_{i,t}|$ since the trend $k_i(t - t_{0,i})$ that separates g^{-1} from f has zero $\partial/\partial Y_{i,t}$.

1.2.2.1 The Bertalanffy-Richards curve log-likelihood

By adding $k_i(t - t_{0,i})$ to both sides of the GFP transformation $f(Y_{i,t}; L_i, \beta)$ from Eq. (13), we obtain a normal distribution

$$x_{i,t} = \frac{1}{\beta} \log \left(\frac{L_i^\beta}{Y_{i,t}^\beta} - 1 \right) + k_i(t - t_{0,i}) = \sigma_i (\epsilon_{i,t} + \rho \epsilon_{i,t-1}) \quad (23)$$

where $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$ for all technologies i . Eq. (23) gives the closed form of the centred latent defined above. The (t, t') element of the Jacobian of f with respect to $Y_{i,1:T}$ is

$$\left(\frac{\partial f(Y_{i,1:T}; L_i, \beta)}{\partial Y_{i,1:T}} \right)_{t,t'} = \frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t'}} = \delta_{t,t'} \frac{-L_i^\beta}{Y_{i,t}(L_i^\beta - Y_{i,t}^\beta)} \quad (24)$$

where $\delta_{t,t'}$ is the Kronecker delta ($\delta_{t,t'} = 1$ if $t = t'$, 0 otherwise). Since the Jacobian is diagonal, its determinant equals the product of diagonal elements. The log-determinant is

$$\log \det \left(\frac{\partial f(Y_{i,1:T}; L_i, \beta)}{\partial Y_{i,1:T}} \right) = \sum_{t=1}^T \left[\beta \log L_i - \log \left(Y_{i,t}(L_i^\beta - Y_{i,t}^\beta) \right) \right]$$

which is independent of the noise model.

i.i.d. noise model First, consider the case where $\rho = 0$, so all noise is i.i.d. Eq. (13) reduces to

$$x_{i,t} = \sigma_i \epsilon_{i,t},$$

which implies

$$x_{i,t} \sim \mathcal{N}(0, \sigma_i^2).$$

The conditional probability of $Y_{i,t}$ given parameters θ_i is normal, yielding the log-likelihood function

$$\log p(Y_{i,t} | \theta_i) = \log \left(\mathcal{N}(x_{i,t}; 0, \sigma_i^2) \mathbb{I}[Y_{i,t} < L_i] \right) + \log \det \left(\frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t}} \right), \quad (25)$$

where $\mathbb{I}[Y_{i,t} < L_i]$ is the indicator function ensuring $Y_{i,t} < L_i$, which is required to ensure Eq. 23 remains well-defined. We next derive the log-likelihood for $\rho \neq 0$.

MA(1) noise model When $\rho \neq 0$, we calculate the joint distribution for the full sequence $x_{i,1}, \dots, x_{i,T}$. The covariance matrix for an MA(1) process is

$$\text{Cov}(x_{i,t}, x_{i,t-j}) = \begin{cases} (1 + \rho^2)\sigma_i^2 & |j| = 0 \\ \rho\sigma_i^2 & |j| = 1 \\ 0 & |j| > 1 \end{cases} \quad (26)$$

The joint probability density function (PDF) is

$$p(x_{i,1}, \dots, x_{i,T}) = \frac{1}{\sqrt{|V_i|}(2\pi)^{T/2}} \exp \left(-\frac{1}{2} X_i^\top V_i^{-1} X_i \right) \quad (27)$$

where

$$X_i = \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,T} \end{bmatrix}$$

is a vector of length T , and

$$V_i = \sigma_i^2 \begin{bmatrix} 1 + \rho^2 & \rho & 0 & 0 & \dots & 0 \\ \rho & 1 + \rho^2 & \rho & 0 & \dots & 0 \\ 0 & \rho & 1 + \rho^2 & \rho & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \rho & 1 + \rho^2 & \rho & 0 \\ 0 & \dots & 0 & \rho & 1 + \rho^2 & \rho \\ 0 & \dots & 0 & 0 & \rho & 1 + \rho^2 \end{bmatrix} \quad (28)$$

is a symmetric $T \times T$ covariance matrix. The first diagonal entry $(1 + \rho^2)\sigma_i^2$ reflects the stationary MA(1) assumption, where $\epsilon_{i,0} \sim \mathcal{N}(0, 1)$ contributes to $x_{i,1}$ even though it occurs before the observed data. For technology diffusion, we assume that $0 < \rho < 1$, indicating positive correlation between noise in consecutive periods. Since $\rho > 0$, V_i is positive definite and therefore invertible. The log-likelihood for the MA(1) model is

$$\log p(Y_{i,1:T} | \theta_i) = \log p(x_{i,1}, \dots, x_{i,T}) + \sum_{t=1}^T \log \det \left(\frac{\partial f(Y_{i,t}; L_i, \beta)}{\partial Y_{i,t}} \right), \quad (29)$$

where the joint density $p(x_{i,1}, \dots, x_{i,T})$ is given by Eq. (27) with covariance matrix V_i , and the Jacobian term is identical to the i.i.d. case.

Posterior distribution under an MA(1) process In an MA(1) model, current noise depends on the previous noise term. Specifically, the noise at time t is $\sigma_i \epsilon_{i,t} + \rho \sigma_i \epsilon_{i,t-1}$. Thus, the first forecast value $Y_{i,T+1}$ depends on both the future unobserved noise draw $\epsilon_{i,T+1}$ and the final in-sample noise term $\epsilon_{i,T}$. $\epsilon_{i,T}$ can be inferred from the data. To best initialize the forecast, we can estimate $\epsilon_{i,T}$ along with the parameters θ_i .

To do this, we define a posterior distribution that explicitly includes $\epsilon_{i,T}$:

$$\pi(\theta_i, \epsilon_{i,T} | Y_{i,1:T}) = \frac{p(Y_{i,1:T} | \theta_i, \epsilon_{i,T}) p(\epsilon_{i,T}) \pi(\theta_i)}{p(Y_{i,1:T})} \propto p(Y_{i,1:T} | \theta_i, \epsilon_{i,T}) p(\epsilon_{i,T}) \pi(\theta_i)$$

where $p(Y_{i,1:T} | \theta_i, \epsilon_{i,T})$ is the likelihood under the DGP, $\pi(\theta_i)$ is the parameter prior, and $p(\epsilon_{i,T}) = \mathcal{N}(0, 1)$ matches the standard-normal distribution of $\epsilon_{i,T}$ under Eq. 14.

For all $s \in \{1, \dots, T-1\}$,

$$x_{i,s} = \sigma_i \epsilon_{i,s} + \rho \sigma_i \epsilon_{i,s-1}$$

as before. For $s = T$,

$$x_{i,T} - \sigma_i \epsilon_{i,T} = \rho \sigma_i \epsilon_{i,T-1}.$$

For the time series $Y_{i,1:T}$, we can perform Bayesian inference on the transformed data $x_{i,1:T}$ to estimate parameters θ_i and $\epsilon_{i,T}$, where

$$\tilde{X}_i := \begin{bmatrix} x_{i,1} \\ \vdots \\ x_{i,T-1} \\ x_{i,T} - \sigma_i \epsilon_{i,T} \end{bmatrix} = \sigma_i \begin{bmatrix} \rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & \rho & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \cdots & 0 & \rho & 1 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \rho \end{bmatrix} \begin{bmatrix} \epsilon_{i,0} \\ \epsilon_{i,1} \\ \vdots \\ \epsilon_{i,T-2} \\ \epsilon_{i,T-1} \end{bmatrix}.$$

The innovation vector includes $\epsilon_{i,0}$ (the implicit pre-sample innovation from the stationary MA(1) assumption), making the matrix $T \times T$ and the innovation vector of length T . We adjust the covariance matrix from Eq. 26 to account for estimating $\epsilon_{i,T}$. The modified covariance matrix is

$$V_i^* = \sigma_i^2 \begin{bmatrix} 1 + \rho^2 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 + \rho^2 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 + \rho^2 & \rho & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho & 1 + \rho^2 & \rho & 0 \\ 0 & \cdots & 0 & \rho & 1 + \rho^2 & \rho \\ 0 & \cdots & 0 & 0 & \rho & \rho^2 \end{bmatrix} \quad (30)$$

where the last diagonal element is now $\rho^2 \sigma_i^2$ instead of $(1 + \rho^2) \sigma_i^2$.

Using the modified covariance matrix V_i^* and vector $\tilde{X}_i$, we could update the joint PDF $p(x_{i,1}, \dots, x_{i,T})$ from Eq. (27) and generate the posterior distribution through Metropolis-Hastings sampling.

Setting $\epsilon_{i,T} = 0$ in practice. We set $\epsilon_{i,T} = 0$ universally. Since $\mathbb{E}[\epsilon_{i,T}] = 0$ by construction, this simplification reduces model complexity without substantially affecting forecast quality (full justification in Section 3.1). The standard covariance matrix V_i from Eq. 26 then applies (rather than V_i^*), and the posterior reduces to $\pi(\theta_i | Y_{i,1:T})$.

1.2.3 Distributional forecasts

We generate distributional forecasts by sampling from the posterior $\pi(\theta_i | Y_{i,1:T})$. We draw N_{samples} samples (after thinning), and for each sample $m \in \{1, \dots, N_{\text{samples}}\}$ use $\theta_{m,i}$ to forecast an adoption path $Y_{m,i,t}$ that the technology could follow.

We forecast for $t \in \{T+1, \dots, T+\tau\}$, where τ is the forecast horizon. Since $\epsilon_{i,T} = 0$, the first forecast step at $T+1$ has noise term $\sigma_{m,i}\epsilon_{i,T+1}$ (as $\rho\epsilon_{i,T} = 0$). For a τ -step forecast over $[T+1, T+\tau]$, the MA(1) noise vector is

$$\epsilon_{T+1:T+\tau,m,i} = \sigma_{m,i} \begin{bmatrix} \epsilon_{i,T+1} + \rho\epsilon_{i,T} \\ \epsilon_{i,T+2} + \rho\epsilon_{i,T+1} \\ \vdots \\ \epsilon_{i,T+\tau} + \rho\epsilon_{i,T+\tau-1} \end{bmatrix} \quad (31)$$

with $\epsilon_{i,t} \sim \mathcal{N}(0,1)$ i.i.d. for $t \in \{T+1, \dots, T+\tau\}$. The first component $\epsilon_{i,T+1} + \rho\epsilon_{i,T}$ shows why $\epsilon_{i,T}$ matters: it propagates into the one-step-ahead forecast through the MA(1) structure.

1.2.4 Implementation: forecasting procedure for a time-series

Algorithm 1 outlines the full procedure for a single technology time-series: data normalisation, Bayesian inference, and forecast generation. Key steps are detailed below.

Data normalization. To assess many technologies simultaneously and fairly, we normalize data to be able to set priors across technologies at each diffusion level. k_i and σ_i are approximately scale invariant across diffusion levels and technologies, so we do not require normalization. However, L_i and $t_{0,i}$ can vary significantly in scale across technologies, making it challenging to define a single prior for all technologies. For example, L_i can range from 10^2 (e.g., fraction of population adopting a behavior) to 10^8 (e.g., number of mobile subscriptions). Similarly, $t_{0,i}$ can vary from being in the 1100s (e.g., adoption of *monasteries*) to the 2000s (e.g., mobile phones). We perform two normalization procedures that help inform prior distributions. First, we normalize adoption data by the maximum observed value,

$$\tilde{Y}_i(t) = \frac{Y_{i,t}}{Y_{i,\max}},$$

where $Y_{i,\max} = \max_t(Y_{i,t})$. Hence, at any diffusion level, $\max_t \tilde{Y}_i(t) = 1$. The value of $\tilde{L}_i \equiv L_i/Y_{i,\max}$ is constant across technologies for a given diffusion level. For example, at 10% diffusion, $\tilde{L}_i \approx 10$ and at 50% diffusion, $\tilde{L}_i \approx 2$.

Second, we shift the time axis so that the first observation occurs at $\tilde{t} = 0$,

$$\tilde{t} = t - t_1.$$

This makes the time of the first observation consistent across technologies, reducing cross-technological variation in the $t_{0,i}$ prior. Some variation in $t_{0,i}$ remains because the inflection point can occur at different times relative to the first observation.

Defining prior distributions. We set weakly informative priors for the parameters $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$ of each technology. Each parameter has a three-parameter shifted Gamma prior. We write $X \sim \Gamma(\alpha, \ell, s)$ to mean $X = \ell + Z$ with $Z \sim \Gamma(\alpha, s)$ [61]. Here ℓ is the lower bound of the support, s is the scale, and α is the shape. The mode is $\ell + (\alpha - 1)s$ for $\alpha \geq 1$. When $\alpha = 1$ the prior reduces to a shifted exponential with origin at ℓ . For numerical stability we sample the logarithm of every parameter except L_i .

We apply the same prior to every technology and at every forecast origin. We first normalise each series so the last observed value $y_{T_{\max}} = 1$, where $T_{\max}$ is the time index of the last observation. The asymptote $\tilde{L}_i$ is then a multiple of current adoption. The prior is

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2, 1, 35), \\ t_{0,i} \sim \Gamma(3, \max(5, T_{\max} - 3/\hat{k}_{\text{ssr},i}), 30), \\ k_i \sim \Gamma(1, 0, 1), \\ \sigma_i \sim \Gamma(1, 0, 1), \end{cases} \quad (32)$$

where $\hat{k}_{\text{ssr},i}$ is a preliminary growth-rate estimate defined below. We call this the *universal* prior, a one-size-fits-all rule, not a prior hand-tuned per technology or diffusion level.

The lower bound $\ell = 1$ on $\tilde{L}_i$ enforces $\tilde{L}_i > 1$. This encodes that the technology has not yet saturated, so its asymptote lies above current adoption.

We anchor the $t_{0,i}$ location to each series through the growth-rate estimate $\hat{k}_{\text{ssr},i}$. We obtain it from a least-squares fit of the GFP transform (Eq. 13) profiled over L_i . If this fit fails, we fall back to $\hat{k}_{\text{ssr}} = 0.15$. Anchoring the location at $T_{\max}$ alone would assume the inflection has not yet occurred. The offset $-3/\hat{k}_{\text{ssr},i}$ relaxes this. A slower growth estimate gives a larger offset, so the prior then allows an earlier inflection. The fallback value of 0.15 is slow, giving an offset of about 20 time steps. The lower bound then does not rule out the true inflection while the technology is below about 83% diffusion (for $\beta = 2/3$). Beyond that level the post-inflection stage of the S-curve is clearly visible, so the prior can be adjusted by hand. Section 4.1 gives the derivation.

The k_i and σ_i priors are unit exponentials. We use this same prior for the backtests of main paper Figure 3 and the solar and wind forecasts of main paper Figure 4.

We keep the prior simple and cross-technology, with no hand-tuning. This shows that a single automatic rule already gives well-calibrated forecasts across the 120-technology set. Section 4.1 gives the full design rationale and sensitivity analyses. It also backtests the prior on a larger 154-technology dataset and shows how more informative technology- or diffusion-stage-specific priors improve forecasts. Section 1.5 discusses this further.

Prior configuration files. We implement the prior distributions of Eq. (32) in `mcmc_setup.py` for the standard model. For forecasting market shares (the fraction of a market rather than absolute adoption levels), we suggest adjusting the L_i prior to reflect that shares are bounded in $[0, 1]$, as implemented in `mcmc_setup_shares.py`. For the asymptotic growth model of Section 4.2, which accounts for market growth via an exogenous exponential parameter, `mcmc_setup_ass_1.py` adjusts the L_i and $t_{0,i}$ location parameters.

Metropolis–Hastings sampling. We perform Metropolis–Hastings MCMC sampling to generate samples from the posterior distribution [62, 63]. Algorithm 1 describes the full forecasting procedure. We base the numbers described in the algorithm on our implementation for the results shown in Figure 3 of the main text, but these can be adjusted as needed.

Algorithm 1: Distributional forecasting procedure for technology diffusion

Input: Technology time series $\{Y_{i,t}\}$ for technology i , forecast horizon τ

Output: Distributional forecast for $t \in \{T+1, \dots, T+\tau\}$

Step 1: Load and normalize data:

Load observed time series: $\{t_1, Y_{i,1}\}, \dots, \{t_T, Y_{i,T}\}$

Calculate normalization constant: $Y_{i,\max} = \max(Y_{i,1}, \dots, Y_{i,T})$

Normalize data: $\tilde{Y}_{i,t} = Y_{i,t}/Y_{i,\max}$ for all $t \in \{1, \dots, T\}$

Shift time: $\tilde{t} = t - t_1$ so first observation occurs at $\tilde{t} = 0$

Step 2: Bayesian inference:

Define prior distributions $\pi(\theta_i)$ for $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$

Optional: Set universal parameters: $\beta = 2/3$, $\epsilon_{i,T} = 0$, and $\rho = 0.8$

Calculate log-likelihood $\log p(\tilde{Y}_{i,1:T} | \theta_i)$ using Eq. (29)

Perform Metropolis–Hastings MCMC sampling:

Generate $N_{\text{MCMC}} = 1,000,000$ samples from posterior

Thin by a factor of 100 to adjust for autocorrelation between MCMC samples

Retain $N_{\text{samples}} = 10,000$ samples

Step 3: Generate forecast trajectories:

foreach posterior sample $m = 1$ to $N_{\text{samples}} = 10,000$ **do**

Extract parameters: $\theta_{m,i} = \{L_{m,i}, k_{m,i}, t_{0,m,i}, \sigma_{m,i}\}$

foreach forecast time $t \in \{T+1, \dots, T+\tau\}$ **do**

Draw $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$

B-R forecast:

$\tilde{Y}_{m,i,t} = L_{m,i} [1 + \exp(-k_{m,i}\beta(t - t_{0,m,i}) + \beta\sigma_{m,i}(\epsilon_{i,t} + \rho\epsilon_{i,t-1}))]^{-1/\beta}$

Update lag: $\epsilon_{i,t-1} \leftarrow \epsilon_{i,t}$

end

Rescale: $Y_{m,i,t} = \tilde{Y}_{m,i,t} \cdot Y_{i,\max}$ for $t \in \{T+1, \dots, T+\tau\}$

end

Step 4: Construct distributional forecast:

foreach forecast time t **do**

The distribution of $Y_{m,i,t}$ across $m = 1, \dots, N_{\text{samples}}$ samples

represents the forecast distribution at time t

end

Step 5 (Optional): Apply shifted correction:

foreach posterior sample m **do**

Generate model value at last training time using $\theta_{m,i}$: $\hat{Y}_{m,i,T}$

Compute GFP shift between observed and model values:

$\delta_{m,i} = \frac{1}{\beta} \log \left(\frac{L_{m,i}}{\hat{Y}_{m,i,T}^\beta} - 1 \right) - \frac{1}{\beta} \log \left(\frac{L_{m,i}}{\tilde{Y}_{m,i,T}^\beta} - 1 \right)$

Apply shift to all forecast times in GFP space and invert to get $Y_{m,i}^{\text{shifted}}(t)$

end

Distributional forecasts for the B-R curve. The τ -step forecast under posterior sample m is the vector

$$\mathbf{Y}_{m,i}(\tau) = \frac{L_{m,i}}{(1 + \exp(-k_{m,i}\beta(\boldsymbol{\tau} - t_{0,m,i}) + \beta\boldsymbol{\epsilon}_{T+1:T+\tau,m,i}))^{1/\beta}}, \quad \boldsymbol{\tau} = \begin{bmatrix} T+1 \\ \vdots \\ T+\tau \end{bmatrix},$$

with $\boldsymbol{\epsilon}_{T+1:T+\tau,m,i}$ the forecast-period noise vector from Eq. 31. The component-wise forecasts $\{Y_{m,i,t}\}_{m=1}^{N_{\text{samples}}}$ at each $t \in \{T+1, \dots, T+\tau\}$ form the predictive distribution at time t .

Shifted and nonshifted forecasts. We consider two forecast variants. The *nonshifted* forecast uses the trajectory $\mathbf{Y}_{m,i}(\tau)$ directly, as described above. The *shifted* variant applies a post-hoc correction in GFP space. For each posterior sample m , let $\hat{Y}_{m,i,T}$ denote the model's fitted value at

the last training time T . We compute the GFP transform (Eq. 13) of both the observed data point and the model value,

$$\delta_{m,i} = \underbrace{\frac{1}{\beta} \log \left(\frac{L_{m,i}^\beta}{Y_{i,T}^\beta} - 1 \right)}_{\text{observed GFP at } T} - \underbrace{\frac{1}{\beta} \log \left(\frac{L_{m,i}^\beta}{\hat{Y}_{m,i,T}^\beta} - 1 \right)}_{\text{model GFP at } T},$$

and add $\delta_{m,i}$ to every forecast-period GFP value before inverting back to adoption levels. Then the model fitted value at the last training observation is equal to the true value (i.e., the residual is zero by construction). The shifted variant can reduce point-forecast error and improve calibration, with the consequence of slightly overdispersed forecast distributions at short forecast horizons (See Section 4.1.)

1.2.5 Adaptive widening of the predictive distribution

This subsection develops the adaptive widening shown in the before/after coverage panels of main paper Figure 3(e,f). Bayesian prediction intervals are often miscalibrated when data are limited or the model is misspecified. A nominal $c\%$ interval then need not contain the outcome $c\%$ of the time, and post-hoc recalibration of the predictive distribution is a standard remedy [64, 65]. Our forecasts show this pattern. On the 120-technology backtesting set the shifted forecast distribution is mildly under-dispersed (Section 4.1). The empirical probability integral transform (PIT) CDF sits below the diagonal at high percentiles, so the prediction intervals are too narrow. Part of this overconfidence is attributable to the noise misspecification discussed in Section 1.4.

We correct it post hoc, without re-specifying the noise model. We apply a median-preserving transform to the per-technology posterior predictive samples, so every point forecast is unchanged. We deliberately use a single-parameter form rather than a flexible nonparametric recalibration. This limits the risk of overfitting the backtesting set. The transform is asymmetric and stretches the upper tail more strongly than the lower tail. The exponent is selected per technology from a small lookup table calibrated on the backtesting set. When applied to forecasts, the procedure uses only the training data and the posterior.

Out-of-sample robustness. The recalibration generalises beyond the technologies used to fit it. Leave-one-technology-out cross-validation reproduces the in-sample calibration (Table 28). On correctly-specified surrogate data the fitted exponent is close to one, so a well-specified model receives almost no widening (Table 29). Section 4.1.4 gives the details.

1.2.5.1 The widening transform

Given posterior predictive samples $\{y_i\}$ with median m , we apply a different exponent above and below the median:

$$y_i^{\text{cal}} = \begin{cases} m (y_i/m)^w & y_i > m, \\ m (y_i/m)^{\sqrt{w}} & y_i \leq m, \end{cases} \quad w \geq 1. \quad (33)$$

In log-space, log-deviations from the median are scaled by w in the upper tail and $\sqrt{w}$ in the lower tail. The median is preserved exactly because w does not affect the median forecast. We apply a smaller exponent to the lower tail because forecasts at lower percentiles are better calibrated. The adoption floor is easier to forecast correctly than the maximum saturation level. The empirical justification for the asymmetric form (versus a symmetric variant in which the same exponent is applied above and below the median) is reported in Section 4.1.4.

1.2.5.2 The diffusion-dependent exponent

The required widening depends on the technology's current diffusion stage. We therefore parametrise w as a function of the estimated current diffusion fraction $\hat{d}_i$.

We estimate technology i 's current diffusion stage from the existing MCMC samples as

$$\hat{d}_i = \frac{1}{\bar{L}_i^{\text{med}}}, \quad \tilde{L}_i^{\text{med}} \equiv \mathbb{P}_{50}[\pi(\tilde{L}_i \mid Y_{i,1:T})], \quad (34)$$

the posterior median of the normalised asymptote. Since the data are normalised so $\max_t Y_{i,t} = 1$, the ratio $1/\tilde{L}_i^{\text{med}}$ is technology i 's diffusion fraction at the last training observation. The estimate $\hat{d}_i$ is read off the existing MCMC samples and requires no additional fitting.

We then look up the widening exponent by log-log linear interpolation (linear in $\log \hat{d}$ and $\log w$) between four anchors calibrated by KS minimisation on the 120-technology backtesting set (Section 4.1.4, Table 26):

$$\log w^*(\hat{d}) = \log w_j^* + \frac{\log \hat{d} - \log d_j}{\log d_{j+1} - \log d_j} (\log w_{j+1}^* - \log w_j^*), \quad d_j \leq \hat{d} \leq d_{j+1}, \quad (35)$$

with anchors

$$w^*(0.05) = 1.93, \quad w^*(0.10) = 2.72, \quad w^*(0.25) = 3.58, \quad w^*(0.50) = 1.73.$$

Outside the anchor range we clip: $w^*(\hat{d}) = 1.93$ for $\hat{d} \leq 0.05$ and $w^*(\hat{d}) = 1.73$ for $\hat{d} \geq 0.50$.

1.2.5.3 Per-technology calibration procedure

For technology i with posterior predictive samples $\{Y_{m,i}(t)\}_{m=1}^{N_{\text{samples}}}$ at any forecast time t :

1. **Estimate the current diffusion level.** Compute $\hat{d}_i = 1/\tilde{L}_i^{\text{med}}$ from Eq. (34).
2. **Look up the widening exponent.** Read $w^*(\hat{d}_i)$ from Eq. (35), clipped to the anchor range.
3. **Apply the asymmetric transform.** For each forecast time t let $m_{i,t} = \text{median}_m Y_{m,i}(t)$ be the predictive median. Each sample is transformed via Eq. (33) with $w = w^*(\hat{d}_i)$:

$$Y_{m,i}^{\text{cal}}(t) = \begin{cases} m_{i,t} (Y_{m,i}(t)/m_{i,t})^{w^*(\hat{d}_i)} & Y_{m,i}(t) > m_{i,t}, \\ m_{i,t} (Y_{m,i}(t)/m_{i,t})^{\sqrt{w^*(\hat{d}_i)}} & Y_{m,i}(t) \leq m_{i,t}. \end{cases}$$

The median is preserved exactly. The upper tail is stretched in log-space by w^* ; the lower tail by $\sqrt{w^*}$.

1.2.5.4 Recommended practice

We recommend applying the adaptive widening transform (Eqs. (33)–(35)) to forecasts produced with this method. The widened intervals are better calibrated on the 120-technology backtesting set, with reductions at every forecast origin (main paper Figure 3f, Section 4.1.4).

We apply the adaptive widening to the solar PV and wind forecasts reported in the main text (Figure 4(d-f)) and in Section 5.2. We do not apply it in two SI contexts where the goal is to characterise the raw model. That is, in the empirical energy backtest of Section 4.2, which displays baseline behaviour against ground truth, and the un-widened calibration diagnostics of Section 4.1. Every figure states explicitly whether the widening is applied.

1.2.6 Software implementation: how to run the forecasting procedure

We implement the procedure in Python using NumPy [66], SciPy [67], Pandas [68], and statsmodels [69]. Below we list the MCMC hyperparameters and outline the run procedure.

1.2.6.1 MCMC hyperparameters.

We set the following hyperparameters for MCMC sampling:

1. Number of MCMC samples: $N_{\text{MCMC}} = 1,000,000$
2. Thinning factor: 100 (to reduce autocorrelation)
3. Target acceptance rate: 0.234 (optimal for Metropolis–Hastings sampling)

4. Warm-up \ burn-in period: 100,000 (to allow the scale to converge to the target acceptance rate)

The values above are the defaults we use, but can be adjusted based on computational resources and convergence diagnostics.

Number of MCMC samples. The number of samples can be increased for more accurate posterior estimates, at the cost of longer computation time.

Number of MCMC chains. We run a single MCMC chain for each technology. Running multiple chains can help assess convergence, but increases computation time.

Target acceptance rate. The optimal acceptance rate of Metropolis–Hastings sampling is 23.4% [70, 71], although this is generally seen as an upper bound, and it may sometimes be lower [72, 73]. In general, an acceptance rate between 10 – 25% is considered acceptable. If the acceptance rate is lower, it means that the proposed distribution is too wide (i.e. there are many samples in low probability regions), and if it is higher, the proposed distribution is too narrow (i.e. the chain is not exploring the parameter space sufficiently).

The thinning factor. The thinning factor can be modified based on the observed autocorrelation in the chains (See Ref. [74] for a discussion on optimal thinning parameter). Thinning is not necessary when conditions are optimal (e.g. sufficiently long sampling run and optimal acceptance rate [75]), but Talts et al. [74] show that generally thinning can significantly improve the posterior predictive performance.

The warm-up \ burn-in period. This period allows adjustment of the scale parameter of the proposal distribution (smaller scale for lower acceptance rates, larger scale for higher acceptance rates) to reach the target acceptance rate.

Sensitivity of hyperparameters. We recommend $\geq 1,000,000$ samples, a thinning factor of ≥ 100 , an acceptance rate of 10–25%, and a burn-in of $\geq 100,000$. Sensitivity tests confirm robustness across reasonable ranges: increasing samples to 5,000,000, raising the thinning factor to 1,000, varying the acceptance rate within 10–25%, and extending burn-in to 500,000 all leave forecast quality essentially unchanged. Burn-in below 100,000 produces poorer mixing and less accurate posteriors; with burn-in above this floor, the thinning factor has only a marginal effect (mainly on the upper tail). Parallel chains (1 vs 4) likewise do not change forecast quality, so we run a single chain per technology.

1.2.6.2 How to run the forecasting procedure.

To run the forecasting procedure, one needs to run two scripts: (1) Either `backtest.py`³ or `forecast.py`⁴ and (2) `plot_forecast.py`. The `backtest.py` or `forecast.py` script performs Bayesian inference to obtain posterior samples, while `plot_forecast.py` generates and plots distributional forecasts using the posterior samples. The steps are as follows.

1. Install the required libraries:

```
pip install numpy scipy pandas statsmodels matplotlib
```

2. Set up the data location pathway for saving and loading files.
3. Load the observed data for the technology of interest and save it in two arrays: one for time points and another for adoption values.

³If backtesting for a specified observed diffusion level

⁴If forecasting only, or backtesting for a specific observed number of data points

4. Define training/testing split based on the desired diffusion level (e.g., 10%, 25%, etc.) or number of observations.⁵
5. Set up the prior distributions for the model parameters using the appropriate configuration file if forecasting (`mcmc_setup.py` for standard models, `mcmc_setup_shares.py` for share models, or `mcmc_setup_ass.py` for asymptotic growth models), following the structure outlined in Eq. (32). For backtesting results presented in the main paper and SI, we always used `mcmc_setup.py` and the universal prior unless otherwise stated.
6. Set specific hyperparameters for MCMC sampling, such as the number of samples, thinning factor and acceptance rate.
7. Run the MCMC sampling to obtain posterior samples (saved in a CSV file).
8. The separate script (`plot_forecast.py`) implements the forecasting procedure:
 - (a) Load the posterior samples from the CSV file.
 - (b) Set the forecast horizon, τ (number of time steps to forecast).
 - (c) Generate forecast trajectories using the posterior samples, as described in Step 3 of Algorithm 1.
 - (d) Construct the distributional forecast from the generated trajectories.
 - (e) Generate a figure to visualize the forecast trajectories (saved as `forecast_ $\tau$ .png`).
 - (f) If backtesting, compare the forecast trajectories to the actual observed data beyond the training period (saved as `backtest_ $\tau$ .png`), where τ defines the forecast horizon.
9. Save the forecast trajectories to a CSV file for further analysis.

⁵To forecast only, set training data to all available data points. That is, define sufficiently large diffusion level or number of observations to include all data points in training.

1.3 Evaluating distributional forecast performance

We assess whether our forecast distributions are *probabilistically calibrated* [65]. A calibrated forecast neither over- nor under-states uncertainty. We measure calibration with the probability integral transform (PIT) [76, 77], the methodology underlying the coverage panels of main paper Figure 3(d-f).

The PIT. Let $F_{Y_{T+\tau}|Y_{1:T}}$ be the CDF of the predictive posterior $p(Y_{T+\tau} | Y_{1:T})$ for a future value $Y_{T+\tau}$ given past data $Y_{1:T}$. The PIT of an observed value $y_{T+\tau}$ is its quantile under this CDF, $F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau})$. If the forecast is calibrated, the PIT is uniform, $F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) \sim \mathcal{U}(0, 1)$.

Estimating the PIT. The predictive posterior is rarely available in closed form. Marginalising over the posterior and exchanging the order of integration writes the predictive CDF as the posterior expectation of the per-parameter CDF. We approximate this expectation by Monte Carlo over posterior draws $\theta^{(n)} \sim \pi(\theta | Y_{1:T})$:

$$\begin{aligned} F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) &= \int_{-\infty}^{y_{T+\tau}} p(Y_{T+\tau} | Y_{1:T}) dY_{T+\tau} = \int F_{Y_{T+\tau}|\theta}(y_{T+\tau}) \pi(\theta | Y_{1:T}) d\theta \\ &\approx \frac{1}{N} \sum_{n=1}^N F_{Y_{T+\tau}|\theta^{(n)}}(y_{T+\tau}), \end{aligned} \quad (36)$$

where $F_{Y_{T+\tau}|\theta^{(n)}}$ is the predictive CDF of $Y_{T+\tau}$ under parameters $\theta^{(n)}$. Equivalently, we take the empirical CDF of the corresponding predictive draws $Y_{T+\tau}^{(n)}$,

$$F_{Y_{T+\tau}|Y_{1:T}}(y_{T+\tau}) \approx \frac{1}{N} \sum_{n=1}^N \mathbb{I}(Y_{T+\tau}^{(n)} \leq y_{T+\tau}), \quad (37)$$

where $\mathbb{I}(\cdot)$ is the indicator function. Both are consistent estimators of the same predictive CDF. We use the empirical-CDF form in the calibration analysis of Section 4.1.

Confidence band. Under perfect calibration the PIT values are $\mathcal{U}(0, 1)$ [78], so a Q-Q plot of the empirical PIT distribution against $\mathcal{U}(0, 1)$ lies on the 45° line. A convex deviation below the diagonal indicates under-dispersion (overconfidence); a concave deviation indicates over-dispersion (underconfidence). We place a 95% confidence band on the plot under this uniform null. For a sample of n independent PIT values, the k -th order statistic of n i.i.d. $\mathcal{U}(0, 1)$ draws follows $U_{(k)} \sim \text{Beta}(k, n + 1 - k)$ [79], and the pointwise 95% band of these Beta quantiles gives the grey envelope in the calibration figures. The binomial counterpart for a PIT histogram is the Clopper–Pearson interval [80], under the same null. When a panel pools several diffusion targets or forecast origins, each technology contributes one correlated PIT value to each, so we size the band with the smallest n entering the panel, the most conservative choice. The empirical Q-Q analysis is in Section 4.1.

Effect of the prior. Calibration depends on how well the prior matches the true parameter distribution. Under a correctly specified model, a prior that is too narrow tends to give overconfident (underdispersed) forecasts, and a prior that is too wide tends to give overdispersed forecasts. The effect grows over the forecast horizon. This is why we keep the priors weakly informative (Section 1.2). On correctly-specified surrogate data forecast with our universal prior, the fitted widening exponent is close to one. This confirms that little recalibration is needed when the model is correct (Section 4.1).

1.4 Empirical surrogate datasets

Purpose. Several analyses compare empirical features to a benchmark generated by the Bertalanffy-Richards model itself. These features include parameter distributions, ensemble growth, residual variance, and forecast errors. We call the benchmark dataset an *empirical surrogate*. The empirical surrogate inherits per-technology parameter estimates from the data, and gives us a null distribution under the assumed DGP. Unlike a forecast for a real technology, where the true parameters are unknown, the surrogate is generated under the model itself, so the model is correctly specified by construction. We then compare the empirical features against this null.

Generating procedure. We work with N empirical technologies (typically $N = 120$) and fix the technology-invariant parameters $(\beta, \rho, \epsilon_{i,T}) = (2/3, 0.80, 0)$ from Section 3.1. We estimate $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ in two steps. First, we fit $\hat{L}_i$ by non-linear least squares of Eq. 3 to $\{(t, Y_{i,t})\}$. Second, conditional on $\hat{L}_i$, we run OLS of the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ on time t . The OLS slope and intercept give $\hat{k}_i$ and $\hat{t}_{0,i}$, and the residual standard deviation gives $\hat{\sigma}_i$ after the $\sqrt{1 + \rho^2}$ correction for MA(1) noise (Section 3.1). We then draw M surrogate datasets, each with N time-series,

$$Y_i^{(s)}(t) \sim \mathbf{P}\left(Y_{i,t} \mid \hat{\theta}_i, \beta, \rho, \epsilon_{i,T}\right), \quad s = 1, \dots, M, \quad (38)$$

where $\mathbf{P}$ is the DGP of Eq. 14. Some sections test whether a technology-invariant choice is empirically justified, and there we vary that parameter locally. For example, Section 3.1 sweeps β and ρ to assess invariance. Section 3.3 substitutes the Gompertz and logistic DGPs for direct model comparison.

Estimating $\hat{\sigma}_i$ on the 1–95% diffusion range. The shock scale $\hat{\sigma}_i$ sets the surrogate’s intrinsic noise floor. It also bounds the forecast accuracy any well-specified model can attain. When we use the surrogate to benchmark forecast performance (Section 4.1), we estimate $\hat{\sigma}_i$ from the GFP-transformed residuals over the central portion of the diffusion range,

$$d_i(t) := Y_{i,t}/\hat{L}_i \in [0.01, 0.95], \quad (39)$$

Here $d_i(t)$ is the empirical diffusion share, the estimated counterpart of the normalised form $Y_i^{(n)}(t) = Y_{i,t}/L_i$ defined in Section 1.1. Two model-misspecification sources motivate this truncation. (i) *Early diffusion.* The B-R model assumes approximately exponential growth at low d (Section 1.1). Empirically, a minority of technologies grow faster than exponential at $d \lesssim 0.5\%$ and decay to exponential thereafter [1] (Section 3.2). (ii) *Near saturation.* The additive-Gaussian-in-GFP noise specification of Eq. 14 forces level-space noise to vanish as $d \rightarrow 1$, but empirical series retain noise there. Including either tail forces the least-squares fit to absorb structurally unmodelled variance into a larger $\hat{\sigma}_i$. We set the lower bound just above the Comin threshold so we retain observations at the forecast origins ($d_{\text{train}} \in \{0.05, 0.10, 0.25, 0.50\}$) used in Section 4.1. On the 120-technology dataset this truncation reduces the cross-technology median of $\hat{\sigma}_i$ from ≈ 0.47 on the full series to ≈ 0.28 on the 1–95% range. We also use two alternatives as sensitivity checks in Section 4.1. A tighter window $\hat{\sigma}_i^{5-95}$ ($d \in [0.05, 0.95]$, cross-technology median ≈ 0.24) drops more low-diffusion data. A wider window $\hat{\sigma}_i^{0-95}$ ($d \in [0, 0.95]$, cross-technology median ≈ 0.33) keeps all observations below the saturation tail. The wider window tests whether the choice of the lower cut matters, and Section 4.1 confirms it does not change the conclusions.

Analytical expected error (interpretive approximation). The decomposition below is a first-order approximation, not a rigorous bound. We use it to build intuition for the empirical error patterns of Section 4.1. We validate all numerical claims there against exact surrogate-based ALE. A first-order Taylor expansion of the absolute log-error $\ell_{i,t} \equiv \log(\hat{Y}_{i,t}/Y_{i,t})$ around the true parameters $(L_i, k_i, t_{0,i})$ yields

$$\ell_{i,t} \approx \frac{\delta L_i}{L_i} + q(d_{i,t})(\delta k_i h - k_i \delta t_{0,i} + \epsilon_{i,t}), \quad q(d) \equiv 1 - d^\beta, \quad (40)$$

where $\delta L_i, \delta k_i, \delta t_{0,i}$ are parameter biases and h is forecast horizon. We denote the three first-order contributions as

$$\begin{aligned} L_0 &\equiv |\delta L_i / L_i| && \text{(level),} \\ U(h) &\equiv q(d) |\delta k_i h - k_i \delta t_{0,i}| && \text{(curvature, grows with horizon),} \\ S(\sigma) &\equiv q(d) \sigma_{\epsilon,i} \sqrt{2/\pi} && \text{(shock),} \end{aligned}$$

with $\sigma_{\epsilon,i} = \sigma_i \sqrt{1 + \rho^2}$ the MA(1)-corrected shock scale: the marginal standard deviation of the GFP-space residual at any single time point. Since the noise satisfies $\eta_{i,t} = \sigma_i(\epsilon_t + \rho\epsilon_{t-1})$ with $\epsilon_t \sim \mathcal{N}(0, 1)$ i.i.d. (Section 1.1), $\text{var}(\eta_{i,t}) = \sigma_i^2(1 + \rho^2)$ and thus $\text{sd}(\eta_{i,t}) = \sigma_{\epsilon,i}$. Operationally, $\sigma_{\epsilon,i}$ is the typical magnitude by which the observed GFP value deviates from the deterministic B-R trend; the underlying innovation scale is $\sigma_i = \sigma_{\epsilon,i} / \sqrt{1 + \rho^2}$. We estimate σ_i from the empirical residual standard deviation after the OLS fit to the GFP-linearised model (see Section 3.1 and the procedure described in this section). Treating the three sources as approximately independent, the expected absolute log error scales as

$$\mathbb{E}[\text{ALE} \mid d] \approx \sqrt{L_0^2 + U(h)^2 + S(\sigma)^2}. \quad (41)$$

Under correct parameters ($L_0 = U = 0$) this reduces to the pure-shock floor

$$\mathbb{E}[\text{ALE} \mid d, \text{correct params}] = (1 - d^\beta) \sigma_{\epsilon,i} \sqrt{2/\pi}, \quad (42)$$

which decreases monotonically from $\sigma_{\epsilon,i} \sqrt{2/\pi}$ at $d \rightarrow 0$ to 0 at $d \rightarrow 1$.

Worked surrogate example. Using the cross-technology median noise scale on the 1–95% range, $\hat{\sigma}^{1-95} \approx 0.28$ (median of $\hat{\sigma}_i$ across the 120 technologies), the MA(1) shock standard deviation is $\sigma_\epsilon = \hat{\sigma}^{1-95} \sqrt{1 + \rho^2} \approx 0.28 \sqrt{1 + 0.8^2} \approx 0.36$. Evaluating Eq. (42) at the typical forecast origin $d = 0.05$ with $\beta = 2/3$ gives the irreducible pointwise ALE floor

$$\mathbb{E}[\text{ALE} \mid d=0.05, \text{correct params}] = \underbrace{(1 - 0.05^{2/3})}_{q(d) \approx 0.86} \cdot \underbrace{0.36}_{\sigma_\epsilon} \cdot \underbrace{\sqrt{2/\pi}}_{\approx 0.80} \approx 0.25,$$

where $q(d)$ is the GFP variance-reducing factor of Eq. (40) (close to 1 at small d), σ_ϵ is the MA(1) shock standard deviation, and $\sqrt{2/\pi}$ converts the standard deviation of a centered Gaussian to its mean absolute value. The product corresponds to a $\approx 28\%$ relative error in level space ($e^{0.25} - 1$). This is the expected pointwise ALE under perfect parameters and noise scale $\hat{\sigma}^{1-95}$, and therefore the best accuracy any model can reach at $d = 0.05$ forecast origin. The tighter alternative $\hat{\sigma}^{5-95} \approx 0.24$ lowers the floor at $d = 0.05$ to ≈ 0.21 , and a wider alternative $\hat{\sigma}^{0-95} \approx 0.33$ increases the floor. We use $\hat{\sigma}^{0-95}$ as a sensitivity check in Section 4.1.

Shape of the well-specified surrogate ALE. The forecasts in Section 4.1 use a per-sample FP shift that pins each posterior curve to the last training observation. After the shift, all parameter-uncertainty contributions vanish at the training boundary and grow with forecast horizon. The noise floor $S(\sigma)$ decreases with diffusion. The empirical surrogate datasets in Section 4.1 are approximately monotone-increasing in diffusion within the observable range up to 95%. The pronounced hump observed in the empirical pointwise ALE in Section 4.1 (main paper Figure 3a, b) is therefore not a feature of the B-R null. It reflects departures of the empirical data from the B-R DGP.

Caveats. Eqs. 40–42 are interpretive approximations, not proven bounds. They break down in three ways. (1) The decomposition is first-order in $(\delta L_i, \delta k_i, \delta t_{0,i})$ and treats the biases as independent. In practice the GFP-space regression couples them. Overestimating L_i implies underestimating k_i and overestimating $t_{0,i}$ (opposite signs for the symmetric case). Second-order terms become non-negligible for large δL_i . (2) The decomposition collapses the MA(1) covariance to its variance contribution $\sigma_{\epsilon,i}^2 = \sigma_i^2(1 + \rho^2)$, dropping the lag-1 covariance $\rho\sigma_i^2$. For averaged residuals over the relevant window lengths, this understates the standard deviation by a factor of ≈ 1.4 . (3) Eq. 42 mixes the L_1 shock scale ($\sqrt{2/\pi}$) with quadratic aggregation. This is convenient for the additive decomposition but is not the exact moment of $|\ell_{i,t}|$. Numerical claims derived from these equations are therefore approximate. The surrogate datasets of Eq. 38 retain the full MA(1) structure and parameter coupling, and we report exact ALE on those surrogates in subsequent figures.

Diagnostic-comparison estimate. For diagnostic comparisons that test whether the DGP can replicate an empirical feature of the ensemble itself (Sections 3.1, 3.2, 3.3), we retain the full-series estimate of $\hat{\sigma}_i$. These comparisons are not benchmarks of forecast-error magnitude, so the truncation argument does not apply. Section 1.5 discusses the broader implications of the noise-model misspecification and its link to the post-hoc widening of prediction intervals in Section 1.2.5.

1.5 Limitations

With finite data, prior distributions influence Bayesian posteriors and forecasts [81,82]. We normalise across technologies so that a single transparent prior applies to all, regardless of the units of L .

We use empirical Bayesian inference [83]: prior hyperparameters are estimated from the same data later used to compute the posterior. This double use can bias results [84,85], but it may also yield more accurate forecasts and improved robustness on short series [83,86]. Section 4.1 shows reasonable out-of-sample performance, indicating that the posterior accurately balances prior and data evidence [87].

Our Bayesian inference relies on MCMC methods, which are inherently stochastic [88–90]. We run a single chain per technology for computational efficiency; in real-data backtesting we confirmed that running parallel chains does not materially change forecast quality (results not shown). Finite computational resources mean errors remain in posterior samples [90], and finite sampling from the posterior predictive distribution adds variability to the computed prediction intervals [81]. In particular, observations near the boundary of the 95% prediction interval may switch between inside and outside across runs: small MCMC-driven shifts in the posterior predictive move the interval edge across them.

Throughout the text, we avoid over-interpreting results that may depend on a particular seed or prior assumption. For example, forecasting from data covering only the first 1% of diffusion provides so little information that the prior dominates the posterior.

Noise model. The DGP uses a single additive Gaussian shock in GFP space (Eq. 14), which is structurally misspecified at the diffusion endpoints (Section 1.4). This contributes to the moderate overconfidence of the B-R prediction intervals reported in Section 4.1 (main paper Figure 3(e,f)), which we correct post hoc via the widening transform of Section 1.2.5. A fully model-based remedy would replace the single shock with a heteroscedastic specification, or augment it with a second source of variance acting in level (or log-level) space, to capture residual variability where the GFP shock vanishes near saturation. Boswijk and Franses [50] propose a heteroscedastic error structure for the Bass diffusion model on essentially these grounds; a similar extension could in principle be applied to the B-R model used here. Such specifications, however, introduce identification challenges and add parameters per technology that are difficult to pin down on the short data sets considered here. We leave this extension to future work.

Universal shape parameter. The B-R curve uses a single technology-invariant exponent $\beta = 2/3$ (Section 3.1). The cross-section supports this value on average (main paper Figure 2d), but true technology-specific exponents would shift the inflection diffusion and the asymmetry of each fitted curve. Technology-specific estimation of β_i worsened PIT calibration in our backtests (Section 4.1.6), indicating that β_i is weakly identified per series under our priors. A coarser grouping by sector or by diffusion stage is a plausible refinement that we leave to future work.

Saturation prior at low diffusion. The universal $\tilde{L}$ prior is wide and has its mode well above the observed maximum. At low forecast origins the prior dominates the likelihood and the cross-technology posterior median of $\tilde{L}$ can sit above the true normalised asymptote on average (Section 4.1.2.3). This is by design. The universal prior demonstrates that a single prior, combined with technology-specific heuristics for the t_0 location, produces well-calibrated forecasts across forecast origins and horizons (Section 4.1, main paper Figure 3(e,f)). It is a transparent default, not a ceiling. We encourage practitioners to exploit any prior information available for their application. For example, those with diffusion-stage knowledge can replace the universal prior with

the tighter, stage-conditional alternative of Section 4.1.7, which lowers MedALE on the calibration set. A complementary direction for future work is to adjust the k prior to compensate for the noise-misspecification bias identified in Section 4.1.2.3.

2

Data

We compile a technology diffusion dataset comprising time-series spanning the 12th to 21st centuries at global, regional (e.g. EU), and country levels. To provide ground truth, we keep only technologies with data on both early and late stages of diffusion. We perform a six-step cleaning procedure that reduces the dataset from 1362 to 120 time-series: (1) remove death phases (post S-curve asymptote), (2) ensure sufficient sample size, (3) require complete S-curve representation (i.e., enough coverage of early and late diffusion), (4) remove duplicated technologies and non-technology time-series, (5) remove outliers, and (6) eliminate highly correlated time-series. We discuss limitations at the end of this section.

Defining technology diffusion time-series. We do not impose strict definitions of *technology* or *diffusion*, to keep as many series as possible and avoid bias. Technology diffusion time-series measure the adoption level of a specific technology over time, represented as stocks (cumulative adoption). In some cases, the data is in shares (percentage of all potential adopters); we keep these but note that in- and out-of-sample performance is broadly similar. Some time-series from our initial dataset arguably do not represent technology diffusion (e.g., *GDP*, a measure of *democracy*, *governmental legislation*), and we remove these. Where the case is borderline, however, we tend to keep the series (e.g., *gold reserves*, *monasteries*). Our results are not sensitive to these choices.

2.0.1 Datasets

Our initial dataset combines data from CHAT [91], HATCH [92], World Bank [93], Our World in Data [94], IEA [95], Grubler and Nakićenović [27, 96, 97], IHS [98], RethinkX [99], and Way et al. [49], with monastery records from Janauschek [100]. Grubler and Nakićenović provide European-level, country-level and global-level data; CHAT and HATCH offer country-level and global data; and the remaining datasets provide global-level data.

Aggregating country-level data to global proxy. For technologies in CHAT and HATCH with only country-level data, we construct a global diffusion proxy by aggregation. Let $Y_{i,t}^{(c)}$ denote diffusion of technology i in country c at time t . We compute a global technology stock as

$$Y_{i,t} = \sum_c Y_{i,t}^{(c)},$$

treating *NaN* values as zero. This is a simplifying assumption: a missing country entry may reflect absent reporting rather than true zero adoption, so the aggregated level may understate actual global diffusion, particularly in early years when reporting coverage is sparse. After aggregation, we replace remaining zeros with *NaN* to exclude them from model fitting.

Converting flow data to stock data. When the data is already in stocks, we keep it as is. When the data is in flows, we convert it into cumulative stocks. For flow data $\Delta Y_{i,t}$, we compute

stocks as

$$Y_{i,t} = \sum_{T=1}^t \Delta Y_{i,T},$$

which assumes zero initial stock. If a technology was already present before the start of the sample, this understates the level and can alter the apparent diffusion fraction. The final dataset contains few aggregated flow series, and the cleaning procedure ensures that most series start near a technology’s introduction, so the effect is small but cannot be ruled out for all series.

2.0.2 Shares and non-shares

The data comprises both shares and non-shares. For example, the datasets contain data on the total km of railroad tracks (a non-share) as well as the share of the population vaccinated with certain vaccines (shares). Shares are a normalized metric between 0 and 100 (or 0 and 1) accounting for differences in potential adopter base, while non-shares are absolute measures of adoption level. We use both types of data in our analysis, noting that about a third of the data is measured in shares. We do not observe any qualitative differences between the two.

Calculating useful energy shares. For energy, we have experimented with estimating S-curves for both levels and shares; this affects results for specific technologies but does not affect the qualitative overall conclusions. To compute shares, we rely on the concept of *useful energy* to put fossil and electricity-based energy sources on a common footing. We use useful energy data from Way et al. [49], who calculate useful energy from primary energy production, and extend their series (ending in 2020) through 2021–2024 using energy data from Our World in Data [101, 102] and efficiency factors from Way et al. For primary energy consumption $P_{i,t}$ of energy technology i with useful energy factor $f_{i,t}$, useful energy is $U_{i,t} = f_{i,t}P_{i,t}$. Total useful energy consumption is $U_t = \sum_i U_{i,t}$. The share of useful energy from technology i is

$$Y_{i,t} = \frac{U_{i,t}}{U_t}.$$

We use useful energy shares in our dataset for backtesting. We only use useful energy (non-share) data for forecasting future energy technology diffusion levels and when discussing energy technology diffusion levels (Sections 4.2, 5.1, and 5.2). We compare results using useful energy shares vs. levels in Section 4.2.

When the technology measure is a share and is considered a proxy aggregate (i.e., summing across countries; for example *HIB3 vaccine* and *RCV1 vaccine*), we compute an average. There are two such cases in the final cleaned dataset. In these two cases, the data contains two countries. For share data $Y_{i,t}^{(c)} \in [0, 100]$, we proxy the global share as

$$Y_{i,t} = \frac{1}{C} \sum_{c=1}^C Y_{i,t}^{(c)},$$

where $C = 2$.

Shares vs. non-shares: statistical comparison. Of the 120 technologies in the cleaned dataset, 42 (35%) are measured as shares and 78 (65%) as absolute levels. Table 14 shows the distribution by sector. Using the same statistical methods applied to the full dataset in Section 3.2, we tested whether shares and non-shares exhibit different diffusion dynamics. Table 15 summarizes the results. We found no statistically significant differences for R^2 , normalized residuals, or growth rates (k_i) (all p-values > 0.05). The inflection point $t_{0,i}$ (measured relative to first observation) occurs earlier ($p < 0.01$) for shares. This indicates that technologies measured in shares have less data at very low adoption levels. Given these findings, we treat shares and non-shares together in our analysis.

Sector	Total	Shares	% Shares
Industry	30	8	27%
Transport	19	7	37%
Agriculture	19	4	21%
Health	17	6	35%
Energy	15	7	47%
ICT	10	4	40%
Household technologies	4	4	100%
Financial services	2	0	0%
Education	2	2	100%
Military	1	0	0%
Other services	1	0	0%
Total	120	42	35%

Table 14: Distribution of shares and non-shares technologies by sector in the cleaned dataset.

Metric	Test	Statistic	p-value
R^2	t-test	-0.47	0.64
	KS-test	0.17	0.39
	Mann-Whitney U	1760	0.50
Normalized residuals	KS-test	0.02	0.62
	Mann-Whitney U	3232347	0.58
Growth rates, k_i	t-test	-0.01	0.99
	KS-test	0.19	0.25
	Mann-Whitney U	1655	0.93
Inflection point $t_{0,i}$ (rel. to first obs.)	t-test	-2.70	0.008
	KS-test	0.35	0.002
	Mann-Whitney U	1111	0.004

Table 15: Statistical tests comparing diffusion dynamics between shares and non-shares. The only statistically significant difference is in the inflection point $t_{0,i}$ (measured relative to first observation), where shares have earlier inflection points ($p < 0.01$).

Computing adjusted R^2 for fit quality assessment. We evaluate S-curve fit quality across all 1362 time-series using adjusted R^2 from Bertalanffy-Richards curve fits (Eq. 3).⁶ For S-curve fit $\hat{y}_{i,t}$ to transformed data $y_{i,t}$ (defined via the GFP transformation in Eq. 13), we compute the adjusted R^2 as

$$\bar{R}_i^2 = 1 - \frac{\text{SSE}/(N_i - p)}{\text{TSS}/(N_i - 1)}, \quad \text{SSE} = \sum_t (\hat{y}_{i,t} - y_{i,t})^2, \quad \text{TSS} = \sum_t (y_{i,t} - \bar{y}_i)^2,$$

where $\bar{y}_i = \frac{1}{N_i} \sum_t y_{i,t}$ is the sample mean, N_i is the number of data points, and $p = 2$ is the number of parameters estimated ($\hat{k}_i$ and $\hat{t}_{0,i}$).

Figure 2 shows the distribution of the adjusted R^2 for the raw dataset. While the fit explains $> 90\%$ of the variance for many technologies, a significant number of technologies exhibit adjusted $R^2 < 0.5$, potentially due to incomplete S-curves, outliers, or data quality issues.

⁶Running `save_raw_data.py` generates the dataset with 1362 time-series and corresponding R^2 values.

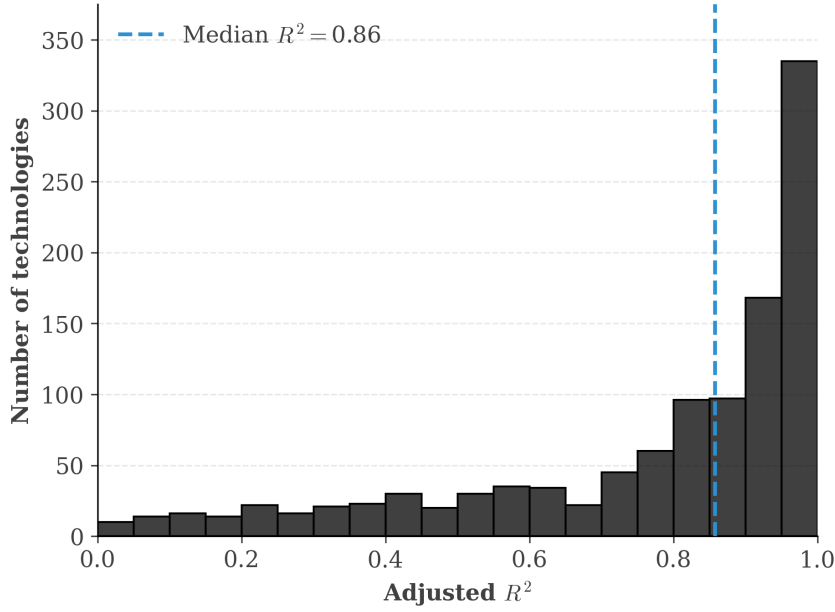


Figure 2: Distribution of adjusted R^2 for the fit of the Bertalanffy-Richards curve to all 1362 raw time-series.

2.0.3 Data cleaning

Our cleaning procedure comprises six systematic steps.⁷ Table 16 summarizes the impact of each cleaning step on the dataset. Earlier steps naturally remove more time-series, as the pool of candidates diminishes progressively. The folder `datasets/` contains the data after each cleaning step.

Step	Description	Removed	Remaining	Data pts.	Unique technologies	% Ret.
–	Raw data	–	1,362	56,014	339	100%
1	Remove death phase	0	1,362	49,605	339	100%
2	Enough data points	404	958	45,665	290	86%
3(a)	Saturation	290	668	31,998	216	64%
3(b)	Early S-curve	211	457	26,125	155	46%
4(a)	Unique technologies	302	155	7,398	155	46%
4(b)	Incorrect saturation	34	121	6,084	121	36%
5	Remove outliers	0	121	6,032	121	36%
6	Correlated time-series	1	120	5,991	120	35%
Total	Final dataset	1,242	120	5,991	120	35%

Table 16: Summary of the data cleaning procedure. We identify *unique technologies* by manually grouping time-series based on title and description. Technologies are classified into 11 sectors (Table 18) using [103–108] as guidance.

2.0.3.1 Computing parameters for cleaning

Estimating asymptote and asymmetry parameters via NLLS. We fit Eq. 3 using NLLS regression⁸, excluding the first and last data points (this makes our estimates more robust, as we

⁷Running `data_cleaning.ipynb` generates the final cleaned dataset with 120 time-series.

⁸We use a `piecewise_regression` package that uses bootstrapping, which may lead to varying results upon replication. Under tests the resulting cleaned dataset remains, but number of data points at each step may vary slightly.

have noticed the estimate of L is very sensitive to these points). This yields $\hat{L}_i$ and $\hat{\beta}_i$. We apply the linearization transformation (Eq. 13) with

$$L = \max[\hat{L}_i, 1.2 \max_t(Y_{i,t})], \quad \beta = \min[\max[\hat{\beta}_i, 0.15], 50].$$

The factor 1.2 provides a 20% margin above maximum observed value (values between 1.1 and 1.3 would yield the same technology set). For plotting the distribution of residuals, we use $L = \hat{L}_i$ where $\hat{L}_i$ is the NLLS estimate.

We set $\beta \geq 0.15$ to avoid extreme parameter estimates. The lower bound $\beta = 0.15$ corresponds to inflection at $\approx 1\%$ of asymptote; the upper bound $\beta = 50$ to $\approx 99\%$ diffusion. As discussed in Section 3.2.1, early technology growth can exhibit initial rapid growth slowing to constant exponential behavior after the first 1% of diffusion. Such behavior may lead to estimating very small $\hat{\beta}_i$, due to model mis-specification rather than actual diffusion dynamics. We have performed a sensitivity analysis, and varying the lower bound between 0.1 and 0.2, or setting $\beta = 2/3$ for all technologies yield similar results.

Computing normalized residuals via OLS. We transform data using the GFP transform (Eq. 13) and regress the transformed data $f(Y_{i,t}; L_i, \beta_i)$ on time t using OLS to estimate $\hat{k}_i$ and $\hat{t}_{0,i}$. We calculate normalized residuals

$$\tilde{R}_{i,t} = \frac{R_{i,t}}{s_i(R_i)}, \quad \text{where} \quad R_{i,t} = y_{i,t} - \hat{k}_i(t - \hat{t}_{0,i}), \quad (43)$$

and $s_i(R_i) = \sqrt{\frac{1}{N_i-2} \sum_t R_{i,t}^2}$ is the sample standard deviation of residuals $R_{i,t}$ (with a $N_i - 2$ degrees-of-freedom correction, for estimating $\hat{k}_i$ and $\hat{t}_{0,i}$).⁹ Algorithm 2 provides the complete procedure.

Algorithm 2: Computing normalized residuals for technology i

Input: Raw time-series data $\{(t, Y_{i,t})\}_{t=1}^{N_i}$ for technology i

Output: Normalized residuals $\tilde{R}_{i,t}$ for $t = 1, \dots, N_i$

Step 1: Initial parameter estimation via NLLS

Fit Bertalanffy-Richards curve (Eq. 3) via NLLS to obtain $\hat{L}_i$ and $\hat{\beta}_i$

Step 2: Adjust parameters for robust transformation

Set $\beta = \min[\max[\hat{\beta}_i, 0.15], 50]$ (avoid extreme asymmetry parameter estimates) Set

$L = \max[\hat{L}_i, 1.2 \max_t(Y_{i,t})]$ (ensure asymptote exceeds max observed value)

Step 3: Apply GFP transformation

Compute $y_{i,t} = -\frac{1}{\beta} \log \left[\left(\frac{L}{Y_{i,t}} \right)^\beta - 1 \right]$ for each $t = 1, \dots, N_i$

Step 4: Estimate linear model parameters via OLS

Fit $y_{i,t} = k_i(t - t_{0,i}) + \epsilon_{i,t}$ by OLS to obtain $\hat{k}_i$ and $\hat{t}_{0,i}$

Step 5: Calculate raw residuals

Compute $R_{i,t} = y_{i,t} - \hat{k}_i(t - \hat{t}_{0,i})$ for each $t = 1, \dots, N_i$

Step 6: Compute residual standard deviation

Calculate $s_i(R_i) = \sqrt{\frac{1}{N_i-2} \sum_{t=1}^{N_i} R_{i,t}^2}$

Step 7: Normalize residuals

Compute $\tilde{R}_{i,t} = \frac{R_{i,t}}{s_i(R_i)}$ for each $t = 1, \dots, N_i$

return $\{\tilde{R}_{i,t}\}_{t=1}^{N_i}$

2.0.3.2 Step 1: Removing death phase data

Some technologies (e.g., *telegrams*, *steamships*) have complete life cycles including phases where the stocks decline. In this paper, we model only the first three phases (up to maturity), omitting

⁹Under the MA(1) noise model (Eq. 13), the OLS residuals $R_{i,t}$ are autocorrelated, so the $N_i - 2$ divisor does not yield an exactly unbiased variance estimator. This does not materially affect the cleaning procedure, as s_i is used only for residual normalization, not for model parameter estimation.

death-phase data. We identify the death phase by analyzing growth trends in transformed data $y_{i,t}$ (under GFP transformation, Eq. 13). To do this, we cannot use $\hat{L}_i$ and $\hat{\beta}_i$ from NLLS fitting directly, as the death phase may distort them. Instead, we use fixed values $L_i = 1.2 \max_t(Y_{i,t})$ and $\beta_i = 2/3$. This L_i includes all data points in the transformation; $\beta_i = 2/3$ ensures plausible asymmetry. Other values ($\beta_i \in \{0.5, 1\}$ or $L_i = 1.1 \max_t(Y_{i,t})$) yield similar results.

For $y_{i,t}$, the start of the death phase represents a shift from positive to negative growth. We detect this via piecewise linear regression with a single breakpoint following Muggeo [109]:

$$y_{i,t} = \begin{cases} k_{i,1}(t - t_{0,i}) + \xi_{i,t}, & t \leq b_i \\ k_{i,1}(b_i - t_{0,i}) + k_{i,2}(t - b_i) + \xi_{i,t}, & t > b_i \end{cases}$$

where $k_{i,1}$ is the slope before breakpoint b_i , $k_{i,2}$ the slope after, and $\xi_{i,t}$ the error term. We estimate b_i and its 95% confidence interval $[b_i^{(0.025)}, b_i^{(0.975)}]$. If $k_{i,1} > 0$ and $k_{i,2} < 0$, we omit data for $t > b_i^{(0.975)}$. If both slopes ≥ 0 , the time-series exhibits no death phase. If both < 0 , the time-series exhibits only the death phase, and we remove the technology.

Rationale for removing death phase before other steps. We remove the death phase first because S-curve models do not address the death phase: including it introduces strong model misspecification. We start with death phase removal so that all further steps can use statistically estimated values of $\hat{L}_i$ and $\hat{\beta}_i$.

Manual correction of misidentified death phases. Piecewise linear regression may misidentify death phases due to noise or few observations in the saturation phase. We use $b_i^{(0.975)}$ as the cutoff and visually inspect all 306 technologies with identified death phases. We override the automated breakpoint for 85 of them. In 80/85 cases we restore data points that the automated procedure had classified as post-saturation, since they represent noise around the asymptote rather than a sustained decline. This occurs because the noise model underlying the regression (GFP transform) assumes homoscedastic errors, which implies noise diminishes to zero as diffusion approaches the asymptote. In reality, some noise may remain in the saturation phase, leading to misidentification of breakpoints. Including additional data points in the saturation phase helps determine the quality of forecasts in later steps. In 5/85 cases we remove additional data points, as the death phase starts earlier than initially identified. An example is shown in Figure 4g.

Figure 6b shows an example of misidentification for *Steamship capacity*, where regression identifies a death phase starting in 1910, but visual inspection indicates 1930 (blue dots are those data points added back into the dataset). Figure 7 shows an example for *cane sugar production (Japan)*, where regression identifies a death phase starting in 2006; visual inspection indicates 2003 (brown dots are additional data points removed in this step). We adjust both breakpoints based on these inspections.

2.0.3.3 Step 2: Enforcing minimum data requirement

We require $N_i \geq 10$ data points. We do this because using fewer points yields more unstable parameter estimates, but the results are not sensitive to the threshold, e.g. analysis with $N_i \geq 8$ or $N_i \geq 12$ yields similar results. Of the 1362 time-series, 958 satisfy the $N_i \geq 10$ requirement.

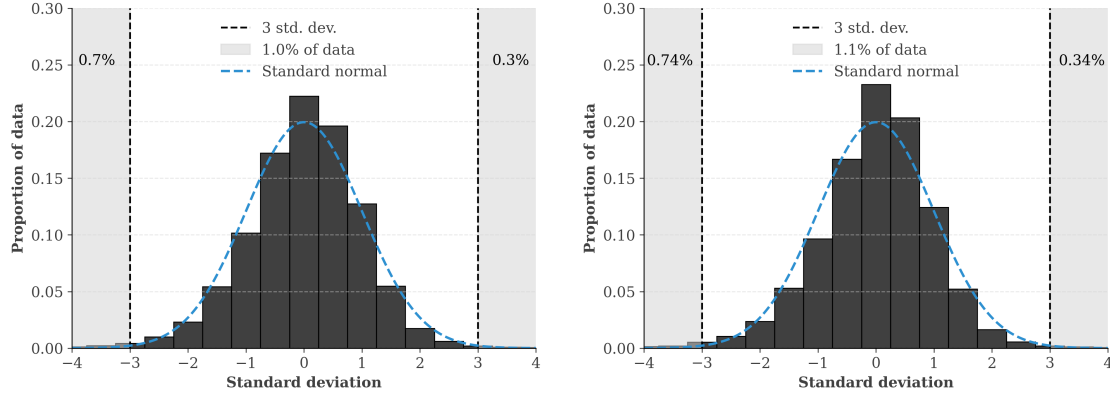
2.0.3.4 Step 3: Requiring complete S-curve coverage

Testing full S-curve dynamics requires both early and late phase data. We require early phases to analyze initial growth patterns (and train the model for backtesting), the full curve to analyze the general shape of adoption, and later phases to analyze saturation dynamics and validate forecasts. Specifically, we require (1) at least 5 data points before 35% diffusion, and (2) diffusion beyond 80% of estimated L .

To find which time series satisfy these requirements, we first estimate the asymptote $\hat{L}_i$ using NLLS. Next, for saturation, we require that $\max_t(Y_{i,t}) > 0.8\hat{L}_i$. This removes 290 time-series. For early diffusion, we keep time series that have at least 5 data points before 35% diffusion,

$|\{t : Y_{i,t} < 0.35\hat{L}_i\}| \geq 5$. This removes 211 time-series. Of 958 time-series, 457 satisfy both requirements.¹⁰

At this point in the data cleaning procedure, all technologies contain both early and late adoption data, no death phase, and at least 10 data points, which should make them suitable for testing hypotheses about S-curves. We observe the distribution of residuals before removing outliers to highlight the effect of outlier removal. Figure 3 shows the normalized residuals after cleaning steps 1-3. Table 19 summarizes statistics of the residuals. As previously noted, these residuals are computed using the GFP transformation with $L = \hat{L}_i$ and $\beta = \hat{\beta}_i$ from NLLS fitting, unlike the residuals used in the cleaning steps which used adjusted L and β for robustness. The normalized



(a) Normalized residuals for 958 time-series with $N_i \geq 10$ (Step 2) (b) Normalized residuals for 457 time-series representing full S-curve (Step 3)

Figure 3: Histograms of normalized residuals under the Bertalanffy-Richards curve fit at different cleaning stages. (a) Residuals for 958 time-series with at least 10 data points (Step 2). (b) Residuals for 457 time-series that have at least 10 data points and cover a majority of the adoption curve (Step 3). In both cases, the residuals are well-behaved with $\approx 1\%$ of the residuals greater than 3 standard deviations from the mean (slightly fatter tails than a standard normal distribution).

residuals (standard deviations on the x-axis) behave similarly between steps 2 and 3, with $\approx 1\%$ beyond 3 standard deviations in both cases (compared to 0.3% in a standard normal distribution). In both cases, there are more extreme negative residuals than positive ones. Table 19 indicates that the median is positive, but extreme left-tail observations give the distribution negative skewness. This is likely due to early-phase diffusion, where some technologies exhibit rapid initial growth before stabilizing to an exponential regime, a phenomenon that is not fully captured by the B-R curve. We discuss this in detail in Section 3.2.

2.0.3.5 Step 4: Selecting unique technologies and removing poor-quality data

Step 4(a): Applying scoring system to select unique technologies. Some technologies appear multiple times under different names or from different sources (e.g., *mobile phones*, *railroads*, *television*). We manually group technologies into unique technologies based on title and description. For the grouping, see the data dictionary in `data/cleaned/data_dictionary_final.csv` for grouping the remaining 457 time-series into 155 unique technologies. See the notebook `data_cleaning.ipynb` for the code used to generate the data dictionary.

After grouping technologies into unique technologies, we retain one representative time-series per unique technology using a scoring system that prioritizes global coverage, number of data points, early-phase representation, and temporal resolution:

- +1 point if the time-series is at the global level (over country-level data)
- +0.5 point if $N_i \geq 20$, +1 point if $N_i \geq 50$

¹⁰Saturation thresholds of $0.7\hat{L}_i$ or $0.9\hat{L}_i$, and early-phase requirements of $0.3\hat{L}_i - 0.5\hat{L}_i$ yield similar results.

- +1 point if $\{Y_{i,t} : Y_{i,t} < 0.05\hat{L}_i\} \geq 5$ (sufficient early data to train on)
- +1 if $\Delta_{\max} = 1$ year; -1 if $\Delta_{\max} \in [5, 10]$ years; -2 if $\Delta_{\max} > 10$ years, where $\Delta_{\max} = \max_{j=1, \dots, N_i-1} (t_{j+1} - t_j)$ for ordered observation times $t_1 < \dots < t_{N_i}$

When multiple time-series for the same technology have the same highest score, we select the time-series with the longest early data period (for training in the validation exercise) or longest overall time-series (for validation forecasts in backtesting). If still equal, we use adjusted $R^2 \geq 0.9$ as a tiebreaker. The selection procedure described here reduces the dataset from 457 to 155 technologies (302 time-series removed).

Exceptions to scoring system selection. In some cases, we manually select lower-scoring time-series. For *Coal*, *Hydropower*, and *nuclear electricity*, we select Way et al. [49] data (useful energy shares) over series with higher-scoring primary energy or absolute-TWh data, because the useful-energy-share representation yields a consistent unit across the energy technologies used in the forecasting exercise (Section 5.2). For *Sulphuric acid production*, we select Canadian data because it has longer time-series and more early-phase data over world data.

Step 4(b): Removal of poor-quality time-series. We manually flag 34 of the 155 unique technologies for potential removal through visual inspection, then validate these using quantitative checks (with some overlap). We consider four rationales for removal:

(1) Non-technology time-series. Remove: *GDP (UK)*. appears in CHAT/HATCH as potential diffusion indicators but do not represent, in our view, technology adoption. Other non-technology time-series exist, such as *Democracy Index*. These were removed in earlier filtering steps.

(2) Incorrectly specified saturation. Remove time-series where the asymptote $\hat{L}_i$ does not reflect true saturation. All checks below use only data statistics, not model outcomes, and are therefore not circular. The 154-technology robustness check (Section 4.1) confirms that including the flagged technologies does not materially affect calibration or accuracy. (a) The asymptote estimate, $\hat{L}_i$, changes by $> 50\%$ when excluding the last 20% of data (flags 7 technologies); this tests for instability in the NLLS fit that signals the saturation level is not well identified from the data. (b) Average year-over-year log growth does not decrease: $\overline{\Delta \log Y} / \Delta t_{\text{final}} \geq 0.8 \cdot \overline{\Delta \log Y} / \Delta t_{\text{mid}}$, where subscripts denote final 5 data points and middle 40–60% of the time-series respectively. This is distinct from Step 3’s requirement that the technology has *reached* 80% diffusion: a technology can satisfy Step 3 yet still show insufficient deceleration in growth rate, which this test detects (flags 7 technologies). (c) Visual inspection: technologies that show growth without any visible curvature toward an asymptote are removed on judgement grounds: *sodium battery storage*, *high-speed rail*, *satellite launches*, *CCS*, *offshore wind (Germany)*, *solar thermal*, *lithium-ion batteries*, *space launches*. This complements criterion (b) — (b) is statistical, (c) is visual — and together they flag cases where neither the data statistics nor the shape of the curve provide evidence of saturation. (d) NLLS cannot identify $\hat{L}_i$ from data confined to the early-exponential regime ($Y_{i,t} \ll L_i$): the B-R model then reduces to $Y_{i,t} \propto \exp(k_i t)$, in which L_i enters only through curvature that has not yet manifested [40]. This is an identifiability constraint, not a model-outcome judgement. We remove technologies whose data show no curvature toward saturation.

(3) Poor data quality. We remove time-series with erratic behavior inconsistent with smooth diffusion, applying two standard outlier-detection thresholds to year-over-year changes that are independent of model outcomes: (a) $|\Delta Y_{i,t}| > 10\sigma_{\Delta Y}$ where $\sigma_{\Delta Y}$ is the standard deviation of year-over-year changes (flags 10 technologies); (b) coefficient of variation $\text{CV}(\Delta Y_{i,t}/Y_{i,t-1}) > 3$ (flags 14, with overlap). The S-curve model assumes smooth, monotone diffusion; time-series with sudden jumps or persistent fluctuations violate this assumption and would produce unreliable fits regardless of model choice. This step removes, for example, *hospital beds* and *cheques* (sudden jumps) and *iron ore*, *cobalt*, and *nickel* (persistent fluctuations). See Figure 4.

Nine technologies pass all automated checks above but are flagged and removed through manual visual inspection: *hernia surgery*, *coronary bypass*, *EFT (UK)*, *sensible heat storage*, plus five others. These are relatively short time series ($N_i < 15$) with insufficient data to distinguish a genuine asymptote from noise, so we opt to remove them. We also remove several battery-, CCS-, and

solar-related technologies. We remove these for two reasons. (1) Our application to solar and wind energy forecasting is a key motivation for this study, so we ensure the dataset does not contain technologies in the same energy storage or solar sector, which we reserve for out-of-sample validation (Section 5.2). (2) These technologies have very few data points and no observable curvature toward an asymptote, making the NLLS asymptote estimate unreliable (the same criterion as (d) above). For example, CCS has fewer than 15 global data points with no sign of saturation [110]. The same applies to sodium battery storage [111] and other early-stage storage technologies such as *lithium-ion battery storage*, *latent heat storage*, and *solar thermal energy* [112].

(4) Removal of individual outlier observations within retained technologies. We remove individual data points that reflect events unrelated to diffusion—wartime disruptions, financial crises, or data-entry errors—while retaining the technology itself. Specific cases are documented in Section 2.0.3.6.

For three technologies (*ships*, *main engine RPM*, *sand/gravel construction*), we retain the technology but remove points that are obvious outliers. Figure 4 shows examples of removed time-series (figures 4g–4i show those where we removed specific data points and used the remaining time-series for analysis). After Step 4(b), 121 technologies remain. See `data/cleaned/step_4b`

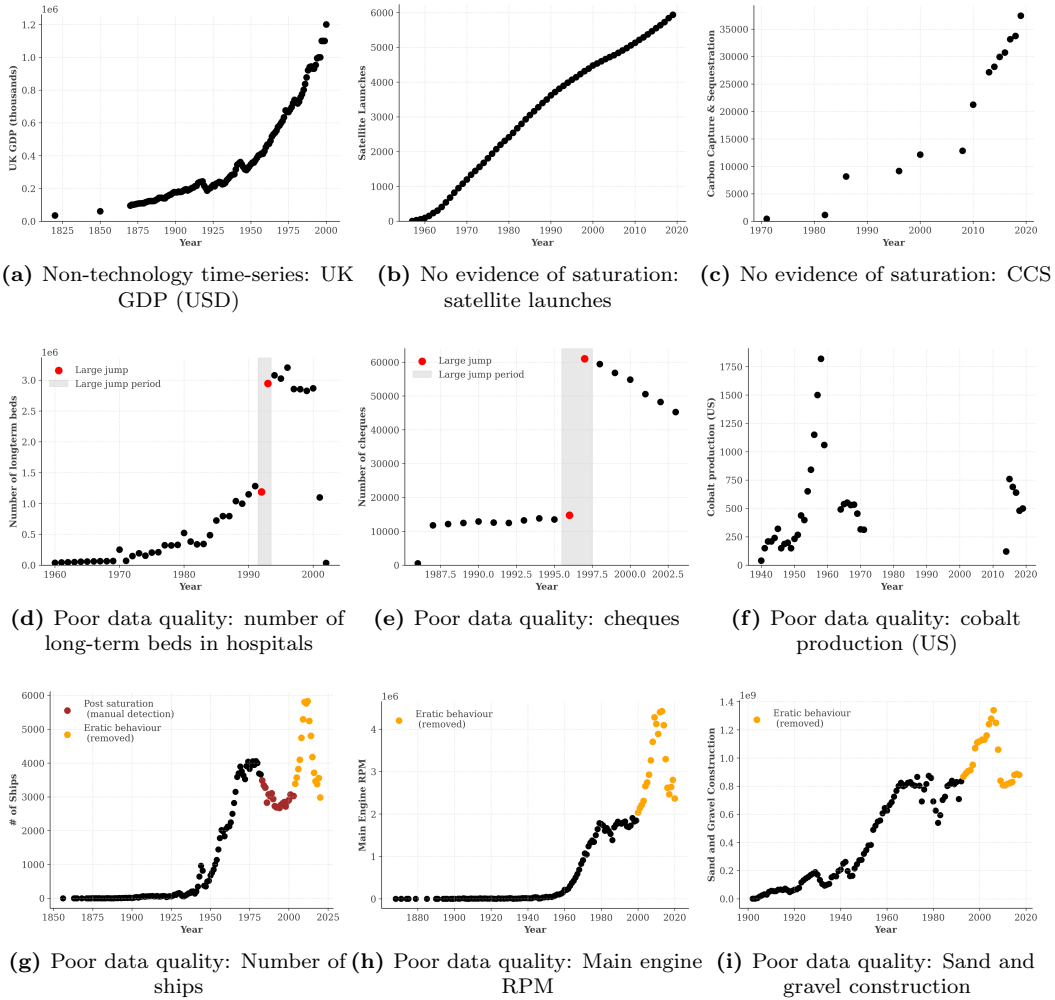


Figure 4: Examples of time-series removed during Step 4(b). (a) Non-technology. (b-c) Growth without clear saturation. (d-e) Sudden jumps to saturation. (f) Persistent fluctuations. (g-i) Erratic fluctuations in final data points (orange: removed in Step 1; black: retained; red: removed manually).

and `data_cleaning.ipynb` for complete removal details and plots.

Comparison of dataset properties before and after Step 4(b). Figure 5 compares the adjusted R^2 distribution and the normalized residuals before Step 4(b) to the final dataset. The left panel shows that removing the 34 technologies marginally increases the median adjusted R^2 (from 0.93 to 0.94), confirming that most removed series had lower fit quality. The right panel shows boxplots of normalized residuals (Eq. 43, under the GFP best fit with $\beta = 2/3$) binned by nondimensional time. The residual distributions are nearly identical between the two datasets, indicating that the removal does not materially alter the overall statistical properties of the data.

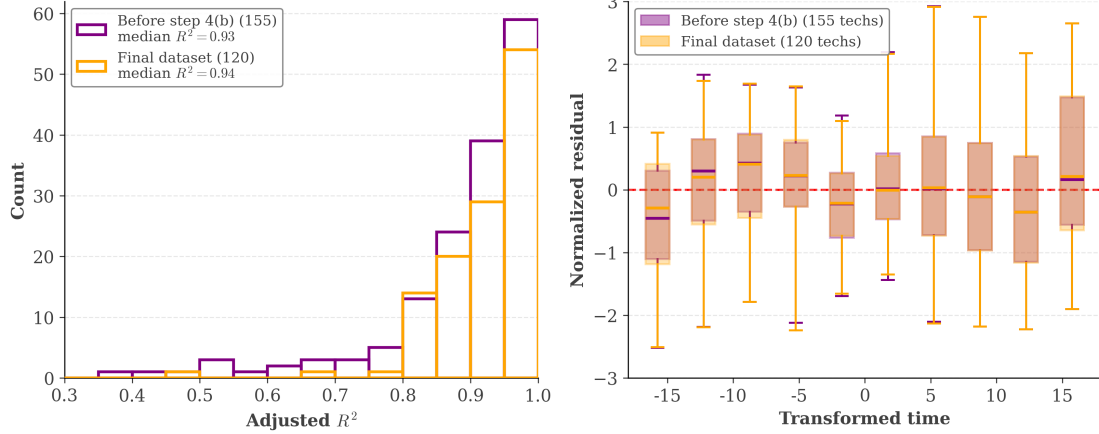


Figure 5: Comparison of dataset properties before and after Step 4(b). **Left:** Histogram of adjusted R^2 for the 155 technologies before Step 4(b) (purple) and the final 120 technologies (orange). The median R^2 increases slightly from 0.93 to 0.94. **Right:** Boxplots of normalized residuals (2.5th–97.5th percentile whiskers) binned by nondimensional time (same transformation as for main paper figure 2(b) and (c)). The large overlap across nondimensional time bins indicates that the distribution of residuals is nearly identical before and after Step 4(b), suggesting that the removal of technologies does not materially alter the overall residual distribution.

2.0.3.6 Step 5: Manual identification and removal of outliers

Manual inspection identifies 9 outlier data points for removal: one from *capacity of ships (UK)*, one from *capacity of steamships (UK)*, three from *primary bauxite production (US)*, and four from *gold reserves (South Korea)*. Figure 6 shows the removed outliers. The ship capacity outlier in 1910 (Figures 6a, 6b) is likely a data entry error. Bauxite production anomalies during 1942–1944 (Figure 6c) reflect WWII aluminum demand [113] and subsequent peacetime conversion [114]. Gold reserve outliers in South Korea (Figure 6d) correspond to financial difficulties [115, 116] preceding the 1997 IMF agreement [117], followed by a 1998 gold-collecting campaign [118]. These outliers reflect extraordinary economic events unrelated to the diffusion process itself; their removal corresponds to criterion (4) of Step 4(b).

11

¹¹The time-series excludes 2011–2013 gold purchases by the South Korean government [119].

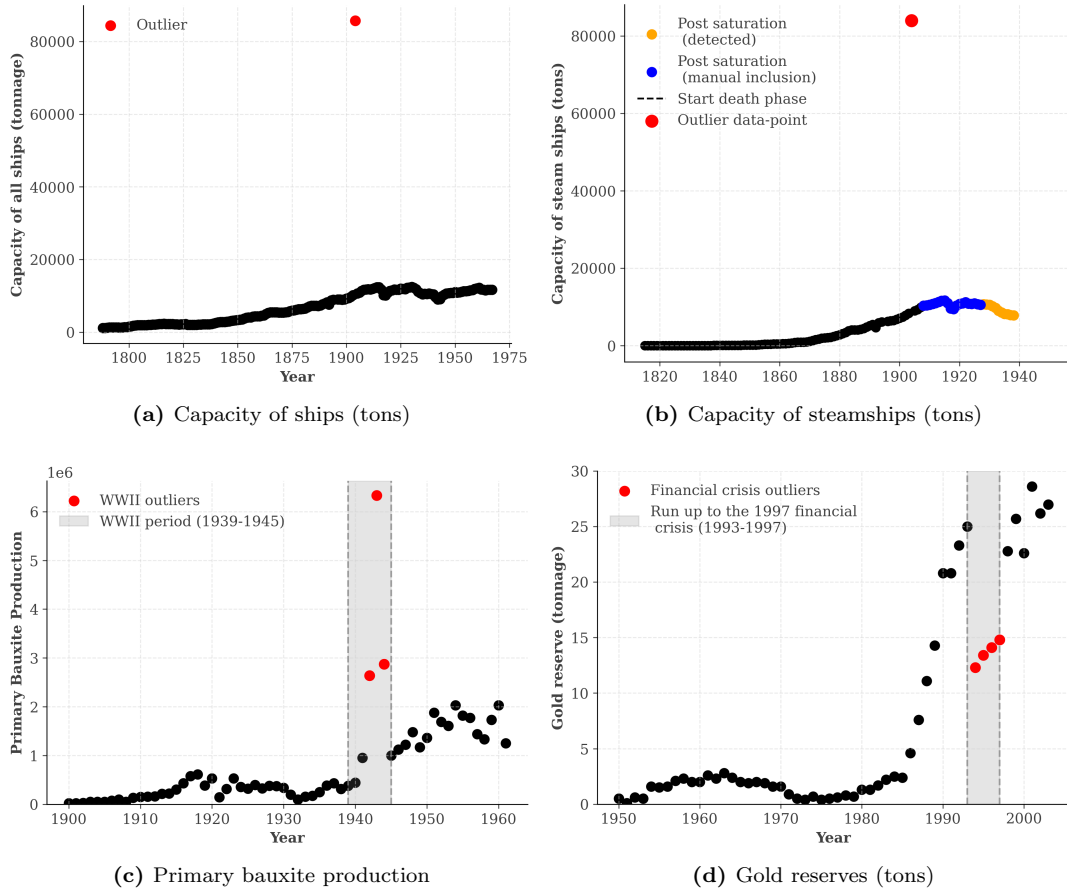


Figure 6: Diffusion curves for technologies where outliers were removed. (a) Capacity of ships (tons) in the UK with one outlier removed in 1910. (b) Capacity of steamships (tons) in the UK with outlier removed in 1910. Outlier is likely a data collection error. Death phase detection was too stringent (blue data are those data-points detected as post-saturation that were manually included again in step 1). (c) Primary bauxite production in the US with three outliers removed: World War II disruption pattern. (d) Gold reserves (tons) in South Korea with four outliers removed, due to 1993-1997 lead up to financial crisis.

Special case: Cane sugar consumption (Japan). For *cane sugar consumption (Japan)*, removal is more extensive. Figure 7 shows the time-series with removed data points highlighted. During WWII, cane sugar consumption in Japan dropped sharply due to supply disruptions and rationing [120,121], falling to a historical low [122] from approximately 15 kg per capita [123]. Post-war recovery increased consumption, reaching pre-war levels by the mid-1950s [122,124]. We remove all data before 1943 and treat diffusion as starting post-WWII. By the 1990s, Japan experienced long-term decline in sugar consumption [125], indicating death phase onset, reaching a per-capita low in consumption in 2010 [126].^{12,13} We analyze data from 1943 to the early 2000s.¹⁴

¹²Death phase is shown to indicate the full extent of removed data.

¹³Cane sugar consumption stabilized around 17 kg per person in 2021 [126] (post-2010 data not included).

¹⁴Per capita sugar consumption peaked in the early 1970s [126,127], then declined, aligning with trends in Figure 7.

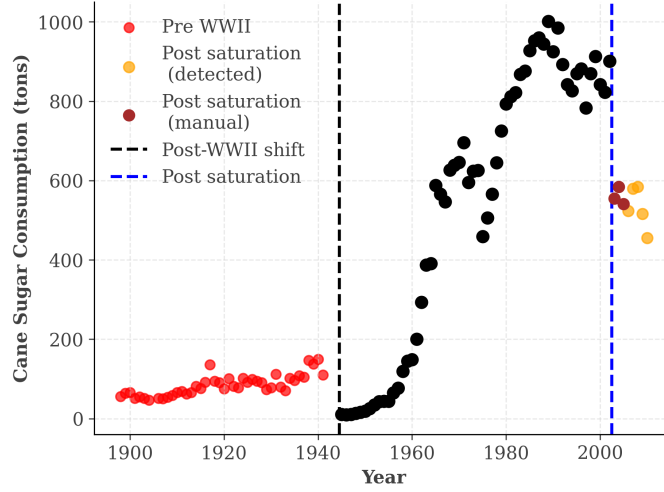


Figure 7: Cane sugar consumption (kg per capita) in Japan. Red dots are pre-WWII removed data points. Orange dots are post-2006 removed data points (post-saturation) that were detected in Step 1. Brown dots are additional post-saturation points that were manually removed in Step 1, for death-phase detection checks.

2.0.3.7 Step 6: Removing highly correlated time-series

Some technologies may be highly correlated (e.g., different vaccines in one country). We calculate Pearson correlation coefficients between residuals of all technology pairs (i, j) :

$$r_{ij} = \frac{\sum_t (R_{i,t} - \bar{R}_i)(R_{j,t} - \bar{R}_j)}{\sqrt{\sum_t (R_{i,t} - \bar{R}_i)^2 \sum_t (R_{j,t} - \bar{R}_j)^2}},$$

where $\bar{R}_i = \frac{1}{N_i} \sum_t R_{i,t}$ is the mean residual for technology i . We investigate technologies where $r_{ij} > 0.95$. We identify one such pair. We remove 1 time-series: *Polio vaccine (India)*.¹⁵

The final dataset includes 120 unique technologies; Table 17 lists each technology with its region, primary data source, observation count, and year span.

Table 17: The 120 cleaned technology time-series. Display names are abbreviated; full identifiers are kept in the data repository. N_i is the number of cleaned observations; the span gives the first and last observed year. Citations are to the database from which each series was drawn; where known, the original primary source is also indicated.

Technology	Region	Source	N_i	Span
Air conditioning (cars)	US	Grübler [96,97]	30	1954–1984
All Biofuels	Germany	HATCH [91]	30	1992–2021
Ammonia synthesis (gas feed-stock)	World	Rouwenhorst et al. [128] / RethinkX [99]	18	1935–2020
Ammonia synthesis (HFO/naphtha)	World	Rouwenhorst et al. [128] / RethinkX [99]	11	1935–1985
Ammonia synthesis (total)	World	Rouwenhorst et al. [128] / RethinkX [99]	22	1915–2020
Aquaculture Production	UK	HATCH [91]	61	1960–2020
ATMs (thousands)	World	Bessen [129] / RethinkX [99]	27	1973–2009

(continued on next page)

¹⁵We retain *DTP3 vaccine (India)* (adjusted $R^2 = 0.01$ higher; other metrics identical).

(continued from previous page)

Technology	Region	Source	N_i	Span
Automobile tire rims (straight side)	US	Klepper [130] / RethinkX [99]	24	1910–1933
Aviation tonne-kilometers	UK	CHAT [91]	47	1945–1991
BCG Vaccine	India	HATCH [91]	41	1981–2021
Beer Production	Japan	HATCH [91]	102	1896–1997
Biogas	Germany	HATCH [91]	22	2000–2021
Bone marrow transplants	World	CHAT [91]	28	1974–2001
Cable TV subscriptions	UK	CHAT [91]	17	1985–2001
Cadmium Refining	US	HATCH [91]	61	1906–1966
Canals	US	Comin & Hobijn [91]	42	1800–1854
Cane Sugar	Japan	HATCH [91]	58	1945–2002
Capture Fisheries	China	HATCH [91]	58	1960–2017
Caustic Soda	South Korea	HATCH [91]	43	1961–2003
Cement	US	HATCH [91]	120	1900–2019
Cigarette consumption per person	US	US DOT / RethinkX [99]	87	1881–1967
Coal (useful energy)	World	Way et al. [49] / OWID	15	1800–1940
Coal Production	China	HATCH [91]	41	1981–2021
Cobalt Mine Production	World	HATCH [91]	22	1995–2016
Coke (Prussia)	Germany	Landes [131] / RethinkX [99]	26	1825–1890
Color TV broadcasting (countries)	World	Wikipedia / RethinkX [99]	54	1950–2003
Commercial vehicles	US	Comin & Hobijn [91]	99	1904–2002
Containerised ports (cumulative)	World	Rua [132] / RethinkX [99]	28	1956–2003
Copper—Refining	India	HATCH [91]	22	1990–2011
Crop Harvester	South Korea	HATCH [91]	41	1961–2001
Crude oil (useful energy)	World	Way et al. [49] / OWID	24	1870–1978
DDT production	World	Schenker et al. [133] / RethinkX [99]	25	1941–1965
Diesel locomotives in service	US	Comin & Hobijn [91]	44	1925–1970
Dishwashers	US	HATCH [91]	40	1922–2011
Disk Brakes	US	HATCH [91]	19	1966–1984
DTP3 Vaccine	India	HATCH [91]	42	1980–2021
Electric steel production	US	USGS / RethinkX [99]	96	1896–1991
Electricity production	UK	CHAT [91]	98	1896–1993
Ethanol in motor fuel	US	US EIA / RethinkX [99]	39	1981–2019
Ethanol Production	World	HATCH [91]	51	1970–2020
Fax machine shipments (thousands)	World	NRC [134] / RethinkX [99]	15	1977–1991
Fertilizer (total)	World	CHAT [91]	42	1960–2001
Filter cigarettes (				
Flush Toilet	US	HATCH [91]	12	1860–1989
Gas (useful energy)	World	Way et al. [49] / OWID	67	1890–2023
GE corn (HT)	US	USDA / RethinkX [99]	28	1996–2023
Geothermal Energy	Germany	HATCH [91]	18	2004–2021
Gold	South Korea	HATCH [91]	50	1950–2003
Grain combines (thousands)	US	RethinkX [99]	24	1910–1960
Heart transplants	World	CHAT [91]	23	1978–2000
HIB3 Vaccine	World	HATCH [91]	31	1991–2021
Home Air Conditioning	US	HATCH [91]	55	1957–2011

(continued on next page)

(continued from previous page)

Technology	Region	Source	N_i	Span
Home dialysis (kidney)	UK	CHAT [91]	21	1970–1991
Household Internet Access	US	HATCH [91]	24	1993–2016
Hybrid corn (				
Hybrid corn acres	US	USDA / RethinkX [99]	26	1933–1958
Hybrid corn farmers (cumulative)	World	Rogers [6]; Ryan & Gross [135] / RethinkX [99]	15	1927–1941
Hydrochloric Acid	Japan	HATCH [91]	74	1930–2003
Hydropower (useful energy)	World	Way et al. [49] / OWID	67	1890–2023
Internet users	UK	FRED / RethinkX [99]	32	1990–2021
Kidney transplants	World	CHAT [91]	39	1963–2001
Lithotriptor units	World	CHAT [91]	20	1981–2000
Liver transplants	World	CHAT [91]	25	1977–2001
Long-term care beds	World	CHAT [91]	41	1960–2000
Lung transplants	World	CHAT [91]	18	1983–2000
Mammography units	World	CHAT [91]	30	1968–2000
Marine engine RPM	World	Mitchell shipping archive	111	1868–1999
Mechanical cotton harvesting	US	USDA / RethinkX [99]	24	1949–1972
Microcomputers	US	HATCH [91]	33	1984–2016
Milk Production	Japan	HATCH [91]	106	1899–2010
Monasteries (Europe)	Europe	Janauschek [100]	51	1113–1165
Motorship tonnage	World	CHAT [91]	69	1906–1974
MRI units	US	Comin & Hobijn [91]	18	1983–2000
Natural gas consumption	UK	RethinkX [99]	17	1962–1978
Nitric Acid	World	HATCH [91]	75	1929–2003
Nitrogen Fertilizer	India	HATCH [91]	60	1961–2020
Nuclear electricity (useful energy)	World	Way et al. [49] / OWID	38	1965–2002
Nuclear Weapons	World	HATCH [91]	66	1945–2010
Number of ships	World	Mitchell shipping archive	117	1856–1982
O-heath	US	Grübler [96, 97]	84	1874–1957
Oil Production	UK	HATCH [91]	23	1965–1987
Oil Refineries	World	HATCH [91]	68	1940–2007
Oxygen production	US	Grübler [96, 97]	26	1958–1983
Petroleum refining (catalytic cracking)	World	Linstone & Sahal [136] / RethinkX [99]	34	1938–1971
Phone subscriptions	World	World Bank	41	1980–2020
Phosphate Fertilizer	India	HATCH [91]	60	1961–2020
Pineapple production	Ivory Coast / Costa Rica	FAO / RethinkX [99]	61	1961–2021
Potash Fertilizer	China	HATCH [91]	59	1961–2020
Power steering (cars)	US	Grübler [96, 97]	34	1951–1984
Primary Aluminum Production	US	HATCH [91]	92	1900–1991
Primary Bauxite Production	US	HATCH [91]	59	1900–1961
Primary school enrollment	UK	CHAT [91]	54	1860–1913
Radial tire shipments (				
Radiation therapy equipment	World	CHAT [91]	41	1960–2000
Radio	South Korea	HATCH [91]	52	1948–1999
Railroad Tracks	US	HATCH [91]	123	1830–1952

(continued on next page)

(continued from previous page)

Technology	Region	Source	N_i	Span
RCV1 Vaccine	World	HATCH [91]	42	1980–2021
Retail locations accepting card	US	Comin & Hobijn [91]	16	1988–2003
ROTAC Vaccine	World	HATCH [91]	16	2006–2021
Sailing-ship tonnage	World	CHAT [91]	102	1788–1889
Salt production	US	HATCH [91]	118	1900–2017
Sand and Gravel Construction	US	HATCH [91]	91	1902–1992
Secondary school enrollment	UK	CHAT [91]	86	1886–1990
Sewing machines per person	US	Hounshell [137] / RethinkX [99]	43	1853–1895
Sexed dairy semen (				
Ship tonnage (all)	UK	CHAT [91]	178	1788–1967
Steamship tonnage	UK	CHAT [91]	111	1815–1927
Stoves	US	HATCH [91]	108	1900–2011
Sulphuric Acid	Canada	HATCH [91]	84	1919–2002
Synthetic filaments	World	HATCH [91]	81	1910–1990
Telegraphs	World	Grübler [96,97]	87	1856–1942
Telephones	France	HATCH [91]	97	1889–1985
Television	China	HATCH [91]	41	1959–1999
Tractor adoption	Canada	Olmstead & Hayami / RethinkX [99]	63	1909–1971
Transatlantic passengers by air (share)	World	Gilbert (2008) / RethinkX [99]	27	1948–1974
Travel by car (share of total)	US	RethinkX [99]	66	1895–1960
Unleaded gasoline (share)	US	Kerr & Newell [138] / RethinkX [99]	26	1970–1995
Wet Flue Gas Desulfurization Systems	Germany	HATCH [91]	28	1972–1999
YFV Vaccine	World	HATCH [91]	25	1997–2021
Zinc	Canada	HATCH [91]	93	1911–2003

Augmenting the dataset with energy carriers in TWh. For the empirical-regularity diagnostics in Section 3.2 and the forecasting exercises in Sections 4.2–5.2, we additionally retain 11 useful-energy time-series in TWh covering the major fossil and low-carbon energy carriers (coal, oil, gas, nuclear, hydropower, biomass, solar PV, and wind, among others; the full list is in the data repository). These series are not part of the 120-technology set used for the calibration analysis in Section 4.1.

Multi-source construction for solar PV and wind. For solar PV and wind we combine several sources. Useful-energy values (PWh) come from Way et al. [49] and Our World in Data [101,102]. This useful-energy series provides the historical data plotted in main paper Figure 4a. Installed capacity (TW) comes from IRENA [139], extended further back using Nemet [140] and Farmer and Lafond [141].

2.0.4 Behavior of the clean data

We present sector representation, normalized residuals, and R^2 values of cleaned time-series.

Sector classification in raw and cleaned datasets. Figure 8 shows sector representation in cleaned and raw datasets. We perform a manual classification into 11 sectors, based on the titles of the time-series. The sectors can be seen in Figure 8 and Table 18. The manual classification procedure can be found in the script, `data_cleaning_analysis.ipynb`. For inspiration on the classification, we use Refs. [103] for help with identifying *industry*, *infrastructure*, and *transport*

technologies; [104] for *energy*; [105] for *health*; [106] for *agriculture*; [107] for ICT; and [108] for minerals in *industry*. We added *financial services*, *military*, and *other services* for miscellaneous technologies (e.g., monasteries).

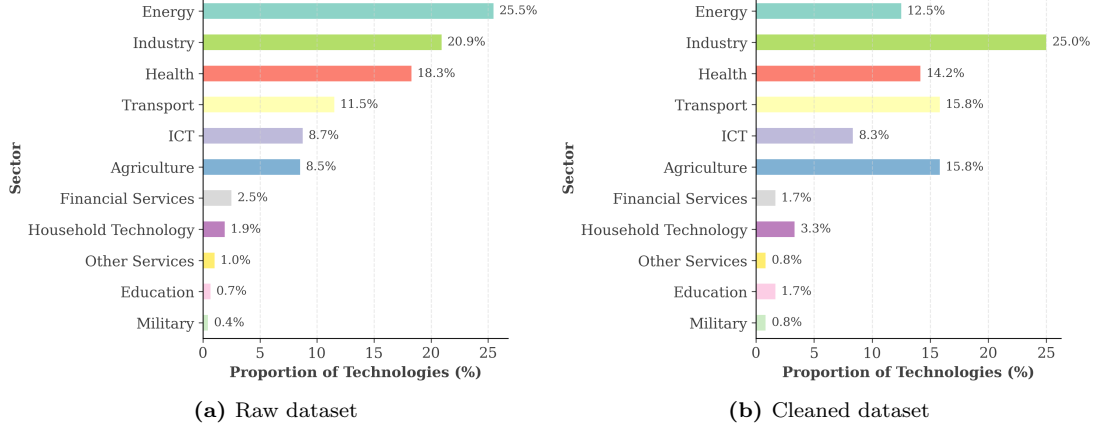


Figure 8: Sector representation in (a) raw dataset of 1362 technologies and (b) cleaned dataset of 120 technologies. *Industry* and *Health* are reasonably large sectors in both datasets. The cleaned dataset has no uncategorized technologies, while the raw dataset contains some. The differences in proportions are likely due to duplicate technologies being removed, uncategorized time-series (*other services*) in the raw dataset, and some sectors having many technologies that have not reached their saturation phase (*Energy*), while others have more mature technologies that properly saturate (*Transport*, *Agriculture*).

Table 18 shows technologies per sector in the final dataset.

Sector	Number of technologies	Examples
Industry	30	textile machinery, steel, and cement production
Transport	19	automobiles, railways, aircraft, shipping capacity
Agriculture	19	fertilizer, tractors, hybrid corn acreage
Health	17	vaccinations, surgeries, medical devices
Energy	15	gas, coal, oil, electricity production
ICT	10	mobile phones, internet users, television, radio
Household technologies	4	dishwashers, stove, flush toilet
Financial services	2	retail locations accepting card, ATMs
Education	2	primary school enrollment
Military	1	nuclear weapons
Other services	1	monasteries

Table 18: Sector representation in the cleaned dataset of 120 technologies.

The raw dataset contains 14 uncategorized technologies, which we place in *other services*, while the cleaned dataset has 1 (*Monasteries - Europe*). The uncategorized technologies in the raw dataset include *GDP*, *public investment*, *private investment*, and *population* for different countries. We removed these in the cleaning process as they do not adequately represent technology adoption.

The cleaned dataset has a more even sector distribution, with a higher proportion of *Industry*, *Transport*, and *Agriculture* technologies, likely because these sectors include more mature technologies. Some sectors, such as *Energy* and *ICT*, have a lower proportion in the cleaned dataset because many technologies have not reached saturation (e.g., solar, wind, CCS). Furthermore, many *ICT* technologies have extensive coverage in the raw dataset due to duplicates (e.g., mobile phones in different countries), which are removed (except for 1) in the cleaned dataset.

Adjusted R^2 and normalized residuals of cleaned dataset. Figure 9 shows the adjusted R^2 distribution for 120 technologies¹⁶, the normalized residuals, and a comparison between observed residuals and normal and Student-t (df = 7) distributions¹⁷. Table 19 summarizes key statistics of normalized residuals for various cleaning steps.

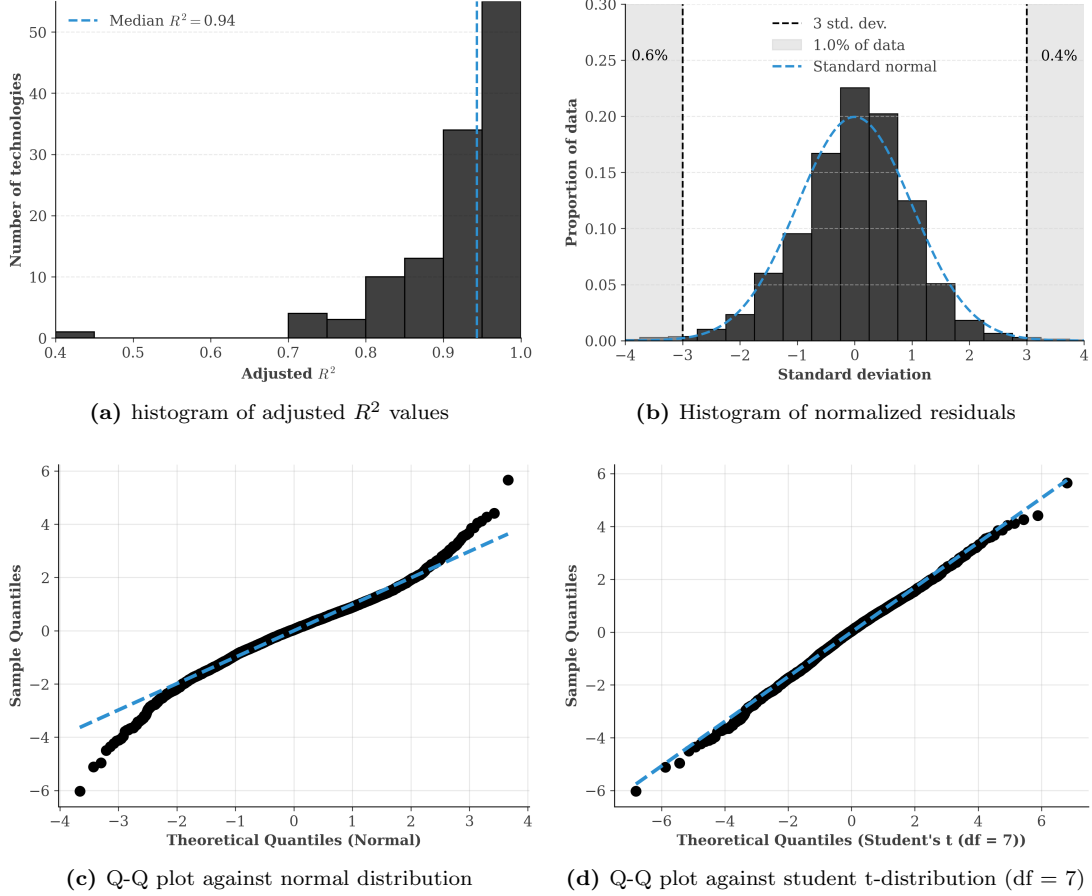


Figure 9: (a) Histogram of adjusted R^2 values for the cleaned dataset of 120 technologies, showing most technologies have high explained variance. (b) Histogram of normalized residuals shows slight negative skew and heavier negative tails. (c) Q-Q plot comparing normalized residuals to standard normal distribution and (d) Student-t distribution with 7 degrees of freedom (df = $N_i - 4$, where $N_i = 11$, is the length of the shortest time-series in the dataset). The residuals deviate from normality in the tails, particularly on the negative side.

Figure 9a indicates that the curve explains $> 90\%$ of the variance for most technologies, with median adjusted $R^2 = 0.94$.¹⁸ One technology has adjusted $R^2 < 0.5$ (*Gold reserves (South Korea)*, $R^2 = 0.45$).

Figure 9b shows a histogram of normalized residuals. Table 19 shows a slight negative skew (-0.3) and a positive excess kurtosis (1.7). A positive excess kurtosis means the residuals have heavier tails than a normal distribution. This is common in real data. Consistent with this, 1% of the residuals lie past 3 standard deviations, where a normal distribution expects 0.3%. Most technologies track the B-R curve closely. A few show anomalous episodes, such as early rapid adoption or structural breaks, that produce occasional extreme residuals. The Q-Q plot against a normal distribution (Figure 9c) confirms this, with deviations from normality in the tails. The residuals match a Student- t distribution with 7 degrees of freedom (Figure 9d) more closely, which

¹⁶For a comparison to other traditional S-curves (Logistic and Gompertz) see Section 3.3.

¹⁷df = $N_i - 4$, where $N_i = 11$, is the length of the shortest time-series in the dataset, and 4 is the number of estimated parameters, $\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\beta}_i$.

¹⁸Results remain unchanged when assuming universal $\beta = \frac{2}{3}$.

Statistic	Step 2	Step 3	Cleaned Dataset	Std. Normal
Mean	0	0	0	0
Median	0.04	0.06	0.05	0
Std. Dev.	1	1	1	1
Skewness	-0.3	-0.3	-0.3	0
Excess kurtosis	2.4	2.9	1.7	0
KS Statistic	0.03	0.05	0.04	0
p-value	< 0.001	< 0.001	< 0.001	N/A

Table 19: Key statistics of the normalized residuals at different cleaning stages: after Step 2 (958 time-series), after Step 3 (457 time-series), and the cleaned dataset (120 time-series). Kurtosis is excess kurtosis, measured relative to the standard normal, which has excess kurtosis 0. The residuals have a slight negative skew and a positive excess kurtosis that reflects heavier tails.

also has heavier tails (excess kurtosis 2). The deviations are larger in the negative tail (black dots below the blue line at negative values in Figure 9d). These may stem from the systematic factors discussed earlier (Figure 3). Some technologies grow faster than the model early on. The model then overestimates the early growth rate and produces extreme negative residuals. It underestimates later diffusion points, which gives a median above 0. Section 3.2 discusses faster-than-expected early growth.

2.1 Limitations

Our data selection procedure faces several limitations.

Selection bias toward successful technologies. While we gather as much data as possible from secondary sources, the data collected in these sources tends to focus on successful technologies.

Subjective selection decisions. Several steps in the cleaning procedure involve subjective judgment: manual correction of death-phase breakpoints (Step 1), manual flagging of poor-quality series (Step 4b), and manual outlier removal (Step 5). While we validate subjective decisions with quantitative checks where possible, these choices introduce analyst degrees of freedom that could in principle affect the results. We have aimed to be transparent about each such decision and its rationale, and show that after each of these steps the change in the distribution of normalized residuals is minimal, and change in quality of fit (measured by adjusted R^2) is negligible.

Data-construction assumptions. When aggregating country-level data to a global proxy, missing values are treated as zeros (Section 2.0.1), which may understate early adoption levels if reporting coverage is incomplete. Similarly, converting flow data to cumulative stocks assumes zero initial stock, which can bias levels if a technology was already present before the sample begins. Both effects are most pronounced in the early part of the time-series—the segment most relevant for the predictability exercise.

Subjective technology definitions. Determining when a time-series is or is not a technology may be difficult. We have not defined a strict criterion for what constitutes a technology, and we have made some subjective decisions in this regard. For example, we have removed *GDP* as it does not represent technology adoption, but we have retained *monasteries* as they represent a form of organizational technology. Other non-technology time-series are present, but were filtered out in earlier steps.

Conditioning on completed S-curves. Our model assumes an S-curve diffusion pattern, so we automatically exclude technologies that do not (or do not yet) feature a slowing down of diffusion. In other words, our dataset includes only technologies that have gone through a full diffusion cycle. The filtering criteria in Steps 2–4 (minimum data points, requiring $> 80\%$ of $\hat{L}$, requiring early-phase coverage) further condition the sample on well-behaved, already-mature technologies. As a consequence, the universality and backtest results reported in this paper are conditional on technologies whose diffusion is well-described by an S-curve. S-curve diffusion is therefore an *assumption* of our model. Conditional on this assumption, it is a *result* of our paper that a technology’s diffusion can be predicted with a well-calibrated distributional forecast using

the B-R model with technology-specific and universal parameters, even from early data alone. We note that this does not exclude technologies that were later substituted out (e.g., *telegrams*). That is, we model only the first three S-curve phases of net-stock growth, not the full birth-to-death life cycle, as the death phase is trimmed in Step 1 and the preceding diffusion is retained (Section 1.1).

3

In-sample data analysis

3.1 Defining technology-invariant parameters

The full B-R model we consider here is a highly nonlinear model with a total of 7 free parameters: L , k , t_0 , σ , β , ρ and ϵ_T . The terminal residual ϵ_T enters the model because the MA(1) noise carries the last observed shock forward into the first forecast step; it could in principle be estimated technology-specifically, but doing so adds a free parameter for each technology. Combined with the small number of data points (often as few as 10), fitting all these parameters may very likely result in overfitting. To improve this situation we take three parameters as technology-invariant, i.e. equal for all technologies: $\beta = 2/3$, $\rho = 0.80$, and $\epsilon_T = 0$. This introduces some bias (real technologies may differ in these parameters) but reduces variance and improves forecasts. We show below that observed variation in these parameters across our data set is small enough that real differences are hard to discern given the statistical error of short series. This section justifies these choices through theoretical considerations, empirical evidence, and support from the literature.

3.1.1 Asymmetry parameter

The asymmetry parameter β determines S-curve skewness: $\beta = 1$ yields symmetric curves (logistic), $\beta < 1$ yields right-skewed curves (inflection before 50% diffusion), $\beta > 1$ yields left-skewed curves (inflection after 50% diffusion). Here, the inflection point refers to the true inflection (maximum growth rate), which occurs at $Y_{\text{inf}}/L = (1 + \beta)^{-1/\beta}$. Note that this differs from $Y(t_0)/L = 2^{-1/\beta}$, since t_0 is not the inflection point for $\beta \neq 1$. Both expressions equal 50% at $\beta = 1$.

Literature context. Mahajan et al. [13] review diffusion models, noting S-curve shape variations across technologies. Meade and Islam [53] conclude that the shape of the S-curve is one of the most important and distinctive features of the diffusion process. Bertalanffy [142] introduced asymmetric growth models, showing right-skewed S-curves fit biological growth data. Rogers [6] and Moore [143] identify that near-exponential adoption occurs in the first 16% of adoption. Grubler [96] documents exponential growth patterns in infrastructure diffusion during initial market penetration phases. Despite available asymmetric models [20], many studies such as the first study of technology adoption using S-curves [5] default to symmetric logistic curves for simplicity.

3.1.1.1 In-sample fit and choice of asymmetry parameter

We first analyze the in-sample fit for various values of a technology-invariant β , and then check how well a DGP with technology-invariant β replicates the empirical distribution of $\hat{\beta}_i$ obtained when we estimate β_i separately for each technology. We use the empirical surrogate datasets defined in Section 1.4, with β fixed to a candidate technology-invariant value (rather than estimated per technology) and the remaining $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ inherited from the empirical fit. We then re-fit

the B-R model with technology-specific β_i to each surrogate series and compare distributions. If the surrogate and empirical distributions of $\hat{\beta}_i$ are similar, the observed heterogeneity in $\hat{\beta}_i$ is consistent with estimation error rather than genuine cross-technology variation, and a technology-invariant β is justified for forecasting given the short data sets.

Goodness of fit for different values of β . Figure 10 shows fit quality across the 120-technology ensemble for four S-curve choices: logistic, Gompertz, B-R with technology-invariant $\beta = 2/3$, and B-R with technology-specific β_i . Figure 10a shows that all S-curve types yield similar adjusted

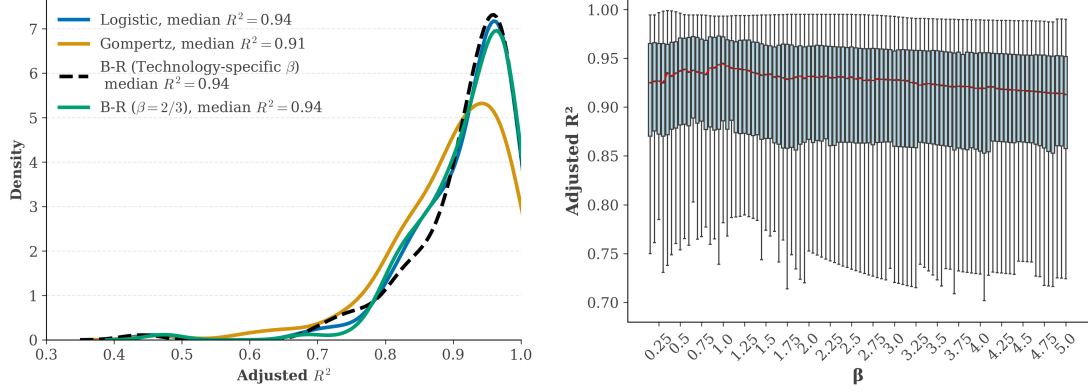


Figure 10: (a) Comparison of R^2 across Gompertz, logistic, and B-R curves (technology-specific β_i and technology-invariant $\beta = 2/3$). (b) Box-plots of adjusted R^2 of 120-technology ensemble for various values of a technology-invariant $\beta \in [0.05, 5]$.

R^2 distributions, with median values above 0.9, indicating high fit quality across models (though the Gompertz model does somewhat worse). B-R with varying β_i yields similar median R^2 to technology-invariant β models ($\beta = 1$ and $\beta = 2/3$), indicating that estimating technology-specific β_i does not substantially improve in-sample fit quality.

Figure 10b focuses on fit quality when using a technology-invariant β . Adjusted R^2 is relatively stable across β , but very small ($\beta < 0.3$) and very large ($\beta \gg 1$) values yield slightly lower R^2 . This implies that a wide range of S-curve models provide reasonable fits, and in-sample R^2 alone is insufficient for model selection. Section 3.3 compares Gompertz, logistic ($\beta = 1$), and Bertalanffy-Richards ($\beta = 2/3$) curves in more detail.

Replicating the distribution of estimated β_i . Given the short data sets, we expect substantial heterogeneity in *estimated* β_i even if all the technologies had the same true value of this parameter, i.e. if the *population* $\beta_i = \beta$, $\forall i$. We assess whether generating data using a technology-invariant $\beta_i = \beta$ can lead to an empirical distribution of estimated β_i similar to what we observe empirically, following Algorithm 3.

Algorithm 3: Assessing whether a DGP with technology-invariant β replicates the empirical distribution of β_i .

Input: Empirical technology time-series $\{Y_{i,t}\}_{i=1}^{120}$

Output: Distribution of estimated β_i from surrogate data

Step 1: Estimate parameters from empirical data

foreach *technology* $i = 1$ *to* 120 **do**

 Fit model with β to obtain:

$\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$

 ($\hat{L}_i$ via NLLS; apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ and regress on time to estimate other parameters)

end

Step 2: Generate and fit surrogate data (M simulations)

foreach *simulation* $s = 1$ *to* M **do**

foreach *technology* $i = 1$ *to* 120 **do**

 Generate $Y_i^{(s)}(t)$ using DGP (Eq. 14) with $\hat{\theta}_i$, β , $\rho = 0.8$, and $\epsilon_T = 0$

 Fit Bertalanffy-Richards curve to $Y_i^{(s)}(t)$ to estimate $\hat{\beta}_i^{(s)}$

end

end

Step 3: Compare distributions

 Compute Kolmogorov-Smirnov distance:

$D_{KS}(\beta) = \sup_x |F_{\text{real}}(x) - F_{\text{surrogate}}(x)|$

Figure 11 compares $\hat{\beta}_i$ values estimated from surrogate data generated under a technology-invariant B-R model with $\beta \in \{0.5, 2/3, 1\}$. We show two probability histograms: (1) the full histogram with all estimated values, and (2) one cropped to $\beta_i < 2.41$. The clip at $\beta_i = 2.41$ corresponds to $Y(t_0)/L = 2^{-1/\beta} < 75\%$ (recall t_0 is not the inflection for $\beta \neq 1$; the true inflection at $Y_{\text{inf}}/L = (1 + \beta)^{-1/\beta}$ corresponds to $\approx 58\%$ diffusion at $\beta = 2.41$). This is consistent with the 80% minimum-diffusion data-selection threshold (Section 2). In the fitting procedure, we set an upper bound $\beta = 100$.

Table 20 shows the Kolmogorov-Smirnov (K-S) distance $D_{KS}(\beta)$ between estimated β_i distributions from real and surrogate data for various values of β , as well as other statistics.

Data source	Mean	Median	D_{KS} (full)	p-value (full)	D_{KS} ($\beta_i < 2.41$)	p-value ($\beta_i < 2.41$)
Empirical data	11.6	1.28	–	–	–	–
Surrogate ($\beta = 1/2$)	7.9	0.57	0.23	$< 10^{-4}$	0.14	0.10
Surrogate ($\beta = 2/3$)	8.6	0.72	0.19	6×10^{-4}	0.10	0.38
Surrogate ($\beta = 1$)	10.7	1.04	0.12	0.06	0.14	0.08
Surrogate ($\beta = 1.28$)	12.7	1.32	0.11	0.16	0.17	0.02

Table 20: Mean and median of estimated $\hat{\beta}_i$ values from empirical data and surrogate data (technology-invariant $\beta = 0.5$, $\beta = 2/3$, $\beta = 1$, and $\beta = 1.28$), along with Kolmogorov-Smirnov distances D_{KS} and p-values comparing surrogate to empirical data distributions. Lower K-S distances and higher p-values indicate better replication of the empirical distribution.

The empirical data shows substantial heterogeneity in estimated β_i values across technologies. The median estimated β_i is 1.28, with a mean of 11.6 due to a few high β_i outliers. Just over half of technologies have $\beta > 1$ (left-skewed S-curves; maximum growth after 50% diffusion) and the rest have $\beta < 1$ (right-skewed; maximum growth before 50%). This implies that there is no clear universal asymmetry pattern.

Figure 11 shows the histogram of individually estimated β_i with a cutoff at $\beta_i < 50$ for visualization purposes). The green histogram shows estimated β_i values from surrogate data generated with technology-invariant value of $\beta \in \{1/2, 2/3, 1.28\}$.

The surrogate data with technology-invariant $\beta = 1.28$ (11c),¹⁹ equivalent to the empirical

¹⁹Similar conclusions apply to $\beta = 1$ (logistic). We do not show the graphs, but we show the statistics in Table 20.

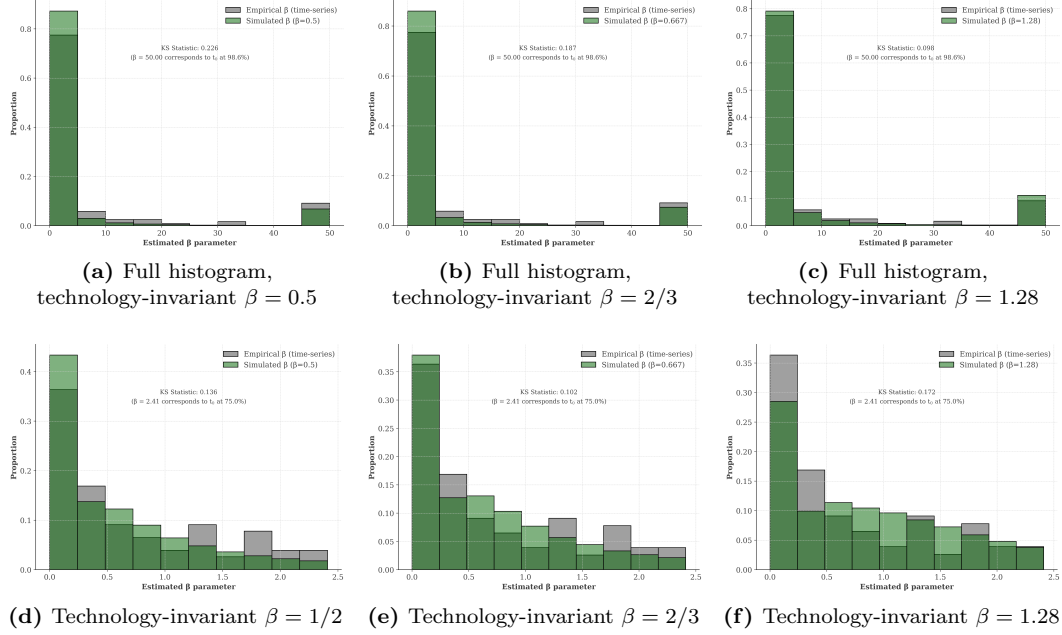


Figure 11: Histograms comparing $\hat{\beta}_i$ estimated from surrogate data (green) with technology-invariant $\beta = 2/3$ based on 1000 simulations for 120 technologies versus β_i estimated from empirical data (gray). The top panels show the histogram with the x-axis cropped at 50 for visualization purposes. The bottom panels are the same, but they are cropped with $\beta_i \leq 2.41$, corresponding to $Y(t_0)/L < 75\%$. The left, middle and right panels correspond to three different values of the technology-invariant β used to create the surrogate data.

median, has the lowest KS distance of those shown (Table 20), but it under-emphasizes the peak at low values of β ($\beta < 0.25$) in the empirical distribution (Figure 11f versus 11d and 11e). $\beta = 0.5$ and $\beta = 2/3$ (Figures 11d and 11e) both reasonably replicate the spread of estimated β_i values, the extreme outliers, and the peak at low β values. In both cases the surrogate slightly underestimates the empirical mean and median (Table 20), shown by the lighter green areas on the left of the histograms in Figures 11a and 11b.

Focusing on the distribution of estimated β_i values with $Y(t_0)/L < 75\%$ ($\beta_i < 2.41$, Figures 11d, 11e, and 11f), surrogate data with technology-invariant $\beta = 2/3$ best replicates the clipped empirical distribution (Table 20). $\beta = 1/2$ slightly over-emphasizes the peak at low β values ($\beta < 0.25$) in the empirical distribution compared to $\beta = 2/3$ (Figure 11d versus 11e).

We conclude that fixing β can reasonably replicate the empirical distribution of estimated β_i values, but no single value dominates across all criteria. When matching the full empirical distribution, $\beta = 1.28$ achieves the lowest KS distance (Table 20). However, the surrogate experiments demonstrate that β estimates are biased upward in the mean: surrogate data with true $\beta = 2/3$ produces mean $\hat{\beta} = 8.6$ (vs. median $\hat{\beta} = 0.72$), and true $\beta = 1.28$ produces mean $\hat{\beta} = 12.7$ (vs. median $\hat{\beta} = 1.3$), due to a heavy right tail of extreme outliers. Additionally, $\beta = 2/3$ best replicates the clipped distribution ($\beta_i < 2.41$, KS $p = 0.38$), while $\beta = 1.28$ is rejected under this clip ($p = 0.02$). $\beta = 0.5$ slightly over-emphasizes the peak at low β values, while $\beta = 1$ and $\beta = 1.28$ under-emphasize it.

3.1.1.2 Growth replication

We assess the performance of models with technology-invariant β in replicating observed growth patterns. The chosen value $\beta = 2/3$ is shown against the data in main paper Figure 2d.

Replicating observed growth patterns. For each candidate technology-invariant β , we apply Algorithm 8 with the B-R DGP at that β in place of the default $\beta = 2/3$, generating 1,000 surrogate

ensembles. From each ensemble and from the empirical pool we compute the cross-technology median growth trajectory (a sequence of medians indexed by diffusion level, not a smooth curve), and compare them via the sum of squared errors (SSE). We repeat the surrogate step 100 times and report the SSE distribution. We also report the proportion of empirical-median points falling within the 95% surrogate-median confidence interval; perfect replication gives 95%.

Figure 12 shows 100 SSEs as solid lines, with shaded regions indicating the 10th–90th percentiles for each β . Figure 12a shows that $\beta = 0.35$ yields the lowest SSE, indicating the best replication

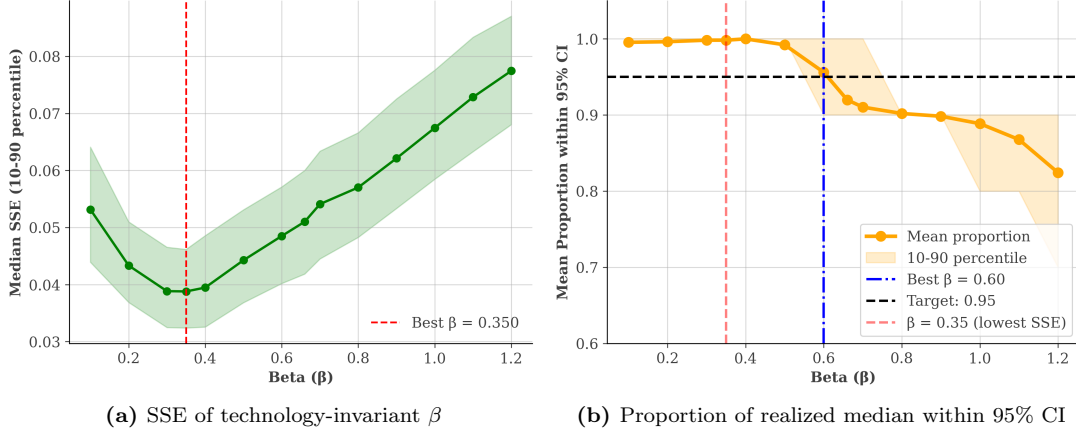


Figure 12: Sum of squared errors (SSE) between median growth trajectories from surrogate data and empirical data across varying technology-invariant β values. Lower SSE indicates better replication of observed growth patterns. **(a)** Median SSE across 100 repetitions of 1000 simulations of the 120-technology ensemble shown as solid line; shaded region indicates the 10th–90th percentile. **(b)** Plot showing the proportion of empirical median growth trajectory points that fall within the 95% confidence interval of the surrogate median growth trajectory for each technology-invariant β value. Shaded region indicates 10th–90th percentile across 100 repetitions. Perfect replication would yield 95% of points within the 95% CI.

of observed median growth patterns. SSE increases consistently for $\beta > 0.4$, with $\beta = 1$ (logistic) yielding a median SSE nearly 2 times higher than $\beta = 0.35$. Similarly for very low β values ($\beta < 0.2$), SSE increases. Figure 12b shows the proportion of points on the empirical median growth trajectory that fall within the 95% confidence interval of the surrogate median growth trajectory for each technology-invariant β value. Lower β yields higher coverage: $\beta < 0.4$ captures essentially all empirical points within the 95% CI, suggesting the surrogate CI is too wide. Higher β yields lower coverage: $\beta > 0.8$ captures fewer points, and $\beta = 1.5$ only 70–80%, indicating poorer replication of observed growth patterns. For $\beta \in [0.5, 0.8]$, the proportion fluctuates between 90–100%, with the median near 95%.²⁰ For $\beta = 0.6$ and $\beta = 0.66$, the mean proportion is 96% and 92% respectively, indicating good replication of observed growth patterns.²¹ Analyzing the 90% CI yields similar results, with $\beta \in \{0.6, 0.66\}$ both yielding mean proportions of 89%.

Conclusion: Justification for a technology-invariant β , with $\beta = 2/3$. Models with predetermined inflection points (Gompertz [28], Logistic [21]) and flexible inflection points (Bass [51], Richards [20]) all fit adoption data reasonably well [26, 27, 41, 144] (Section 3.3). However, the large variation in estimated β_i values indicates estimation uncertainty, which increases overfitting risk and degrades out-of-sample forecasts [40].

Multiple technology-invariant values of β yield similar fit quality (Figure 10b). When replicating the distribution of estimated β_i values (Figure 11), multiple technology-invariant β values work well, e.g., $\beta = 2/3$. Lower β slightly over-emphasizes the peak at low β values in the distribution.

²⁰The empirical median growth trajectory has 10 points, so we expect 9-10 points within the 95% CI. Over 1000 simulations, this yields a mean coverage of approximately 95%.

²¹The last empirical growth data-point tends to fall slightly outside the 95% CI, indicating potential systematic model misspecification. See Figure 16b.

$\beta > 2/3$ have lower K-S distances for the full distribution, but under-emphasize the peak at low β values. Out-of-sample growth replication (Figure 12a) indicates $\beta \in [0.2, 0.4]$ best replicates observed growth patterns (lowest SSE) favoring $\beta < 1$ (right-skewed S-curves), where $\beta = 0.35$ yields the lowest SSE. $\beta \in [0.5, 0.8]$ tends to yield the expected coverage of 95% of points within the 95% CI (Figure 12b), across the 1000 simulations.

No single value of β is best on every criterion. The full-distribution KS test favors $\beta = 1.28$, the growth-replication SSE favors $\beta \approx 0.35$, and the clipped-distribution KS test favors $\beta = 2/3$. We choose $\beta = 2/3$ as a pragmatic compromise: (i) the upward estimation bias documented above means the empirical median 1.28 overstates the true β ; (ii) $\beta = 2/3$ provides the best match to the clipped distribution; (iii) $\beta \in [0.5, 0.8]$ yields approximately the expected 95% coverage in the growth-replication exercise. Section 4.1 confirms that $\beta = 2/3$ yields reasonable out-of-sample forecasting performance: $\beta = 1/3$ and $\beta = 1.5$ are significantly worse, while $\beta = 0.5$ and $\beta = 1$ (logistic) perform similarly but slightly worse.

3.1.2 Moving average parameter

The MA(1) parameter ρ in our DGP (Eq. 14) controls temporal correlation in the noise: $\sigma_i(\rho\epsilon_{i,t-1} + \epsilon_{i,t})$, where $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$. This parametrization means the noise at time t depends on both the current white noise draw $\epsilon_{i,t}$ and the white noise draw from the previous time step $\epsilon_{i,t-1}$, weighted by ρ .

Literature support and justification. Baumgartner et al. [40] discuss positive autocorrelation ($\rho > 0$) in residuals of technology diffusion. Whether observed autocorrelation arises from unmodeled factors, model misspecification, or inherent adoption dynamics remains unclear. Autocorrelation is difficult to estimate accurately and is sometimes technology-invariant for parsimony [141]. We use a technology-invariant ρ rather than estimate technology-specific ρ_i values.

Estimating the MA(1) parameter from empirical data. To estimate ρ_i from empirical data, we fit the B-R model (Eq. 3) via NLLS to estimate $\hat{L}_i$ and $\hat{\beta}_i$, transform the coordinate to a linear model by applying the GFP transform $f(Y_{i,t}; \hat{L}_i, \hat{\beta}_i)$, regress the transformed data on time, and obtain residuals $S_i(t)$. We estimate ρ_i using Maximum Likelihood Estimation (MLE) via an ARIMA(0,0,1) model fitted to the residuals. We then assess how well a technology-invariant ρ replicates the empirical distribution of $\hat{\rho}_i$ using empirical surrogate datasets (Section 1.4), generated with the candidate technology-invariant ρ rather than the per-technology $\hat{\rho}_i$. Algorithm 4 summarises the procedure.

Algorithm 4: Assessing whether a technology-invariant DGP parameter ρ replicates the empirical distribution of estimated MA(1) parameters.

Input: Empirical technology time-series $\{Y_{i,t}\}_{i=1}^{120}$
Output: Distribution of estimated ρ_i values from surrogate data

Step 1: Estimate parameters from empirical data
 foreach *technology* $i = 1$ **to** 120 **do**
 Fit B-R curve to obtain $\hat{\rho}_i$:
 Fit $\hat{L}_i$ and $\hat{\beta}_i$ via NLLS; apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \hat{\beta}_i)$ and estimate other parameters via OLS
 Calculate residuals $S_i(t)$
 Estimate MA(1) parameter $\hat{\rho}_i$ using MLE via ARIMA(0,0,1) model fitted to residuals
 end

Step 2a: Generate surrogate data (M simulations)
 foreach *simulation* $s = 1$ **to** M **do**
 foreach *technology* $i = 1$ **to** 120 **do**
 Generate $Y_i^{(s)}(t)$ using the DGP (Eq. 14) with parameters $\hat{\theta}_i = \{\hat{L}_i, \hat{k}_i, \hat{t}_{0,i}, \hat{\sigma}_i\}$ (estimated using $\beta = 2/3$ and technology-invariant $\rho = \tilde{\rho}$), $\beta = 2/3$, technology-invariant ρ , and $\epsilon_T = 0$
 end
 end

Step 2b: Estimate $\hat{\rho}_i^{(s)}$ from surrogate data
 foreach *simulation* $s = 1$ **to** M **do**
 foreach *technology* $i = 1$ **to** 120 **do**
 Estimate $\hat{L}_i^{(s)}$ and $\hat{\beta}_i^{(s)}$ from NLLS regression on $Y_i^{(s)}(t)$
 Apply $f(Y_i^{(s)}, \hat{L}_i^{(s)}, \hat{\beta}_i^{(s)})$ (GFP transform) and regress on time
 Calculate residuals $S_i^{(s)}(t)$
 Estimate MA(1) parameter $\hat{\rho}_i^{(s)}$ using MLE via ARIMA(0,0,1) model fitted to residuals
 end
 end

Step 3: Collect estimated $\hat{\rho}_i^{(s)}$ values
 foreach *simulation* $s = 1$ **to** M **do**
 Store $\hat{\rho}^{(s)} = \{\hat{\rho}_i^{(s)}\}_{i=1}^{120}$
 end

Step 4: Collect estimated $\hat{\rho}^{(s)}$ values across simulations
 Store $\hat{\rho}_M = \{\hat{\rho}^{(s)}\}_{s=1}^M$

Step 5: Compare MA(1) parameter distributions
 Compute Kolmogorov-Smirnov distance:
 $D_{KS}(\rho) = \sup_x |F_{\text{empirical}}(x) - F_{\text{surrogate}}(x)|$

Results: replicating estimated MA(1) parameter distribution. Figure 13 presents the results of applying Algorithm 4 for $M = 1000$ simulations across technology-invariant ρ values ranging from 0 to 0.95 in steps of 0.05. Figure 13a shows the KS distance between the empirical distribution of estimated MA(1) parameters $\hat{\rho}_i$ from empirical data and that from surrogate data for varying technology-invariant ρ values. Using technology-invariant $\rho = 0.8$ (Figures 13a, 13b) reasonably replicates the estimated MA(1) parameter distribution, with Kolmogorov-Smirnov distance $D_{KS} = 0.08$ and p-value 0.35, making it the only tested value not significantly different from the empirical distribution. The empirical histogram shows more mass between -0.2 and 0.4 , which is not well captured by any tested ρ value.²²

²²Results remain consistent when using $\hat{L}_i$ from NLLS fitting using technology-varying $\hat{\beta}_i$, which has a minimal effect on the $\hat{L}_i$ estimate. In that case $\rho = 0.8$ has lowest KS-distance, with only $\rho = 0.8$ and $\rho = 0.85$ yielding non-significant p-values (p-value > 0.05).

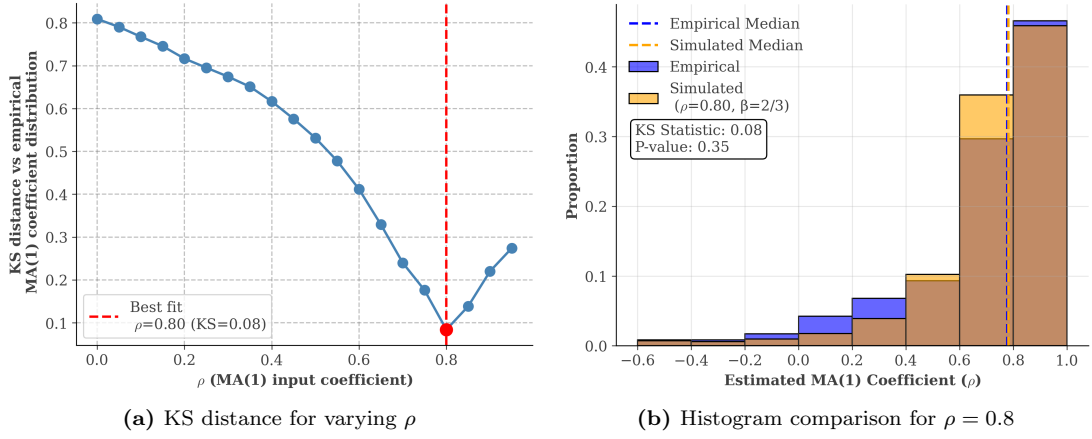


Figure 13: (a) KS distance between estimated MA(1) parameter distributions from surrogate data versus empirical data for varying technology-invariant ρ values. Lower KS distances indicate more similar distributions. (b) Histograms comparing estimated MA(1) parameter distributions from surrogate data (orange) versus empirical data (blue) for technology-invariant $\rho = 0.8$.

Impact of ρ on forecasts. Autocorrelation and stochastic shocks increase the uncertainty of forecasts, but play a relatively minor role in forecast variance compared to parameter uncertainty, as discussed in Section 4.1. The choice of ρ therefore has limited impact on the overall forecast distribution.

3.1.3 Terminal residual

The terminal residual $\epsilon_{i,T}$ is the standardised noise at the final observation time T . In the MA(1) process, the noise at time T is $\sigma_i(\rho\epsilon_{i,T-1} + \epsilon_{i,T})$, which propagates into the forecast at $T+1$ through $\sigma_i(\rho\epsilon_{i,T} + \epsilon_{i,T+1})$. One could estimate technology-specific $\epsilon_{i,T}$ values as additional parameters during Bayesian inference, but this adds a free parameter per technology and increases overfitting risk.

By construction of our noise model, $\epsilon_{i,T} \sim \mathcal{N}(0, 1)$, so $\mathbb{E}[\epsilon_{i,T}] = 0$. We therefore set $\epsilon_{i,T} = 0$ universally: for each technology, we take the realized noise at the final observation to be at its expected value.

In practice, the forecast distribution is also anchored to the last observed value via a level-shift applied at the forecast origin (Section 1.2). This is distinct from the MA(1) terminal residual and does not contradict the choice $\epsilon_{i,T} = 0$.

3.1.4 Summary and limitations

We use three technology-invariant parameters to reduce overfitting and simplify the model: (1) $\beta = 2/3$, (2) $\rho = 0.80$, and (3) $\epsilon_T = 0$. Empirical data shows substantial heterogeneity in estimated β_i and ρ_i values. However, surrogate data generated with technology-invariant parameters reasonably replicates the empirical distributions of estimated β_i and ρ_i values (Figures 11 and 13). This suggests that observed heterogeneity in parameter estimates may reflect estimation uncertainty rather than true technological heterogeneity, though the technology-invariant parameters chosen here may not fully capture all observed heterogeneity. Section 4.1 compares forecasting performance using these technology-invariant values versus technology-specific estimates.

3.2 Testing for empirical regularities

This section examines regularities in the 120 time-series dataset. First, we test whether technologies grow exponentially at early diffusion stages, as reported in [40, 49, 96, 145]. Second, we compare empirical growth-rate behavior against the empirical surrogate datasets of Section 1.4 generated with $\beta = 2/3$. Third, we compare the stochastic variation in the empirical data to the same surrogates across diffusion levels. Finally, we test three additional regularities noted in the literature.

Note on box plots. All box plots in this section (and the main text) are binned for equal time intervals or diffusion intervals. Therefore, box plots do not necessarily contain the same number of data points. We make this choice to account for the non-linear diffusion process. Binning similar time or diffusion points is more informative than binning equal numbers of points, which may span large ranges of time or diffusion levels and obfuscate trends.

Box plots follow the standard definition: the box spans the interquartile range (IQR, 25th to 75th percentile), a horizontal line within the box indicates the median, and whiskers extend to the smallest and largest values within $1.5 \times \text{IQR}$ of Q1 and Q3, unless otherwise stated. We exclude points beyond the whiskers.

3.2.1 Do technologies grow exponentially?

The B-R family of models has exponential growth at early stages. Approximate exponential growth has been observed in many cases [145], but there are also cases in which this assumption is violated [1]. Here we examine the extent to which the technologies in our database exhibit approximate exponential behavior in early stages and flag the technologies for which this seems to be a poor assumption.

Setup. We define the early diffusion stage as the first 10% of diffusion. We identify this by estimating the asymptote L_i via NLLS and keeping all observations $Y_{i,t}$ such that $Y_{i,t}/L_i \leq 0.10$. Results are similar when using $L_i := \max_t Y_{i,t}$. When fewer than 6 observations exist in the first 10%, we use the first 5 observations of the series.

For each technology, we then calculate yearly growth rates as the annualized log-differences:

$$\Delta \log Y_{i,h}(t) = \frac{\log Y_i(t) - \log Y_i(t-h)}{h}, \quad (44)$$

where h is the number of years between consecutive observations (typically $h = 1$; we use larger h only when data gaps prevent adjacent-year pairs). Under deterministic exponential growth, $\Delta \log Y_{i,t}$ is constant. To test this, we regress $\Delta \log Y_{i,t}$ on time as

$$\Delta \log Y_{i,t} = \alpha_i + \gamma_i t + \epsilon_{i,t}, \quad (45)$$

and test $H_0 : \gamma_i = 0$ using the t -statistic. We use $p < 0.01$ as the threshold for rejecting exponential growth. The sign of significant γ_i values is informative: negative γ_i would indicate the rapid-then-stabilising pattern reported by Comin et al. [1].

Results. Figure 14a shows p-values for the slope from regressing $\Delta \log Y_i(t)$ on time, as per (45), against R^2 .²³ That is, we test whether growth rates are constant (exponential growth) at early diffusion stages. Testing at the 1% significance level, 89% of the 131 technologies have slopes not significantly different from zero. This means 15 technologies have significant slopes (14 from the 120-technology set plus *Batteries* from the TWh energy data). Figure 14b shows all the estimated slopes plotted against R^2 . All but one significant slope are negative. This is in line with Comin et al. [1], who report fast initial growth that levels off to exponential growth. Figure 14c shows that the significant technologies span multiple industries, with slopes predominantly negative and near 0. Figure 14d shows the same significant technologies after excluding the first 0.5% of diffusion,

²³All analysis uses the 120-technology set. For Figure 14a, we additionally include the 11 useful-energy technologies in TWh introduced in Section 2 (131 series total).

revealing that most become non-significant.²⁴ This suggests that the initial rapid growth occurs at very low diffusion levels ($< 0.5\%$), after which growth rates stabilize.

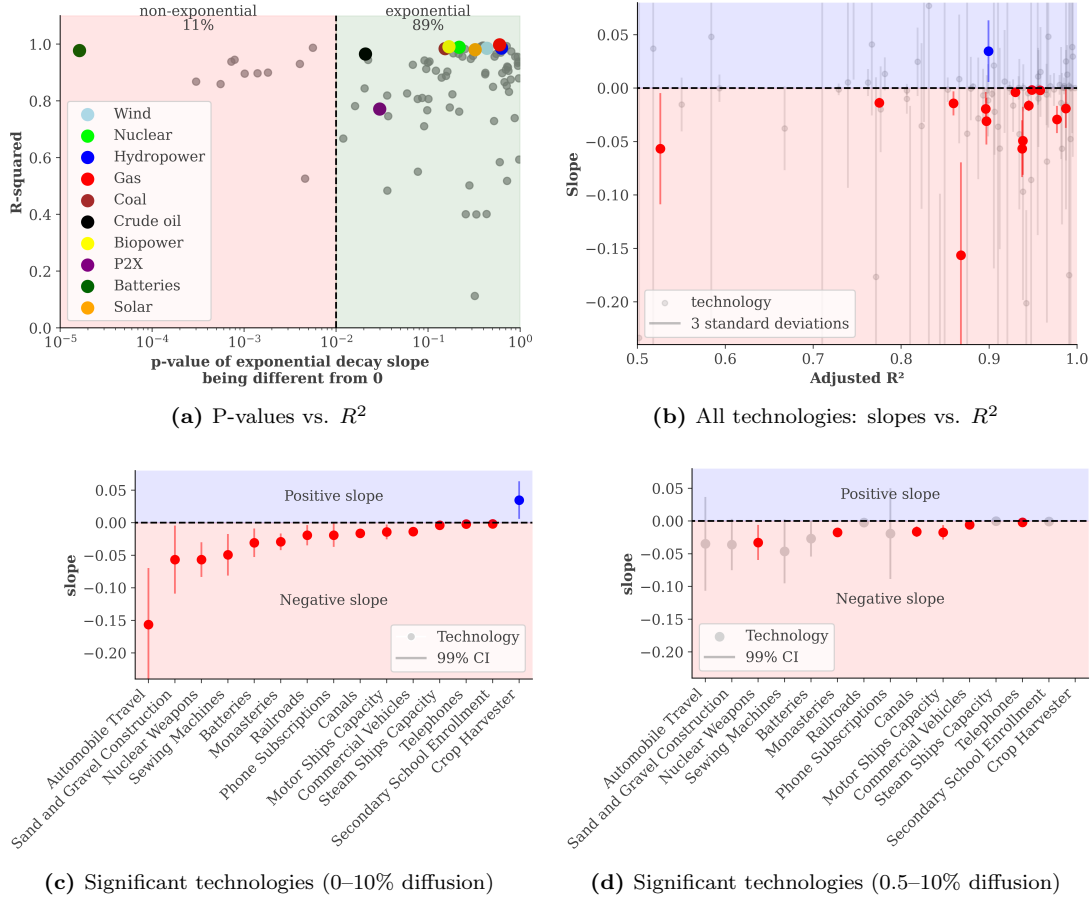


Figure 14: Testing whether technologies exhibit exponential growth at early diffusion stages (0–10%). (a) P-values vs. R^2 for 131 technologies (120-technology set plus 11 TWh energy technologies, highlighted). (b) Slopes vs. R^2 for all technologies; 89% have non-significant slopes (gray). (c) Slopes for the 15 significant technologies. (d) Same technologies after excluding first 0.5% of diffusion; most become non-significant. Those with fewer than 5 data points after exclusion are omitted.

Figure 15 further zooms in on 9 of the 15 technologies with significant slopes,²⁵ showing the annual growth rates against the adoption level.

Figure 15 shows that technologies (a–f) grow rapidly at $< 0.5\%$ diffusion and slow to exponential by $\sim 1\%$ diffusion, consistent with [1]. *Canals* and *Monasteries* (Figures 15g, 15h) show marginally significant decreasing growth throughout. *Crop harvesters* (Figure 15i) shows a different pattern: accelerating growth, then decelerating, before reaching exponential behavior.

Limitations of the model at very early diffusion stages. The B-R curve assumes exponential growth at early diffusion stages. Figure 15 shows that 10% of technologies exhibit accelerated growth at $< 1\%$ diffusion, decaying to exponential. The B-R model does not capture this behaviour, but the period is short and occurs at very low diffusion, so the impact on model fit and forecasts is minimal. Comin et al. [1] suggest the pattern may occur more broadly. Whether our 10% figure reflects reality or systematic data omission at very early stages is unclear.

²⁴Those technologies with less than 5 data points after excluding the first 0.5% are omitted.

²⁵The other 6 technologies (including *Batteries* from the TWh energy data) have slopes near 0 or behave similarly to those shown.

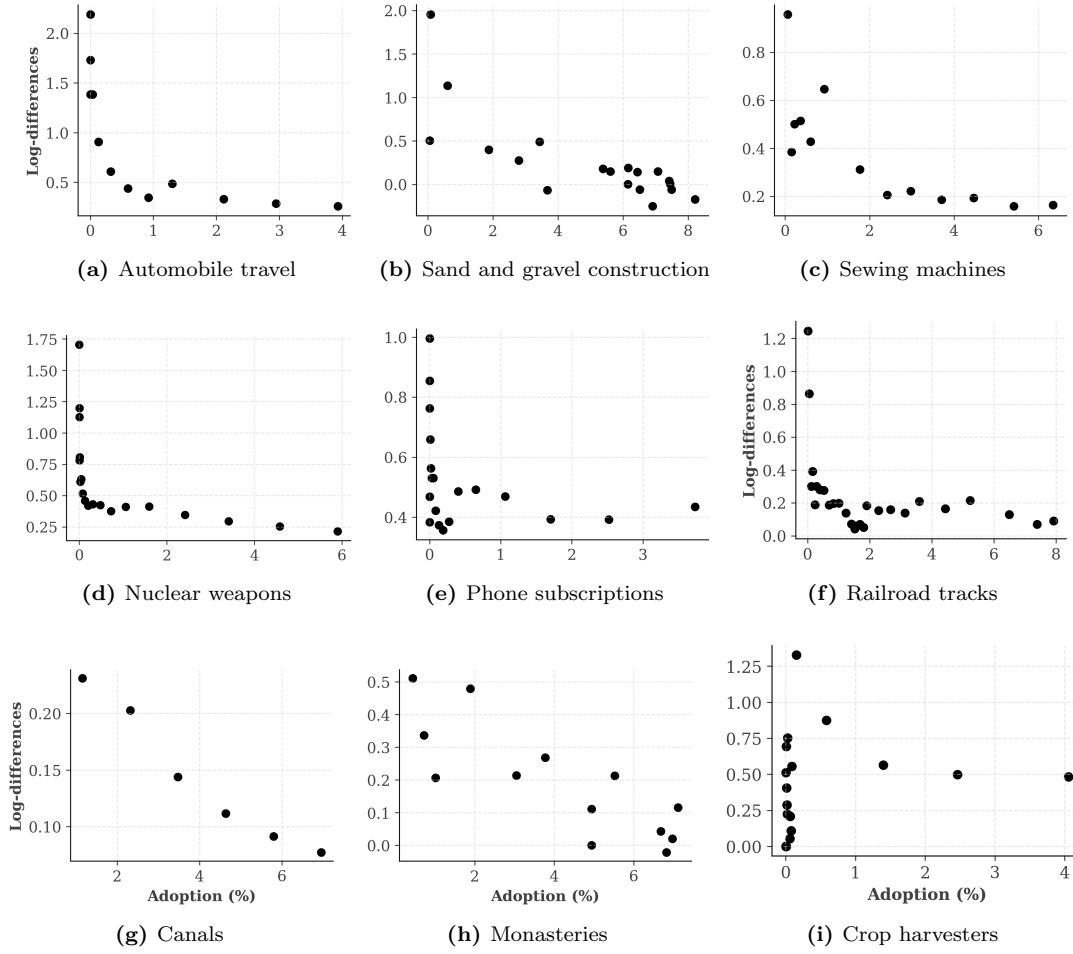


Figure 15: Log-differences for technologies with significant slopes. 9 of 15 technologies shown (14 from the 120-technology set plus *Batteries* from TWh energy data; *Batteries* not shown). (a)–(f) have negative slopes under regression. Growth is rapid at $< 0.5\%$ diffusion, then decreases toward constant growth (exponential). (g) *Canals* and (h) *Monasteries* have decreasing growth (marginally significant). (i) *Crop harvesters* has positive slope: accelerating growth at $< 0.2\%$ diffusion, decelerating from 0.5% to 1.5% , then approximately exponential growth.

3.2.2 Replicating observed growth behavior

This subsection provides the SI methodology and evidence behind the in-sample universality of growth paths shown in main paper Figure 2. Here we test whether the B-R curve with $\beta = 2/3$ approximates observed growth behavior by normalizing log-differences and time to pool technologies, and compare the empirical pool to the surrogate ensemble generated under the same DGP (Section 1.4).

Normalizing growth rates and time. At early diffusion ($t \ll t_0$), the B-R curve has $\Delta \log Y_{i,t} \approx k_i$ independently of β . We normalize log-differences by $\hat{k}_i$ (median growth rate from first 10% diffusion) and time by $\hat{k}_i(t - \hat{t}_{0,i})$:

$$\Delta \log^{(n)} Y_{i,t} = \frac{\Delta \log Y_{i,t}}{\hat{k}_i}, \quad t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$$

This collapses all technologies onto a universal curve: $\Delta \log^{(n)} Y_{i,t} \approx 1$ at early diffusion, ≈ 0 at late diffusion.

Mathematical justification. All B-R curves with the same β have identical shape in GFP-coordinates but differ in three technology-specific parameters: growth rate k_i , inflection time $t_{0,i}$, and asymptote L_i . At early diffusion ($t \ll t_0$, approximately $Y_{i,t} < 0.1L_i$), the GFP-transformed data satisfies $\log(Y_{i,t}/L_i) \approx k_i(t - t_{0,i}) + \epsilon_{i,t}$, giving $\Delta \log Y_{i,t} \approx k_i + \epsilon_{i,t+1} - \epsilon_{i,t}$. Thus:

$$\mathbb{P}_{50}[\Delta \log Y_{i,t}] \approx k_i \quad \text{for } Y_{i,t} \leq 0.1L_i$$

This provides a consistent estimator $\hat{k}_i = \mathbb{P}_{50}[\Delta \log(Y_{i,t}/L_i)]$ at early diffusion.

For time alignment, rearranging the early-diffusion approximation gives $k_i t_{0,i} = k_i t - \log(Y_{i,t}/L_i) + \epsilon_{i,t}$. Taking medians:

$$\hat{k}_i \hat{t}_{0,i} = \hat{k}_i \mathbb{P}_{50}[t] - \mathbb{P}_{50}[\log(Y_{i,t}/L_i)], \quad \text{so} \quad \hat{t}_{0,i} = \mathbb{P}_{50}[t] - \frac{\mathbb{P}_{50}[\log(Y_{i,t}/L_i)]}{\hat{k}_i}$$

Normalized time $t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$ aligns all technologies at their inflection points, removing rate and timing differences. Normalizing growth rates by $\hat{k}_i$ removes rate scaling, collapsing all curves onto a universal shape determined solely by β .

We use medians rather than OLS for robustness to outliers and deviations from exponential growth at very early diffusion. Comin et al. [1] report rapid initial growth that decays toward exponential behavior. Figure 15 confirms this pattern: $\sim 10\%$ of technologies exhibit non-exponential growth at $< 0.5\%$ diffusion before stabilizing. Median estimation is robust to these transient dy-

Algorithm 5: Replicating growth behavior with universal β

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$

Output: Comparison of empirical vs. simulated growth behavior

Step 1: Normalize empirical data:

foreach technology $i = 1$ to 120 **do**

 Calculate $\Delta \log Y_{i,t}$ using Eq. 44; define $L_i = \max_t Y_{i,t}$

 Estimate $\hat{k}_i = \mathbb{P}_{50}[\Delta \log(Y_{i,t}/L_i) : Y_{i,t} \leq 0.1L_i]$

 Estimate $\hat{t}_{0,i}$ from median: $\hat{t}_{0,i} = \mathbb{P}_{50}[t] - \mathbb{P}_{50}[\log(Y_{i,t}/L_i)]/\hat{k}_i$

 Normalize: $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t}/\hat{k}_i$, $t^{(n)} = \hat{k}_i(t - \hat{t}_{0,i})$

end

namics.

 Calculate median growth trajectory: $\mathbb{P}_{50}[\Delta \log^{(n)} Y_{i,t}]$

Step 2: Generate and normalize surrogate data:

foreach simulation $s = 1$ to $M = 1,000$ **do**

 Generate $\{Y_i^{(s)}(t)\}_{i=1}^{120}$ using DGP (Eq. 14) with $\hat{\theta}_i$, $\beta = 2/3$, $\rho = 0.8$

 Apply Step 1 normalization to obtain $\{\Delta \log^{(n)} Y_i^{(s)}(t)\}_{i=1}^{120}$

 Calculate surrogate median: $\mathbb{P}_{50}[\Delta \log^{(n)} Y_i^{(s)}(t)]$

end

Step 3: Compare distributions:

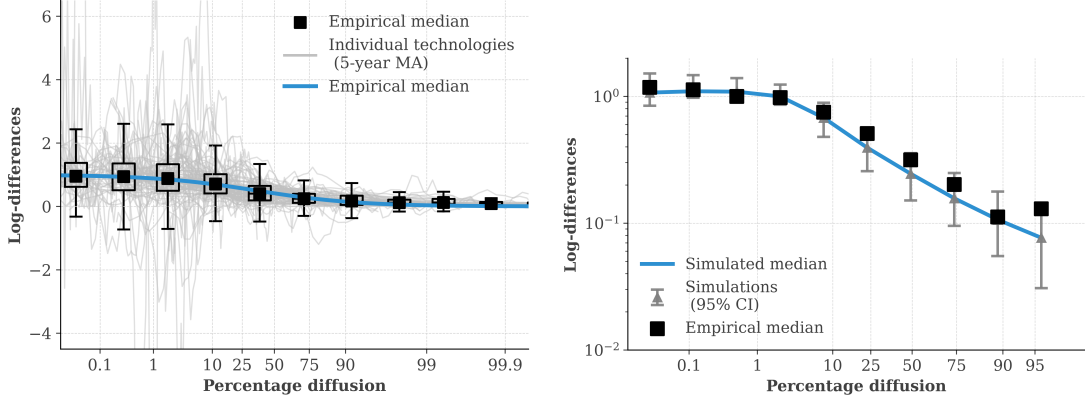
 Compare empirical median $\mathbb{P}_{50}[\Delta \log^{(n)} Y_{i,t}]$ to distribution of surrogate medians

$\{\mathbb{P}_{50}[\Delta \log^{(n)} Y_i^{(s)}(t)]\}_s$

Algorithm 5 compares median growth behavior from empirical data to surrogate data generated with $\beta = \frac{2}{3}$. The algorithm outputs the distribution of median growth trajectories from $M = 1,000$ surrogate datasets for comparison to the empirical median growth trajectory.

Results. Figure 16 shows the representative growth trajectory (median time-binned log-differences of empirical data) compared to the Bertalanffy-Richards growth-rate profile with $\beta = \frac{2}{3}$ and the distribution of median normalized log-differences from 1,000 surrogate data simulations (as per Figure 2d in the main text). We also show the 5-year moving average of log-differences for all technologies. Figure 16a shows that growth behaviour is volatile across technologies, even with 5-year moving averages. The median log-differences (blue line) closely follow the B-R curve with $\beta = \frac{2}{3}$, with small deviations at high diffusion levels. Figure 16b compares the empirical median to surrogate data, showing minimal differences.

Figure 16b shows that all empirical median log-differences lie within the 2.5 – 97.5th percentiles: any point outside would be marked as a red square, but none are present. Notably, after 10%



(a) Collapse plot of log-differences (empirical data) (b) Collapse plot of log-differences (simulated data)

Figure 16: Collapse plot of log-differences against diffusion level. (a) Collapse plot of log-differences of 120 technologies (empirical data). Grey lines are 5-year moving averages of log-differences for each technology. Box-and-whisker plot shows binned median and interquartile range across technologies. Blue line shows median. Black dashed line shows Bertalanffy-Richards curve with $\beta = \frac{2}{3}$. (b) Collapse plot of log-differences (log-scale) from 1,000 surrogate datasets generated with $\beta = \frac{2}{3}$. Vertical lines show the 2.5 – 97.5th percentiles. Blue line (and triangles) shows median across simulations. Black dots show median from empirical data (from (a)).

adoption, empirical median boxes consistently lie above the simulated median triangles. In particular, although it still lies within the error bars, the last empirical median (rightmost black square) lies in the upper tail of the surrogate distribution. This pattern is consistent with the saturation-noise misspecification of Section 1.4: the surrogate band tightens toward zero at high d while empirical log-differences retain variability, biasing their cross-technology median into the surrogate’s upper tail. Section 4.1 confirms the same pattern in forecast errors. We use the full-series estimate of $\hat{\sigma}_i$ here, as is appropriate for diagnostic replication tests (Section 1.4).

3.2.2.1 Collapse plot of technology adoption curves

We collapse technology time-series onto a single plot with $L = 1$, $k = 1$, $t_0 = 0$ using Eqs. 16, 17, and 18. We then create collapse plots for both levels of adoption, and their GFP transformation (Figures 2a and 2b in main text).

We estimate parameters as follows. For β , we assume $\beta = \frac{2}{3}$ in both cases. For the collapse of the levels, we use NLLS estimates of $\hat{L}_i$, $\hat{k}_i$, and $\hat{t}_{0,i}$. For the GFP collapse, we first estimate $\hat{L}_i$ by NLLS, then apply the GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ (Eq. 13) and regress $f(Y_{i,t}; \hat{L}_i, \beta)$ on time to estimate $\hat{k}_i$, $\hat{t}_{0,i}$, and $\hat{\sigma}_i$. We then apply normalization per Eqs. 16, 17, and 18. For the GFP collapse plot, we additionally normalize time by dividing by $\hat{\sigma}_i$. That is, the GFP-transformed data satisfies $y_{i,t} = k_i(t - t_{0,i}) + \sigma_i \eta_{i,t}$, and after setting $k = 1$ and $t_0 = 0$ (via the standard normalization), the residual scale σ_i still varies across technologies. Dividing time by $\hat{\sigma}_i$ removes this last technology-specific quantity, ensuring complete collapse.

Figure 17 shows collapse plots for GFP transformation and levels of adoption, with selected technologies highlighted. Figure 17 shows that the data collapses onto a single plot. The median of the ensemble (black line) tracks the model (red dashed line) closely, and technologies across sectors (highlighted) collapse onto similar diffusion paths.

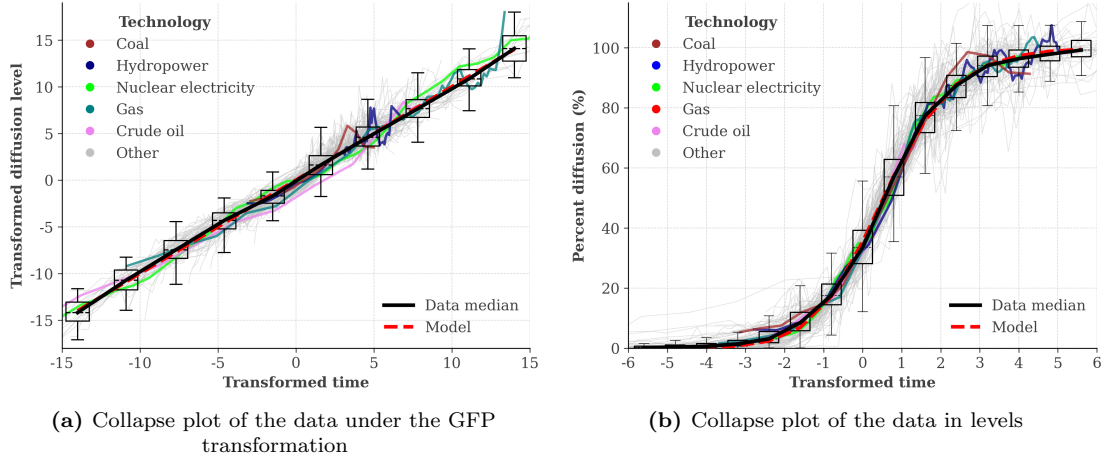


Figure 17: Collapse and binned box plots for Bertalanffy-Richards ($\beta = 2/3$) curve fits. Black line: median of ensemble. Red dashed line: model. (a) Data under the generalized Fisher-Pry transformation with selected technologies highlighted. (b) Untransformed data with energy technologies highlighted.

Test to replicate observed technology S-curves. We test whether the DGP (Eq. 14) approximates the stochastic variation observed in the data, comparing the normalized empirical ensemble to the normalized empirical surrogate ensemble (Section 1.4) via Algorithm 6.

Algorithm 6: Testing homoskedasticity and normality of noise

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$

Output: Assessment of homoskedasticity and normality of residuals

Step 1: Calculate normalized residuals from empirical data:

foreach *technology* $i = 1$ to 120 **do**

 Fit Bertalanffy-Richards curve via NLLS to obtain $\hat{L}_i$ for each technology

 Apply GFP transform $f(Y_{i,t}; \hat{L}_i, \beta)$ to estimate remaining $\hat{\theta}_i$

 Normalize data per Eqs. 16, 17, 18

 Calculate standard deviation of residuals $\text{std}(\hat{\epsilon}_i) = \sqrt{\frac{1}{T_i-1} \sum_t (\hat{\epsilon}_{i,t} - \bar{\epsilon}_i)^2}$

 Normalize data: $f_i^{(n)}(t) = f(Y_{i,t}; \hat{L}_i, \beta) / \text{std}(\hat{\epsilon}_i)$

end

Define set of normalized data: $f^{(n)}(t) = \{f_i^{(n)}(t)\}_{i=1}^{120}$

Step 2: Generate surrogate data and calculate residuals:

foreach *simulation* $s = 1$ to $M = 1,000$ **do**

foreach *technology* $i = 1$ to 120 **do**

 Generate $Y_i^{(s)}(t)$ using DGP (Eq. 14)

 with $\hat{\theta}_i$, $\beta = \frac{2}{3}$, and $\rho = 0.8$

end

 Calculate normalized data for surrogate data as in Step 1

 Define the set of normalized data: $f_s^{(n)}(t) = \{f_i^{(n)(s)}(t)\}_{i=1}^{120}$

end

Define set of simulations: $\{f_s^{(n)}(t)\}_{s=1}^M$

Step 3: Compare distributions:

 Compare ensemble normalized data (Step 1)

 to distribution of normalized ensemble from surrogate data (Step 2)

 across all normalized time units

Under the DGP, the distributions of normalised data from real and surrogate data should match across all diffusion levels. We calculate residuals as

$$\hat{\epsilon}_{i,t} = f(Y_{i,t}; \hat{L}_i, \beta) - \hat{k}_i(t - \hat{t}_{0,i}), \quad (46)$$

and normalised residuals as

$$\hat{\epsilon}_{i,t}^{(n)} = \frac{\hat{\epsilon}_{i,t}}{\text{std}(\hat{\epsilon}_i)}, \quad (47)$$

where $\text{std}(\hat{\epsilon}_i)$ is the standard deviation of residuals for technology i under the GFP transform.

Replicating ensemble behavior. Figure 18 shows box-plots of $f^{(n)}(t)$ (normalized GFP-transformed data from Step 1 of Algorithm 6) with surrogate data $\{f_s^{(n)}(t)\}_{s=1}^M$ shaded. This is Figure 2c in main text. Further, we complement this with a similar collapse using technology-specific β_i . That is, we apply Algorithm 6, but use β_i instead of $\beta = 2/3$ in step 1. We keep $\beta = 2/3$ in step 2 to generate surrogate data (i.e., estimate parameters of empirical data using $\beta = \frac{2}{3}$ for surrogate generation), since we want to test whether the DGP with a universal β can approximate the stochastic variation observed in the data. We run 1000 simulations for 120 surrogate technology time-series and plot the distribution of normalized data from surrogate data against the normalized data from empirical data. The median of the ensemble (black dots) is highly consistent with the

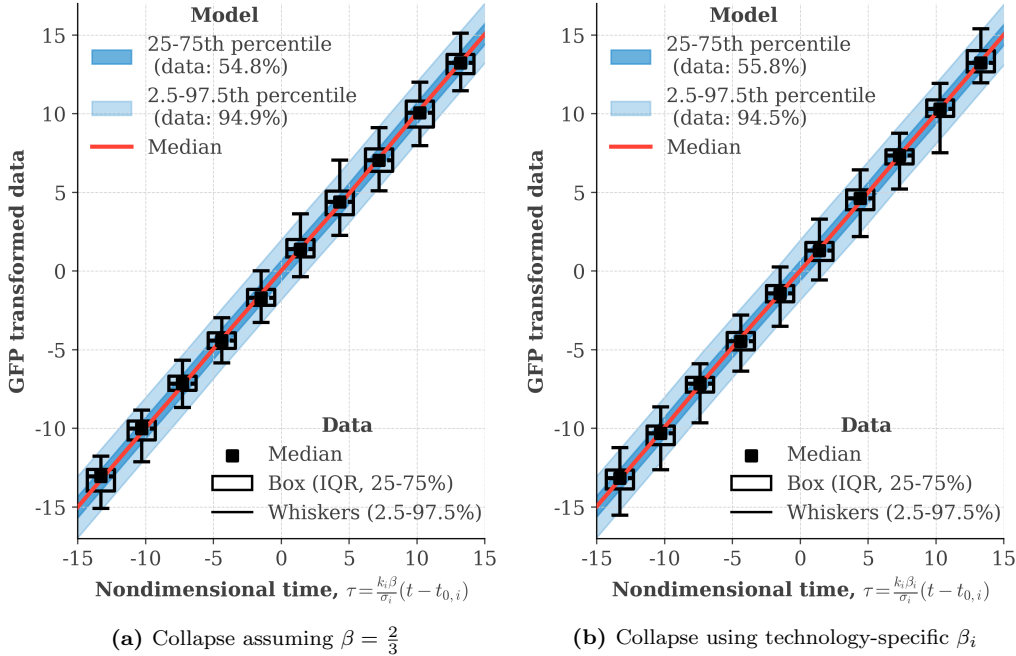


Figure 18: Collapse plot of normalized GFP-transformed data $f^{(n)}$ against normalized time, overlaid with 1,000 ensembles of 120 technology surrogate data from DGP (Algorithm 6). Box-plots: distribution of normalized data from empirical data ensemble. Whiskers are 2.5 – 97.5th percentile. Shaded regions: 25 – 75th and 2.5 – 97.5th percentile ranges of surrogate data. Proportions shown indicate empirical data within each percentile range. Black dots: median of ensemble in each box-plot. Red line: median of surrogate data. (a) Collapse assuming $\beta = \frac{2}{3}$. (b) Collapse using technology-specific β_i .

median of surrogate data (red line) in both Figure 18a and Figure 18b. In both cases, $\approx 95\%$ of the data lie within the 2.5-97.5th percentile range of surrogate data (94.9% when assuming $\beta = \frac{2}{3}$, 94.5% when using technology-specific β_i). However, the proportions of empirical data outside the 25–75th percentile range are higher than expected. This suggests that the data exhibit heavier tails than the DGP, which assumes normality after the generalized Fisher-Pry transformation. These deviations (heavier tails, leptokurtic distribution) represent limitations of the normality assumption.

The results are similar when using technology-specific β_i for collapse or assuming $\beta = \frac{2}{3}$. Note that while Figure 18b uses technology-specific β_i for the empirical collapse, we still generate surrogate data with $\beta = 2/3$ to test whether a universal β can approximate the observed stochastic variation. The similarity in collapse indicates that it can. Despite the heavy-tailed deviations from the normality assumption, the replication shown in Figure 18 indicates that the DGP is a reasonable approximation for forecasting, and we use it in subsequent sections.

3.2.3 Other regularities in technology diffusion from literature

Here we test three additional regularities reported in previous studies: (1) Are technologies diffusing faster over time [146]? (2) Are ICT technologies diffusing faster than other technologies [147]? (3) Do technologies with higher growth rates have more uncertainty?

Are technologies diffusing faster over time? Grubler [146] reports technologies diffuse faster over time. Figure 19 shows²⁶ estimated $\hat{k}_i$ against year of first observation (Figure 2f in main text) and against $\hat{t}_{0,i}$. Red squares: ICT technologies. Colored lines: linear regression lines. $\log \hat{k}_i$

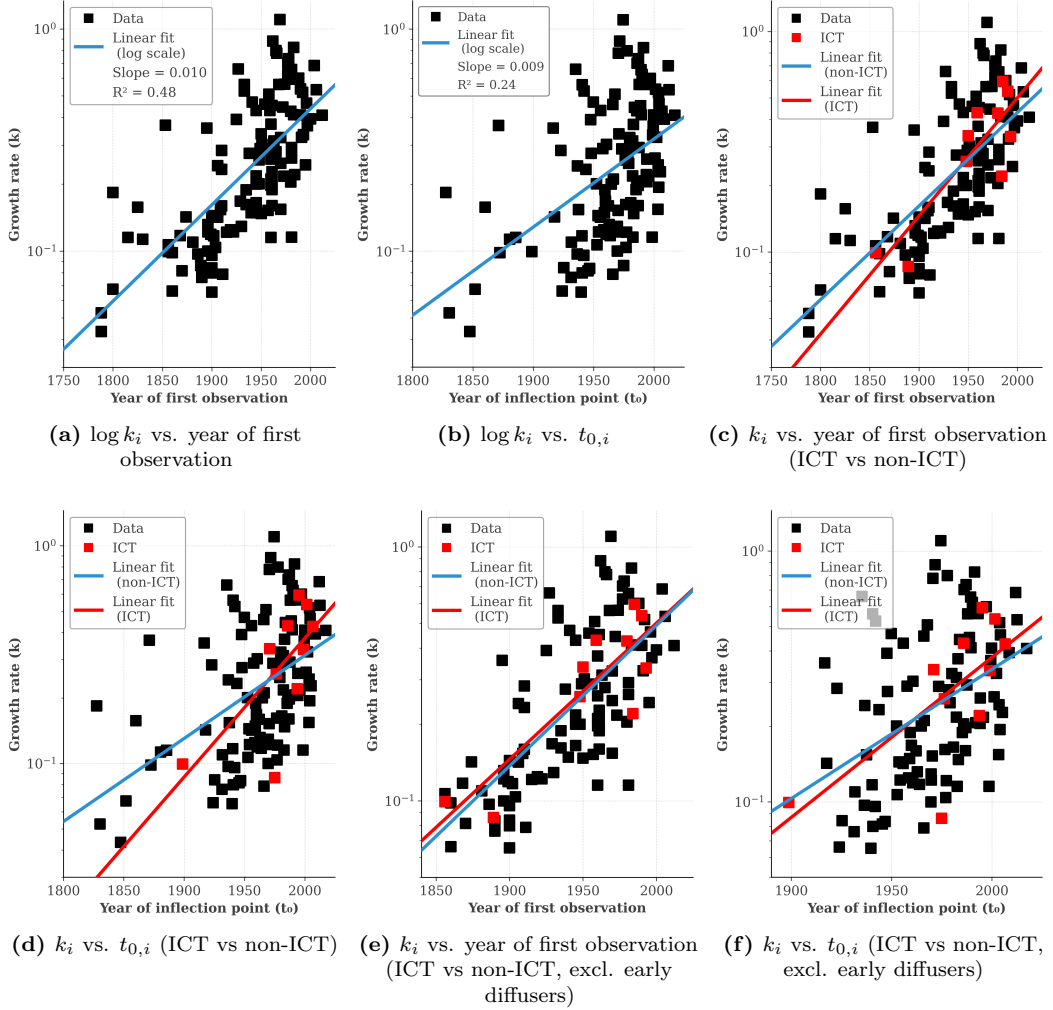


Figure 19: Estimated growth rate $\hat{k}_i$ against time (log-scale). Red dots: ICT technologies. Colored lines: regression lines. (a) $\hat{k}_i$ vs. year of first observation. (b) $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$. Blue line: linear regression line with positive slope. (c) $\hat{k}_i$ vs. year of first observation, ICT (red) vs non-ICT (blue). (d) $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$, ICT (red) vs non-ICT (blue). (e) $\hat{k}_i$ vs. year of first observation, ICT (red) vs non-ICT (blue), excluding technologies before 1900. (f) $\hat{k}_i$ vs. estimated $\hat{t}_{0,i}$, ICT (red) vs non-ICT (blue), excluding technologies before 1900.

and the year of first observation are positively correlated ($R^2 = 0.48$, slope = 0.01; Figure 19a). The correlation between $\log \hat{k}_i$ and $\hat{t}_{0,i}$ is positive ($R^2 = 0.24$, slope = 0.009; Figure 19b). Newer technologies diffuse faster in this dataset, consistent with [146]. This result may reflect selection bias: the dataset includes only technologies that diffused sufficiently. Slow-diffusing recent technologies

²⁶Monasteries (1100s) excluded.

may not yet have reached required thresholds, skewing the sample toward faster-growing technologies in recent years.

Are ICT technologies diffusing faster? Tankwa et al. [147] report that ICT growth rates exceed non-ICT when analyzing country-level data, attributing this to modularity, consistent with [148]. Figures 19c–19f compare ICT and non-ICT growth rates. Table 21 shows statistical tests. Comparisons: ICT vs. all non-ICT, and ICT vs. non-ICT in similar time-period (after first ICT technology observation). Table 21 indicates that ICT technologies have higher median growth

Metric	Non-ICT (all)	p-value (vs ICT)	ICT	p-value (vs ICT)	Non-ICT (same time-period)
Median k_i	0.21	–	0.34	–	0.22
Mann-Whitney U test	–	0.29	–	0.40	–
K-S test	–	0.21	–	0.32	–
Slope of OLS ($\log k_i$ vs. year of first observation)	0.010	–	0.012	–	0.012
Slope of OLS ($\log k_i$ vs. $t_{0,i}$)	0.009	–	0.015	–	0.009

Table 21: Statistical tests comparing growth rates of ICT vs. non-ICT technologies, generated by `make_tab_S21.py` in the replication package. Technologies first observed before 1700 (one technology, monasteries) are excluded; the same time-period subset comprises non-ICT technologies first observed in 1850 or later (the earliest ICT technology, telegraphs, appears in 1856). ICT median $k_i = 0.34$, non-ICT median $k_i = 0.21$, non-ICT (same time-period) median $k_i = 0.22$. Mann-Whitney U test p-values: 0.29 (ICT vs. all non-ICT), 0.40 (ICT vs. non-ICT same period). K-S test p-values: 0.21 (ICT vs. all non-ICT), 0.32 (ICT vs. non-ICT same period). OLS slopes ($\log k_i$ vs. year of first observation): 0.010 (non-ICT all), 0.012 (ICT), 0.012 (non-ICT same period). OLS slopes ($\log k_i$ vs. $t_{0,i}$): 0.009 (non-ICT all), 0.015 (ICT), 0.009 (non-ICT same period).

rates than non-ICT technologies (both when compared to all non-ICT and to the same time-period subset). The difference is not statistically significant (Mann-Whitney U p-values: 0.29, 0.40). The small ICT sample size ($n = 10$) may limit statistical power.

Table 21 and Figures 19c, 19d: ICT growth rates increase faster over time than non-ICT growth rates. The slope difference is smaller when comparing $\log k_i$ vs. year of first observation (0.012 vs. 0.010), but larger when comparing $\log k_i$ vs. $t_{0,i}$ (0.015 vs. 0.009).

Excluding technologies that diffused before 1900, the slope difference decreases (Figures 19e, 19f). For k_i vs. year of first observation (Figure 19e), non-ICT growth rates increase marginally faster, though not significantly. The difference observed when including all technologies may be driven by varied growth patterns among early diffusers. When controlling for time period, ICT and non-ICT growth rates may increase at similar rates.

Notably, these results are subject to dataset limitations. The dataset only includes technologies that have surpassed certain diffusion thresholds, which may bias the sample toward faster diffusing technologies, especially in time-periods closer to the present day. Thus, we should interpret conclusions regarding ICT vs. non-ICT growth rates with caution.

Do technologies with higher growth rates have more uncertainty? Farmer and Lafond [141] report that technologies with faster cost improvements have higher volatility. Here we test a similar phenomenon, but for technology diffusion. Figure 20 shows estimated $\hat{\sigma}_i$ against estimated $\hat{k}_i$.

Figure 20 shows that the correlation between $\hat{k}_i$ and $\hat{\sigma}_i$ is only slightly positive, with $R^2 = 0.06$ and a slope of 0.33. The data are noisy, suggesting the relationship may be spurious. This dataset shows a weak positive relationship between growth rate and uncertainty, weaker than the learning rate-volatility relationship reported in [141].

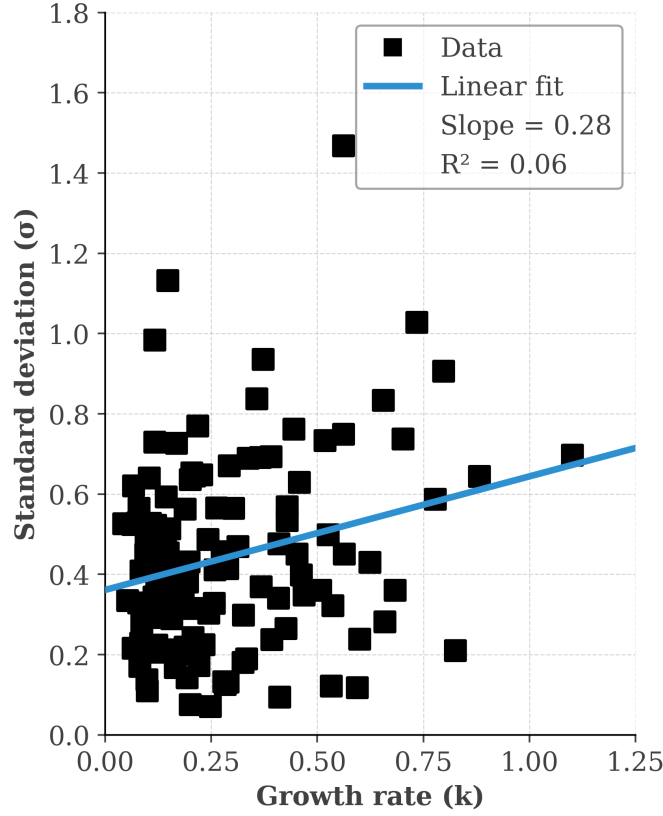


Figure 20: Estimated growth rate $\hat{k}_i$ against estimated noise parameter $\hat{\sigma}_i$. Correlation: $R^2 = 0.06$, slope = 0.33.

3.3 Comparison of S-curve models

The literature is unclear on S-curve model selection [13, 29, 30, 41]. We show that Gompertz, logistic, and Bertalanffy-Richards models produce reasonable in-sample fits ($R^2 \approx 0.93$) despite assuming different S-curve functional forms. In this section, we analyze which model best captures empirical growth dynamics. We then show that ensemble averaging across technologies with heterogeneous parameters (i.e. k_i and $t_{0,i}$ drawn from a distribution) creates apparent Gompertz-like decay even when individual series follow logistic or B-R dynamics. This helps explain why empirical studies sometimes favor Gompertz and sometimes logistic, while B-R is often not considered (Section 1.1).

We perform three analyses: (1) Compare R^2 values to confirm similar in-sample fit quality. (2) Compare the growth collapse plot and show that ensemble averaging across time-series can make the Gompertz curve appear accurate even when individual technologies exhibit logistic- or B-R-like exponential growth at early stages. (3) All three models fail to varying degrees to describe early diffusion dynamics, consistent with Comin et al. [1]. We conclude that Gompertz exhibits systematic early-stage misspecification. Ensemble averaging explains why pooled data may indicate a good fit, even though the underlying growth is actually better explained by another model. The B-R model (with $\beta = 2/3$) is the least misspecified, but the logistic model ($\beta = 1$) is also reasonable. The backtesting comparison in Section 4.1.1 provides the out-of-sample evidence underpinning the choice of B-R as the default forecasting model.

3.3.1 Theoretical differences between Gompertz and logistic curves

We compare growth behaviour under Gompertz and logistic models. Both provide reasonable fits despite prescribing fundamentally different early-stage dynamics: Gompertz assumes exponentially decaying growth, while the logistic assumes approximately constant exponential growth at early

stages.

As derived in Section 1.1, the Gompertz curve features exponentially decaying growth at all diffusion levels:

$$\frac{d \log Y}{dt} = k \exp(-k(t - t_0))$$

In contrast, the logistic curve exhibits approximately constant exponential growth at early diffusion, then exponentially decaying growth at late diffusion:

$$\begin{aligned} \frac{d \log Y}{dt} &= \frac{k}{1 + \exp(k(t - t_0))} \\ &\approx k \quad \text{for } t \ll t_0 \quad (\text{early diffusion: exponential growth}) \\ &\approx k \exp(-k(t - t_0)) \quad \text{for } t \gg t_0 \quad (\text{late diffusion: Gompertz-like decay}) \end{aligned} \quad (48)$$

Under the Gompertz model, growth rates decay exponentially throughout diffusion, implying the logarithm of log-differences (log-log-differences) should be linear in time. Under the logistic (and Bertalanffy-Richards) model, growth rates remain approximately constant at early stages, implying log-differences should be roughly constant initially before decaying later.

3.3.1.1 Goodness of fit

Figure 21 compares adjusted R^2 values for Gompertz, logistic, and B-R curves ($\hat{\beta}_i$ and universal $\beta = 2/3$) across both datasets. All four model variants — Gompertz, logistic, B-R with fixed

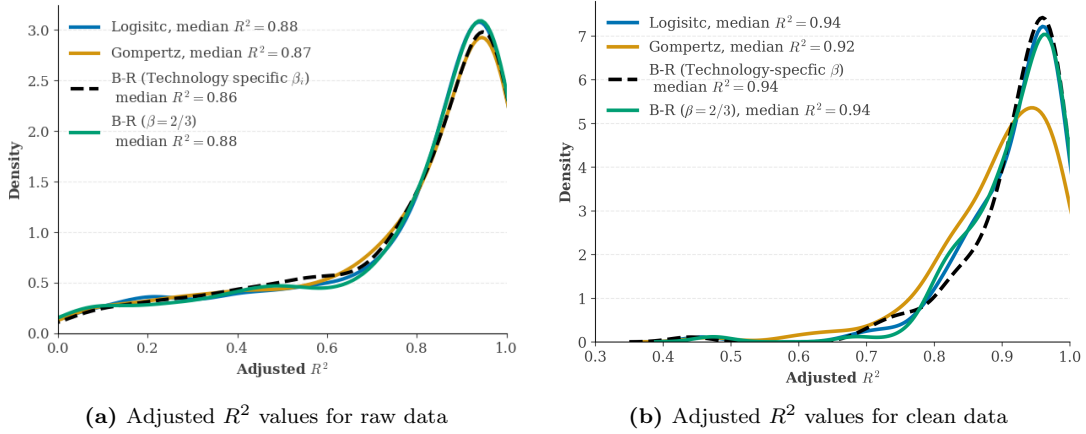


Figure 21: Comparison of R^2 values across Gompertz, logistic, and Bertalanffy-Richards models. (a) Comparison for raw data and (b) comparison for final dataset. All curves yield similar R^2 distributions with similar median R^2 values.

$\beta = 2/3$, and B-R with estimated $\hat{\beta}_i$ — provide similar ad-hoc fits on both the raw (Figure 21a) and clean data (Figure 21b), with Gompertz having the lowest median adjusted R^2 of 0.91.

3.3.2 Ensemble growth behavior under different S-curve assumptions

We compare how well the Gompertz and logistic models capture ensemble growth behavior using collapse plots of log-differences (cf. B-R collapse in Figure 16b). Algorithm 7 details Gompertz normalization; see Section 3.2 for logistic and B-R curves.

Mathematical derivation of normalization factors. For the Gompertz curve with growth rate $\frac{d \log Y}{dt} = k \exp(-k(t - t_0))$, integrating over discrete time step $h = 1$ yields:

$$\Delta \log Y_{i,t} = [1 - e^{-k_i}] \exp(-k_i(t - t_{0,i})). \quad (49)$$

Taking logarithms yields

$$\log \Delta \log Y_{i,t} = \log(1 - e^{-k_i}) - k_i(t - t_{0,i}).$$

This expression is linear in time throughout the entire diffusion curve, allowing parameter estimation on the full sample. We regress $\log \Delta \log Y_{i,t}$ on time t using OLS to obtain slope $m_i = -k_i$ and intercept $b_i = k_i t_{0,i} + \log(1 - e^{-k_i})$, giving:

$$k_i = -m_i, \quad t_{0,i} = \frac{b_i - \log(1 - e^{-k_i})}{k_i}.$$

The normalization factor $(1 - e^{-k_i})$ arises from the discrete-time integration. Setting time $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$ and then $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t} / (1 - e^{-\hat{k}_i})$ collapses all technologies onto the universal curve $\log(\Delta \log^{(n)} Y_{i,t}) = -\tilde{t}_i$.

Algorithm 7: Gompertz collapse plot construction

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$

Output: Gompertz collapse plot parameters $\{k_i, t_{0,i}\}$

Step 1: Calculate log-log-differences:

```

foreach technology  $i = 1$  to 120 do
    Calculate log-log-differences:
     $\log \Delta \log Y_{i,t} = \log \left( \frac{\log Y_i(t+h) - \log Y_{i,t}}{h} \right)$ 
end

```

Step 2: Estimate parameters via OLS:

```

foreach technology  $i$  do
    Regress  $\log \Delta \log Y_{i,t}$  on time  $t$  using OLS:  $\log \Delta \log Y_{i,t} = m_i t + b_i$ 
    Extract parameters:
     $\hat{k}_i = -m_i$ 
     $\hat{t}_{0,i} = \frac{b_i - \log(1 - e^{-\hat{k}_i})}{\hat{k}_i}$ 
end

```

Step 3: Normalize and construct collapse plot:

```

foreach technology  $i$  do
    Normalize time:  $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$ 
    Normalize log-differences:  $\Delta \log^{(n)} Y_{i,t} = \frac{\Delta \log Y_{i,t}}{1 - e^{-\hat{k}_i}}$ 
end
    Plot  $\Delta \log^{(n)} Y_{i,t}$  vs  $\tilde{t}_i$  on log scale

```

3.3.2.1 Replicating ensemble growth behavior via simulation

We construct empirical surrogate ensembles (Section 1.4) under each of the three candidate DGPs (B-R with $\beta = 2/3$, Gompertz, logistic) and compare median normalized growth trajectories to the empirical pool. The procedure follows the standard fit-then-resample protocol; the only deviation is the substitution of the candidate DGP at the generation step.

Parameter estimation procedures and normalization factors. For the B-R curve, the GFP-transform at early diffusion ($t \ll t_{0,i}$) gives:

$$-\frac{1}{\beta} \log \left[\left(\frac{L_i}{Y_{i,t}} \right)^\beta - 1 \right] \approx k_i(t - t_{0,i})$$

At early diffusion, $\frac{d \log Y}{dt} \approx k$, so:

$$\Delta \log Y_{i,t} = \log Y_{i,t+1} - \log Y_{i,t} \approx k_i$$

Normalizing by k_i gives $\Delta \log^{(n)} Y_{i,t} = \Delta \log Y_{i,t} / k_i \approx 1$ at early diffusion.

We estimate $\hat{L}_i$ via NLLS fitting of the B-R model, then regress the GFP-transformed variable on time t using OLS on the first 10% of data ($Y_{i,t} < \hat{L}_i/10$) to obtain slope m_i and intercept b_i :

$$\hat{k}_i = m_i, \quad \hat{t}_{0,i} = -b_i/m_i$$

For the **Gompertz model**, we apply OLS to the full sample, regressing $\log \Delta \log Y_{i,t}$ on time t . As derived above, this yields $\hat{k}_i = -m_i$ and $\hat{t}_{0,i} = (b_i - \log(1 - e^{-\hat{k}_i}))/\hat{k}_i$. The normalization factor $(1 - e^{-\hat{k}_i})$ arises from integrating the exponential decay. Unlike logistic and B-R curves where $\Delta \log Y \approx k$ only at early diffusion, the Gompertz curve has $\log \Delta \log Y$ linear in time, allowing full-sample estimation.

Thus, normalized log-differences are defined as

$$\Delta \log Y_{i,t}/\hat{k}_i \quad (\text{logistic and B-R}), \quad \Delta \log Y_{i,t}/(1 - e^{-\hat{k}_i}) \quad (\text{Gompertz})$$

The normalized time is $\tilde{t}_i = \hat{k}_i(t - \hat{t}_{0,i})$ for all models.

Algorithm 8: Surrogate data generation under three candidate DGPs (Gompertz, logistic, and B-R with $\beta = 2/3$) for replicating empirical ensemble growth behaviour. Per-technology parameters are estimated on the early-data window—first 10% for logistic and B-R, full sample for Gompertz (which has analytic OLS-on-log-log form). $M = 1000$ surrogate ensembles are drawn per DGP and compared to the empirical pool via the cross-technology median growth trajectory.

Input: Real technology time-series $\{Y_{i,t}\}_{i=1}^{120}$

Output: $M = 1000$ surrogate ensembles under both DGPs

Step 1: Estimate parameters from empirical data:

foreach *technology* $i = 1$ *to* 120 **do**

 Estimate $\hat{L}_i$ via NLLS

 Regress $f(\hat{L}_i)$ on time t to estimate $\hat{k}_i, \hat{t}_{0,i}$

 Estimate noise: $\sigma_i = \sigma_{\epsilon,i}/\sqrt{1 + \rho^2}$ with $\rho = 0.8$,
 where $\sigma_{\epsilon,i}$ is OLS residual standard deviation

end

Step 2: Generate surrogate data:

foreach *simulation* $s = 1$ *to* $M = 1000$ **do**

foreach *technology* $i = 1$ *to* 120 **do**

 Generate time-series under each DGP (B-R, Gompertz, logistic) with parameters
 $\{L_i, k_i, t_{0,i}, \sigma_i, \rho = 0.8\}$, using $\beta = 2/3$ for B-R and $\beta = 1$ for logistic

end

end

Step 3: Calculate ensemble median growth rates:

foreach *simulation* s **do**

 Calculate log-differences

Logistic and B-R (technology-invariant β): Estimate parameters on first 10%
 using $\hat{L}_i = \max_t Y_{i,t}$ and normalize

Gompertz: Estimate parameters on log-log-differences and normalize

 Calculate median across 120 technologies at each normalized time

end

 Calculate median of medians across $M = 1000$ simulations

Figure 22 shows log-differences collapse plots under Gompertz, logistic, and B-R ($\beta \in \{1/2, 2/3\}$) curves. It shows that the Gompertz DGP captures ensemble growth behavior when fitting the full sample, with empirical median marginally below the 2.5th percentile at high diffusion levels (red squares). The logistic DGP captures early diffusion growth when fitting the first 10% but underestimates growth in the last 5–10% (red square). Otherwise it is comparable to the B-R $\beta = 2/3$ model, but assuming a slightly faster decay in growth rate at later diffusion. The B-R curve with $\beta = 0.5$ performs comparably to $\beta = 2/3$, but slightly better captures growth at later 90–95% diffusion (the black square is further within the 95% CI of simulation). The logistic and two B-R models tend to underestimate growth rates, observed by the empirical medians (black

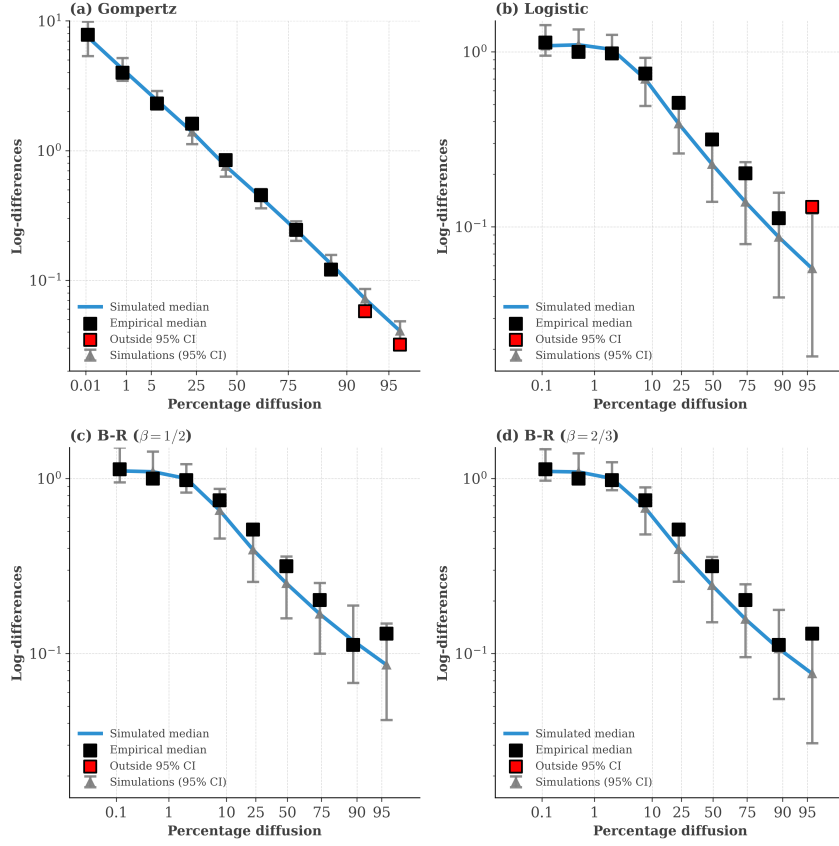


Figure 22: Collapse plots of log-differences under (a) Gompertz (parameter estimates on full sample), (b) logistic (parameter estimates on 10% of the data), (c) B-R ($\beta = 0.5$, parameter estimates on 10% of the data), and (d) B-R ($\beta = 2/3$, parameter estimates on 10% of the data). Black squares: median observed growth rates. Blue lines (and gray triangles): simulation ensemble medians. Gray whiskers: 2.5 – 97.5th percentiles. Red squares: empirical median outside simulated 2.5 – 97.5th percentiles.

squares) often being above the surrogate medians (blue lines), from 15% diffusion onward. The same direction of bias appears across these three GFP-additive-Gaussian-noise DGPs despite their different functional forms, which points to the saturation-noise misspecification of Section 1.4 (shared across them) rather than to functional form; the Gompertz model shows the opposite-direction deviation, reflecting its exponentially decaying growth law rather than the noise specification. As in Section 3.2, $\hat{\sigma}_i$ is estimated on the full series for this diagnostic.

These results illustrate the difficulty of identifying the true DGP from ensemble growth behaviour alone. In-sample fit quality (adjusted R^2) is high for all models, indicating that each captures the overall growth pattern. The Gompertz DGP fits the full sample but slightly overestimates the median growth at high diffusion; the logistic and B-R DGPs fit the first 10% well but tend to underestimate growth thereafter. Ensemble growth behaviour alone therefore does not conclusively identify the true DGP.

3.3.3 Ensemble averaging effects

The conclusion that Gompertz captures ensemble growth may be misleading. Heterogeneous growth rates k_i [5, 6, 149], combined with pooling, could create apparent exponential decay even when individual technologies grow exponentially, which is called an *ensemble averaging artifact*.

Heterogeneous growth rates. We test whether k_i follows a (shifted) exponential distribution (Algorithm 9). A shifted exponential is defined as $k_i = \xi + \text{Exp}(s)$, where ξ is the location shift and s is the rate parameter.

Algorithm 9: Testing exponential distribution of growth rates

Input: Technology time-series $\{Y_{i,t}\}_{i=1}^{120}$ **Output:** Test if growth rates follow exponential distribution**B-R curve parameter estimates****foreach** *technology* $i = 1$ **to** 120 **do** Fit Bertalanffy-Richards curve via NLLS with $\beta = \frac{2}{3}$ to obtain $\hat{L}_i$ Apply GFP transformation to $Y_{i,t}$ to get $y_i(t) = f(Y_{i,t}; \hat{L}_i, \beta)$ Estimate $\hat{k}_i$ by regressing $y_i(t)$ on time t using OLS**end**Fit shifted exponential distribution to $\{\hat{k}_i\}_{i=1}^{120}$

Draw 10,000 samples from fitted exponential distribution

Compare empirical distribution to ensemble distributions

Figure 23 compares shifted versus non-shifted exponential distributions fitted to empirical growth rates $\{\hat{k}_i\}_{i=1}^{120}$. The left panel shows a shifted exponential distribution, $k_i = \xi + \text{Exp}(s)$ (as

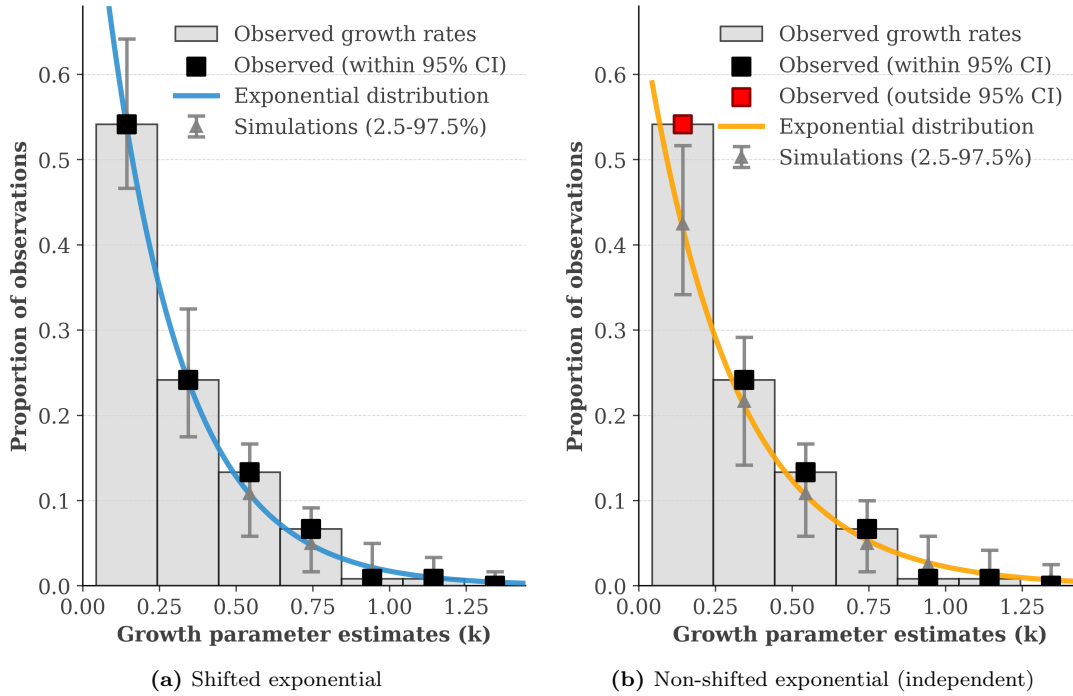


Figure 23: Comparison of (a) the shifted exponential distribution, $k_i = \xi + \text{Exp}(s)$, and (b) the non-shifted exponential distribution, $k_i = \text{Exp}(s')$, both fitted to empirical growth rates. Black squares: observed proportions. Red squares: observed proportions outside the surrogate distribution. Solid curves: theoretical distributions. Triangles with error bars: median and 95% confidence intervals from 10,000 Monte Carlo simulations.

per Figure 2e in main text), where ξ is the location shift and s is the rate parameter, fitted via maximum likelihood. The shifted exponential is theoretically justified because S-curves cannot have zero growth rates (i.e., $k_i > 0$ by construction). The location shift ξ captures the minimum growth rate. The right panel shows a non-shifted exponential distribution, $k_i = \text{Exp}(s')$, with independently fitted rate parameter s' . The shifted exponential shows close visual agreement with the empirical data, with all observed proportions falling within the 95% confidence intervals from 10,000 Monte Carlo simulations. The non-shifted exponential underestimates the lowest growth-rate bin, whose observed proportion falls outside the interval (red square).

Table 22 presents comprehensive statistical tests comparing three exponential models. The shifted exponential generally outperforms both non-shifted alternatives across most statistical tests (KS test, log-likelihood, AIC, BIC). The sole exception is the chi-square test, where the non-shifted

Table 22: Comparison of three exponential models for the growth-rate distribution. Figure 23 plots the shifted exponential and the non-shifted (independent) models; the non-shifted (same rate) is a nested model included here only for the statistical comparison.

Metric	Shifted exponential	Non-shifted (same rate)	Non-shifted (independent)
Location shift c	0.04	0.00	0.00
Rate λ	4.02	4.02	3.42
<i>Goodness-of-fit tests</i>			
KS statistic	0.09 [†]	0.22***	0.19***
KS p -value	0.269	0.000	0.000
χ^2 statistic	43.71***	8.80 [†]	16.29**
χ^2 p -value	0.000	0.117	0.006
<i>Model comparison criteria</i>			
Log-likelihood	47.00 [†]	26.06	27.69
AIC	-89.99 [†]	-50.11	-53.39
BIC	-84.42 [†]	-47.32	-50.60

Note: [†] indicates best value. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Lower KS/chi-square statistics and lower AIC/BIC indicate better fit. Shifted exponential: $k_i = c + \text{Exp}(\lambda)$ ($c = 0.04$ matches the observed minimum growth rate $\min_i \hat{k}_i = 0.043$). Non-shifted (same rate): nested model with $c = 0$, same λ as shifted. Non-shifted (independent): $c = 0$, λ fitted independently.

model using the rate parameter from the shifted fit (nested model) exhibits the lowest chi-square statistic ($p = 0.12$). Note that S-curve time-series must have $k_i > 0$ by construction. The observed minimum growth rate $\min_i \hat{k}_i = 0.04$ confirms this, and closely matches the fitted location shift $\xi = 0.04$. We adopt the shifted exponential distribution $k_i = \xi + \text{Exp}(s)$ for subsequent analysis.

Ensemble averaging artifact. We use the approximate distribution of the empirical growth rates (shifted exponential) to demonstrate how ensemble averaging across technologies with heterogeneous growth parameters can create apparent Gompertz-like decay patterns even when individual technologies follow logistic or Bertalanffy-Richards dynamics with early-stage exponential growth. That is, pooling the log of log-differences across technologies with heterogeneous k_i and $t_{0,i}$ can distort ensemble growth behavior.

Mechanism. When pooling log-log-differences at common calendar times t , two factors create systematic bias:

- **Technologies are at different diffusion stages:** Pooling poses the risk of mixing different diffusion stages, which can bias the ensemble average.
- **Measurement noise adds bias (Jensen’s inequality):** Taking logarithms of noisy growth rates systematically underestimates the true average because the logarithm is a concave function—technologies with high noise relative to their growth rate contribute disproportionately to downward bias [150].

Pooling $\log[\Delta \log Y]$ across technologies at common time coordinates creates systematic bias in the ensemble average.

Simulation demonstration. We simulate an ensemble where the true DGP is B-R ($\beta = 2/3$, $\rho = 0.8$), then demonstrate that pooling produces apparent Gompertz-like decay (Algorithm 10).

Algorithm 10: Demonstrating ensemble averaging artifact. The noise parameter $\sigma = 0.4$ is technology-invariant in this simulation.

Input: simulation parameters

Output: Demonstration of ensemble averaging artifact

Setup: $N = 120$, $k_i \sim 0.02 + \text{Exp}(2)$, $t_{0,i} \sim \Gamma(2, 10, 10)$, $L_i = 100$, $\sigma = 0.4$, $t \in [0, 50]^a$

Step 1: Generate time-series: foreach technology $i = 1$ to N do

$$Y_{i,t} = \frac{L_i}{[1 + \exp(-\beta k_i(t - t_{0,i}) + \beta \sigma(\epsilon_t + \rho \epsilon_{t-1}))]^{1/\beta}}, \beta = 2/3, \rho = 0.8$$

end

Step 2: Calculate transformations:

Calculate collapse plot transformations per algorithm 7

^aOther parameter choices yield similar results, including heterogeneous σ_i drawn from a Gamma distribution.

Results. Figure 24 shows Gompertz-curve growth collapse plots (Algorithm 7) for empirical data and 120 simulated B-R time-series (Algorithm 10). Pooling across B-R processes with heterogeneous k_i and $t_{0,i}$ creates apparent exponential decay, mimicking Gompertz behavior. Figure 24a shows

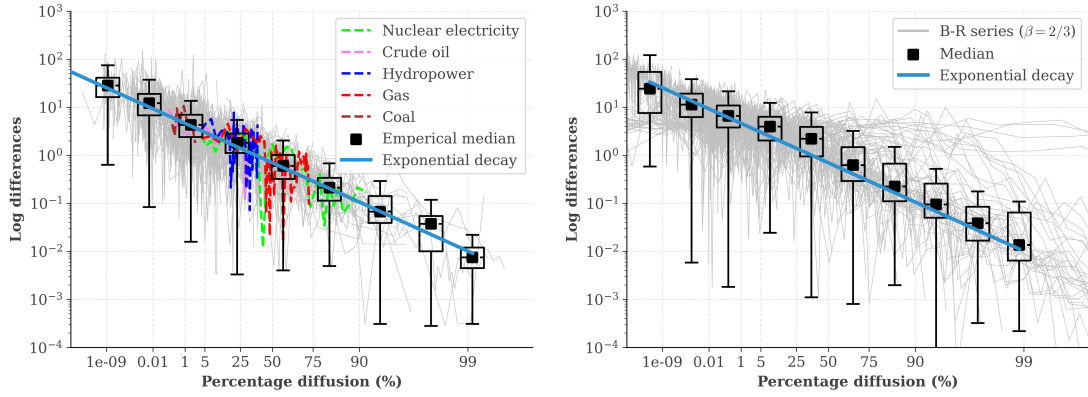


Figure 24: Effects of ensemble averaging assuming Gompertz growth when individual time-series follow logistic or Bertalanffy-Richards curves. Individual time-series: gray, black squares: median, blue line: Exponential decay. (a) Collapse plot of log-differences under Gompertz normalization. We highlight energy technologies. Individual technology clustering at specific diffusion stages indicates potential ensemble averaging bias. (b) Collapse plot across 120 Bertalanffy-Richards series ($\beta = 2/3$) simulated via Algorithm 10. Pooling creates apparent exponential decay, mimicking Gompertz-like behavior.

the collapse plot of log-differences for Gompertz. While ensemble behaviour appears Gompertz-like, individual technologies cluster at specific time and diffusion stages, indicating potential ensemble-averaging bias.

Figure 24b shows the same averaging effect using B-R ($\beta = 2/3$) collapsed via Algorithm 7. The ensemble median (black dots) lies slightly above the true Gompertz curve (green line), unlike empirical data (Figure 24a). Some individual time-series (gray lines) exhibit exponential growth (horizontal lines at low diffusion). Ensemble averaging across heterogeneous B-R processes creates Gompertz-like decay patterns. The artifact is robust to changes in simulation size and parameters: it persists for different β values (e.g. $\beta = 1$), different ρ , and heterogeneous σ_i drawn from a Gamma distribution rather than a fixed σ , as long as noise is sufficiently large relative to growth rates. We do not attempt to delineate the exact parameter region where the artifact occurs.

3.3.4 Comparing Gompertz, logistic, and B-R collapse plots

We test for model misspecification using FP-transform collapse plots. Under correct specification, parameters estimated on post-5% data should predict pre-5% trajectories. Comin et al. [1] show S-curves fail to capture very early growth. We omit the first 5% of diffusion, estimate parameters

on the remainder, then assess whether omitted observations align with predictions. Systematic deviations indicate fundamental misspecification.

Defining early diffusion threshold. The 5% threshold is defined analytically. For B-R with parameters $\{L_i, k_i, t_{0,i}, \beta_i\}$, solve $Y_{i,t} = 0.05L_i$:

$$t_{5\%,i} = t_{0,i} - \frac{1}{\beta_i k_i} \log \left[\left(\frac{1}{0.05} \right)^{\beta_i} - 1 \right] \quad (50)$$

Similarly, for Gompertz and logistic curves, we solve $Y_{i,t} = 0.05L_i$ to find $t_{5\%,i}$.

Results and comparison. Figure 25 presents FP-transform collapse plots for all three models. Figure 25 shows that all models exhibit early-stage deviations, revealing misspecification. Gompertz

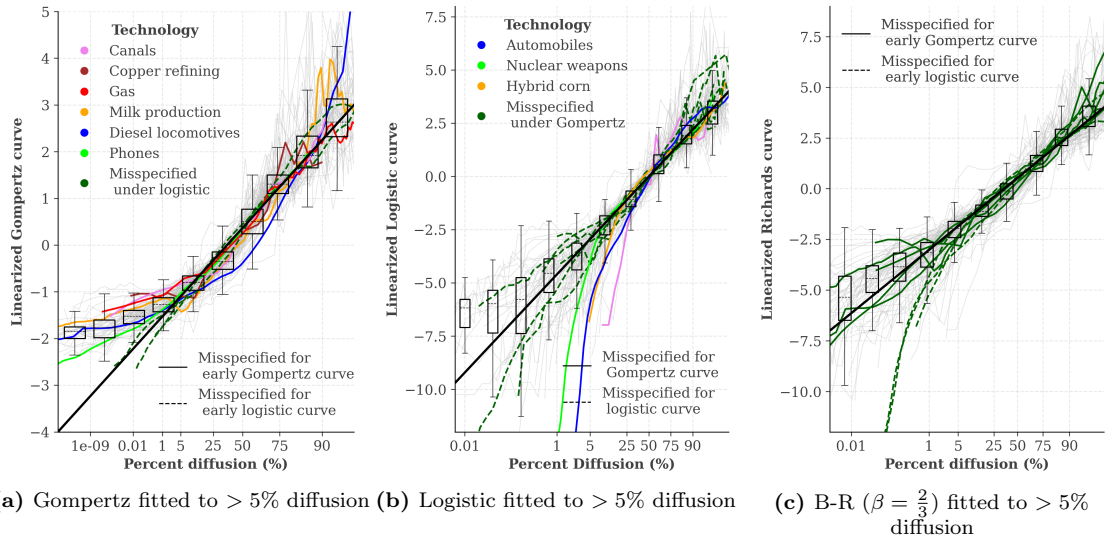


Figure 25: (G)FP-transform collapse plots comparing Gompertz, logistic, and B-R ($\beta = 2/3$) curves fitted to > 5% diffusion (i.e., omitting early < 5% diffusion). **(a)** Gompertz curve. We highlight technologies that deviate from the model description. Green dashed lines show technologies fitted better by logistic than by Gompertz. **(b)** Logistic curve. We highlight technologies that deviate from the model description. Green dashed lines show technologies fitted better by Gompertz than by logistic. **(c)** B-R curve ($\beta = 2/3$) omitting first 5% diffusion. Green solid lines fit poorly under Gompertz; green dashed lines fit poorly under logistic.

shows the most systematic deviations (Figure 25a), with many technologies consistently deviating from predictions. Logistic (Figure 25b) shows fewer deviations; specific technologies (*automobiles*, *nuclear weapons*) deviate before 5–10% diffusion. Technologies deviating under Gompertz often fit better under logistic (e.g., *steam turbines*, *air conditioners*), and vice versa. B-R (Figure 25c) exhibits the least systematic deviations.²⁷

Limitations. All models provide reasonable ad-hoc fits (high R^2), but this test ranks them: B-R first, logistic second, Gompertz third. No single model captures early adoption (< 5% diffusion) for all technologies. Technologies that fit poorly under Gompertz often fit well under logistic and vice versa, suggesting technology-specific idiosyncrasies that none of the models investigated here fully capture. This quasi-out-of-sample test (estimating on post-5% data, testing on pre-5% data) reveals fundamental heterogeneity in early diffusion dynamics across technologies. This represents a limitation of assuming universal functional forms. However, our backtesting analysis (Section 4.1) demonstrates that the B-R model with $\beta = 2/3$ produces robust out-of-sample forecasts:

²⁷*Automobiles* and *nuclear weapons* deviate before $\approx 1\%$ diffusion.

despite early-stage heterogeneity, the model captures sufficient structure for medium- to long-term prediction.

4

Out-of-sample data analysis

4.1 Testing forecast performance on empirical data

This section gives the detail behind the cross-technology backtest of main paper Figure 3, in which we train the model up to a specified diffusion level and forecast the rest of the adoption trajectory. We first set out the forecasting setup and the accuracy and calibration metrics (Section 4.1.1), then report the backtest results and explain the shape of the error curve. We also benchmark against the Bass model (Section 4.1.3). We calibrate the widening transform that corrects the prediction intervals (Section 4.1.4), and check robustness to the subjective data-cleaning step (Section 4.1.5), to the prior, and to the technology-invariant β and ρ (Section 4.1.6). We close by bracketing the prior between an uninformed and a diffusion-aware specification, to bound the achievable forecast quality (Section 4.1.7).

Throughout, we use the *shifted* forecast variant (Section 1.2.4). This variant adjusts the fitted model so the residual is zero at the last observation. Results are robust to the choice of shifted versus non-shifted forecast. Surrogate medians, confidence bands, and PIT distributions in this section come from the empirical surrogate datasets of Section 1.4, using the truncated $\hat{\sigma}_i$ described there.

4.1.1 Methods underlying main paper Figure 3

4.1.1.1 Priors

All models use technology-specific parameters $\theta_i = \{L_i, k_i, t_{0,i}, \sigma_i\}$. We use the universal prior defined in Eq. (32) (Section 1.2.4), shown in Figure 26. The same hyperparameters apply to every technology. Only the t_0 location is anchored to each series, through the preliminary growth-rate estimate $\hat{k}_{\text{ssr},i}$. This section explains how we obtain that estimate, justifies the hyperparameter choices, and reports robustness checks.

Growth-rate estimate. We obtain $\hat{k}_{\text{ssr},i}$ by minimising the sum of squared residuals of the GFP transform (Eq. 13) over L_i :

$$\hat{L}_{\text{ssr}} = \arg \min_{L_i} \sum_t [f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i})]^2, \quad L_i \in [y_{i,\max}, 10,000].$$

The OLS slope at $L_i = \hat{L}_{\text{ssr}}$ gives $\hat{k}_{\text{ssr},i}$. Conditional on L_i , this slope is approximately unbiased, so the estimate is a fast and reliable anchor for the t_0 location. If the fit returns a *NaN* value we use a fallback $\hat{k}_{\text{ssr}} = 0.15$.

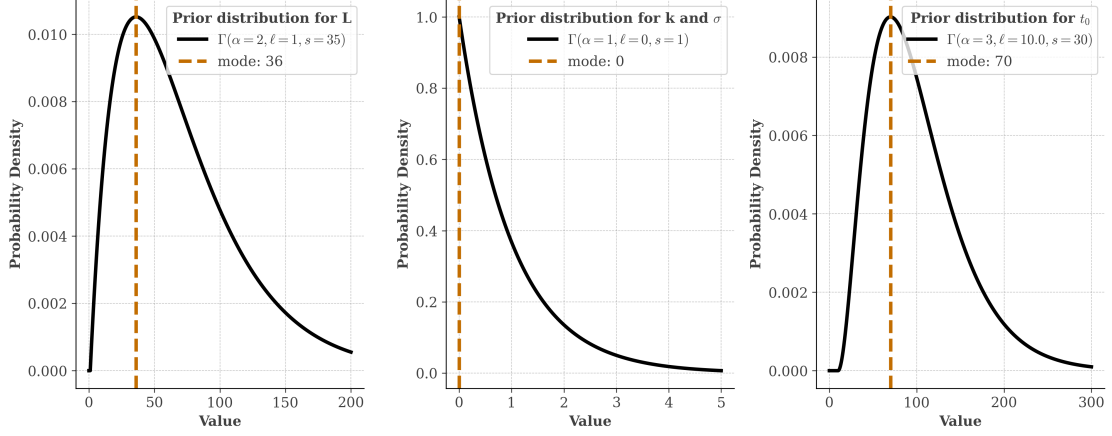


Figure 26: Priors used in the main text (universal prior). The t_0 prior is shown for a technology with $\hat{k}_{\text{ssr},i} = 0.3$ and $T_{\text{max}} = 20$, so that $\ell_{t_0} = \max(5, 20 - 3/0.3) = 10$.

Design rationale. *L prior* ($\alpha_L = 2$, $\ell_L = 1$, $s_L = 35$). The lower bound $\ell_L = 1$ forces $\tilde{L}_i > 1$, so the technology has not yet saturated. The shape $\alpha_L = 2$ places the mode at 36 times current adoption. This encodes that, for a technology observed early, the saturation level usually lies well above current adoption. Because $\alpha_L = 2$ rather than 1, the prior peaks at this mode rather than at its lower bound, so it does not assume the technology is already near saturation. The scale $s_L = 35$ keeps the prior wide but retains mass at low values, so it still admits technologies that saturate early. Forecasts stay well-calibrated at a 50% origin, where the true asymptote is twice the current value and far below the mode (Section 4.1.6).

t₀ prior ($\alpha_{t_0} = 3$, $\ell_{t_0} = \max(5, T_{\text{max}} - 3/\hat{k}_{\text{ssr},i})$, $s_{t_0} = 30$). The location anchors the inflection to the data. A fixed small location would be unreasonable for a long series, placing most mass far before the observed window. The offset $3/\hat{k}_{\text{ssr},i}$ scales with the inverse growth rate, so faster-growing technologies expect the inflection closer to T_{max} . The floor at 5 keeps the location positive. The shape $\alpha_{t_0} = 3$ places the mode 60 steps ahead of the location, encoding that at early diffusion the inflection lies well ahead. The scale $s_{t_0} = 30$ keeps the prior wide.

k and σ priors ($\alpha = 1$, $\ell = 0$, $s = 1$). Both are unit exponentials (Section 1.2.4). Unlike L and t_0 , they are well identified even at early diffusion, so we do not anchor them to the data.

The numerical choices ($s_L = 35$, $s_{t_0} = 30$, and the offset $3/\hat{k}$) are validated by backtesting on the full cross-section (Section 4.1.6). The prior is well-calibrated across a diverse set of technologies. We therefore take it as a reasonable default for technologies the model has not seen, such as solar PV and wind.

When the t_0 location is valid. The location ℓ_{t_0} is valid when it lies at or below the true t_0 . We derive the diffusion level at which this holds. In the deterministic B-R model (3) with $\beta = 2/3$, the noise-free generalised Fisher–Pry transform (Eq. 13) is

$$\frac{1}{\beta} \log(d^{-\beta} - 1) = -k(t - t_0),$$

where $d = Y/L$ is the diffusion level. The time from t_0 to diffusion d is

$$T_{\text{max}} - t_0 = -\frac{1}{\beta k} \log(d^{-\beta} - 1).$$

The condition $\ell_{t_0} \leq t_0$ (with $\hat{k} \approx k$) requires $T_{\text{max}} - t_0 \leq c/k$ for offset c . Both sides scale with $1/k$, so the growth rate cancels and the threshold depends only on c and β :

$$d \leq (1 + e^{-\beta c})^{-1/\beta}.$$

For $\beta = 2/3$ and our default $c = 3$, this gives $d \leq (1 + e^{-2})^{-3/2} \approx 83\%$. So ℓ_{t_0} lies at or below the true t_0 whenever observed diffusion is below about 83%. Above that level the data is highly

informative about t_0 , and in-sample diagnostics readily reveal the inflection, so the prior can be adjusted by hand. We keep $c = 3$. Results are robust across $c \in \{2, 3, 4\}$, which shift the threshold to about 70%, 83%, and 90% (Section 4.1.6).

Empirical Bayes. The universal prior is an empirical Bayes prior. Two of its hyperparameters depend on the series. First, the t_0 location uses the preliminary estimate $\hat{k}_{\text{ssr},i}$. Second, the normalisation by $y_{T_{\max}}$ sets the scale of the L prior, so the saturation level is a multiple of current adoption. After this the sampler treats the prior as fixed. The prior does not depend on the forecast origin, so the forecasts are out-of-sample. However, the functional form and the technology-invariant parameters $\beta = 2/3$ and $\rho = 0.8$ were calibrated on the full cross-section (Section 3.1), so the backtests are not fully independent of the test data.

4.1.1.2 Logistic and Gompertz models

Here we provide the data-generating processes and likelihoods for the two competing S-curve models. Both use the same priors as B-R, with technology-invariant $\rho = 0.8$ and $\epsilon_{i,T} = 0$.

The logistic data-generating process is

$$Y_{i,t}^{(l)} = \frac{L_i}{1 + \exp(-k_i(t - t_{0,i}) + \sigma_i(\rho \epsilon_{i,t-1} + \epsilon_{i,t}))}. \quad (51)$$

Under the Fisher–Pry (FP) transformation, adding $k_i(t - t_{0,i})$ to both sides yields

$$x_{i,t}^{(l)} = \log\left(\frac{L_i - Y_{i,t}}{Y_{i,t}}\right) + k_i(t - t_{0,i}) = \rho \sigma_i \epsilon_{i,t-1} + \sigma_i \epsilon_{i,t}, \quad (52)$$

with $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$. The Jacobian is diagonal with log-determinant

$$\log \det\left(\frac{\partial f}{\partial Y_i}\right) = \sum_{t=1}^T \left[\log L_i - \log [Y_{i,t}(L_i - Y_{i,t})] \right].$$

The τ -step forecast for posterior draw m is

$$\mathbf{Y}_{m,i}^{(l)}(\tau) = \frac{L_{m,i}}{1 + \exp(-k_{m,i}(\tau - t_{0,m,i}) + \epsilon_{T+1:T+\tau,m}^{(l)})}.$$

The Gompertz data-generating process is

$$Y_{i,t}^{(g)} = L_i \exp(-\exp(-k_i(t - t_{0,i}) + \sigma_i(\rho \epsilon_{i,t-1} + \epsilon_{i,t}))). \quad (53)$$

Adding $k_i(t - t_{0,i})$ to the GFP-transformed observations yields

$$x_{i,t}^{(g)} = \log \log\left(\frac{L_i}{Y_{i,t}}\right) + k_i(t - t_{0,i}) = \rho \sigma_i \epsilon_{i,t-1} + \sigma_i \epsilon_{i,t}, \quad (54)$$

with $\epsilon_{i,t} \sim \mathcal{N}(0, 1)$. The Jacobian log-determinant is

$$\log \det\left(\frac{\partial f}{\partial Y_i}\right) = - \sum_{t=1}^T \log \left[Y_{i,t} \left(\log L_i - \log Y_{i,t} \right) \right].$$

The τ -step forecast for posterior draw m is

$$\mathbf{Y}_{m,i}^{(g)}(\tau) = L_{m,i} \exp\left(-\exp\left(-k_{m,i}(\tau - t_{0,m,i}) + \epsilon_{T+1:T+\tau,m}^{(g)}\right)\right).$$

4.1.1.3 Measuring point forecast errors

Let $\{Y_{i,t}\}_{t=1,\dots,T_i}$ be technology i 's observed time-series. We evaluate forecast accuracy at every diffusion target $d^* \in \{1\%, 2\%, \dots\}$ as follows. The fractional year $t_i^*(d^*)$ at which the observed series crosses $d^* \cdot L_i$ is obtained by linear interpolation between the first observation below and the first observation above this level,

$$t_i^*(d^*) = t_k + \frac{d^* \cdot L_i - Y_{i,t_k}}{Y_{i,t_{k+1}} - Y_{i,t_k}} (t_{k+1} - t_k), \quad Y_{i,t_k} \leq d^* \cdot L_i \leq Y_{i,t_{k+1}}. \quad (55)$$

The observed value at t_i^* equals $d^* \cdot L_i$ by construction. The predicted value $\hat{Y}_i(t_i^*)$ is obtained as follows. For each posterior MCMC sample $(L_s, k_s, t_{0,s})$ we evaluate the deterministic shifted S-curve (Section 1.2.4) at $t = t_i^*$, and take the cross-sample median. From this we build two error statistics.

Pointwise absolute log error (ALE) and its cross-technology median (MedALE). **Primary metric, used in main paper Fig. 3a, b and Fig. 27 of this SI.** Per technology, the absolute log error at diffusion target d^* is

$$\text{ALE}_i(d^*) = \left| \log \left(\frac{\hat{Y}_i(t_i^*)}{Y_i(t_i^*)} \right) \right|. \quad (56)$$

We aggregate across technologies by the cross-technology median, denoted MedALE,

$$\text{MedALE}(d^*) = \text{median}_{i \in \mathcal{I}(d^*)} \text{ALE}_i(d^*), \quad (57)$$

where $\mathcal{I}(d^*)$ is the set of technologies whose diffusion reaches d^* within the observed window. The cross-technology median is robust to a small number of outlier-fit technologies whose per-horizon log error sits substantially above or below the typical level. It pairs naturally with the median posterior forecast $\hat{Y}_i$ to give a *typical technology, typical forecast* summary. We also report the cross-technology *mean* as a robustness check (Table 24).

The noise floor near the forecast origin. A correctly-specified model still makes errors, because the data carry irreducible noise. The expected absolute log error of a single forecast is bounded below by this noise floor (Eq. (42), Section 1.4). At $d = 0.05$ the floor is about 0.25, the best accuracy any model can reach there. The surrogate MedALE sits at the floor for short forecast horizons.

From about $d^* = 15\%$, where the noise floor has dropped enough to no longer dominate, the empirical MedALE for B-R rises above the surrogate 5–95% band, and the gap grows through the middle of the diffusion range. This gap is the model's error beyond noise: parameter-estimation error and structural departures of the data from the B-R process. We estimate $\hat{\sigma}_i^{1-95}$ on the diffusion range where the B-R model is well specified, excluding the early super-exponential regime and the saturation tail (Section 1.4). Re-estimating it on the wider 0–95% or tighter 5–95% window shifts the band only slightly and leaves the empirical curve above it from $d^* \approx 15\%$ until it re-enters near saturation (Section 4.1.6).

4.1.1.4 Distributional calibration at each diffusion target

We assess distributional calibration with Q-Q plots of the probability integral transform (PIT), following the methodology of Section 1.3. For technology i at diffusion target d^* , let F_i be the forecast CDF of the posterior predictive distribution at the crossing year $t_i^*(d^*)$ given by the interpolation method of Section 4.1.1.3, and $Y_i(t_i^*)$ the realised value there. The PIT value is

$$u_i(d^*) = F_i(Y_i(t_i^*)).$$

This interpolation yields one PIT value per technology per diffusion target. We treat distinct technologies as independent (pairs such as *sailing ships/steamships* or *telegrams/telephones* are a small minority of the cross-sector dataset), so at each d^* the PIT values form an i.i.d. sample of size n equal to the number of technologies whose data crosses d^* . This n sets the Beta-order-statistic confidence band of Section 1.3. A Q-Q curve on the 45° line is calibrated.

4.1.2 Backtest results

4.1.2.1 Point-forecast accuracy

Main paper Fig. 3a, b reports the MedALE for the B-R, Logistic, and Gompertz models on the 120-technology calibration set, peaking near $d^* \approx 0.5$ and declining toward saturation. Here we add three views of point-forecast accuracy: the same metric resolved by forecast horizon in years (Fig. 27), the MedALE and bias by forecast origin (Table 23), and the pointwise error by diffusion target (Table 24).

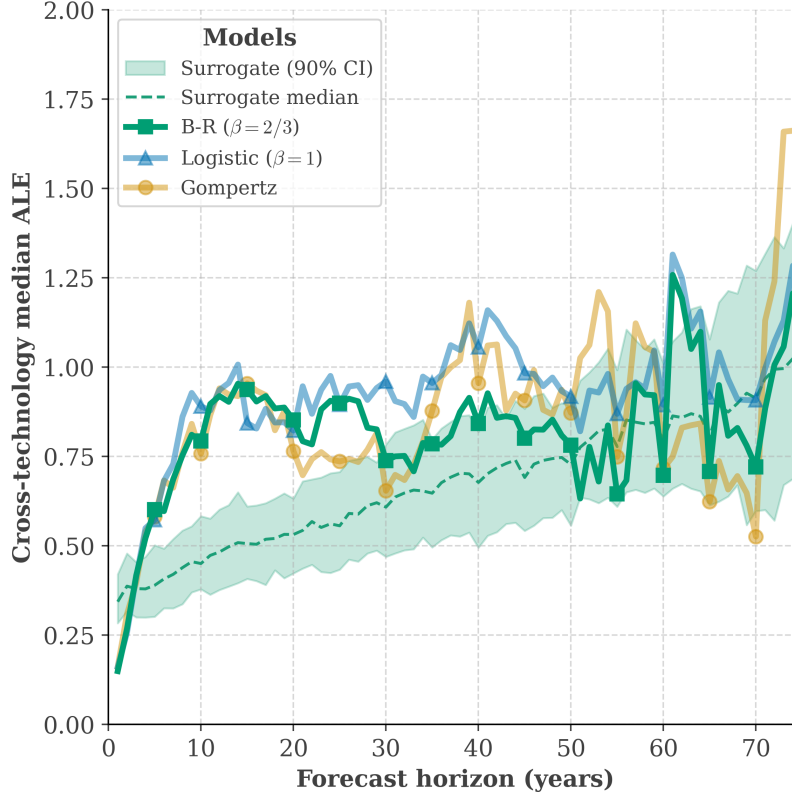


Figure 27: MedALE on the calibration set as a function of forecast horizon in years, at the 5% forecast origin. The teal band is the 5–95% surrogate envelope under the same cross-technology median aggregation. The empirical B-R curve exits the surrogate upper bound at about year 5. It oscillates around 0.7 to 0.9 through year 30. It then re-enters the surrogate band from about year 30 to 35 onward. The band widens at long horizons because fewer technologies remain in sample. Markers are placed at horizons 5, 10, 15, ... for readability.

Table 23 reports, for each model and forecast origin, the MedALE averaged over diffusion targets above the origin and the pooled median signed log error (a measure of bias). B-R has the lowest averaged MedALE at the 5% origin. Gompertz is lowest at the 10%, 25%, and 50% origins, but only through a large negative bias (-0.64 at 5%, shrinking at later origins). That is, under-prediction in level space partially cancels noise-driven over-prediction, lowering the absolute log error in a way a bias-free metric cannot. The pointwise ALE confirms this (Table 24): B-R is lowest at the high targets (50%, 70%, 100%), and Gompertz wins only at 10% and 25%. B-R’s and Logistic’s pooled bias are closest to zero at the 5% origin. B-R’s bias remains low at the 10% and 25% origins and trends more positive at the 50% origin.

Table 24 reports the pointwise ALE at five diffusion targets ($d_{\text{train}} = 5\%$) under both cross-technology median (MedALE) and mean (MALE) aggregations. The model rankings are the same under both. Absolute values are higher under the mean by about 0.1 to 0.2, reflecting moderate right skew, where a few technologies are harder to fit and pull the mean above the median, with no single outlier dominating. The gap is largest at $d^* = 100\%$, where N is smallest.

Table 23: MedALE, averaged across diffusion targets d^* on a 1% grid from the training origin up to 100%, and pooled median signed log error by model and forecast origin. This is the same metric reported per d^* in main paper Figure 3a, b, c, summarised across diffusion targets. The signed log error is the median of $\ln(\hat{Y}/Y)$ pooled across (technology, integer year horizon) pairs at each origin. Lower averaged median ALE is better. Bias closer to zero is better. The default universal prior of Eq. (32) is used throughout.

Forecast origin d_{train}	Averaged cross-tech median ALE			Median signed log error		
	B-R	Logistic	Gompertz	B-R	Logistic	Gompertz
5%	0.90	1.05	1.01	-0.07	-0.02	-0.64
10%	0.72	0.77	0.63	+0.07	+0.20	-0.33
25%	0.56	0.63	0.38	+0.11	+0.15	-0.13
50%	0.34	0.48	0.26	+0.30	+0.24	+0.12

Table 24: Pointwise absolute log error of the median forecast at forecast origin $d_{\text{train}} = 5\%$ under the universal prior of Eq. (32), evaluated at five fractional diffusion targets d^* . We report the cross-technology median as the primary aggregation (used in main paper Figure 3a, b and throughout this SI) and the cross-technology mean as a robustness check. Lower is better. Bold marks the best model at each d^* within each aggregation.

Aggregation	Model	$d^* = 10\%$	25%	50%	70%	100%
Cross-tech median	B-R ($\beta = 2/3$)	0.374	0.907	1.023	1.005	0.868
Cross-tech median	Logistic	0.457	1.032	1.220	1.209	0.948
Cross-tech median	Gompertz	0.345	0.804	1.092	1.285	1.171
Cross-tech mean	B-R ($\beta = 2/3$)	0.518	0.931	1.139	1.155	1.091
Cross-tech mean	Logistic	0.551	1.023	1.232	1.235	1.166
Cross-tech mean	Gompertz	0.489	0.890	1.180	1.257	1.240

4.1.2.2 Calibration results

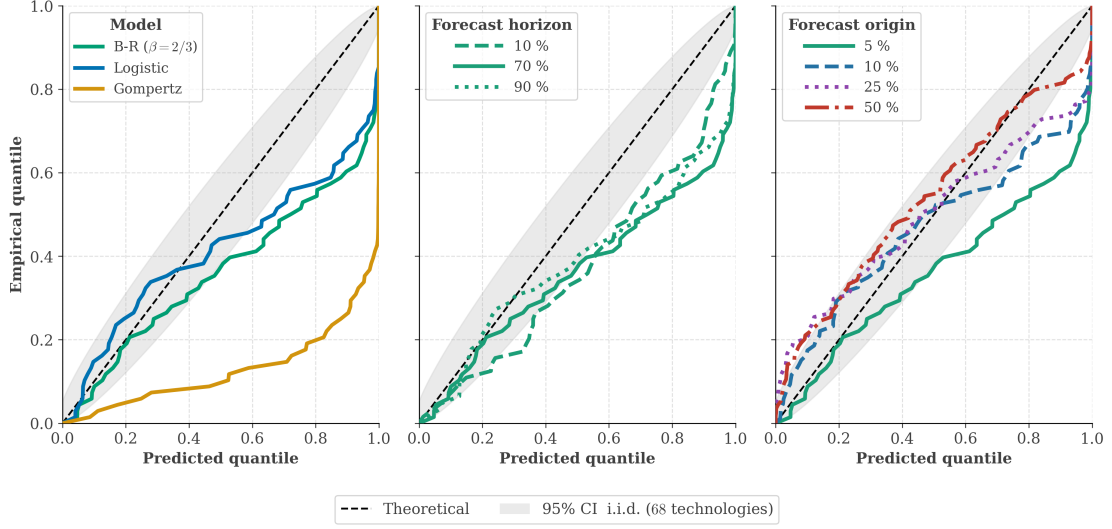
We now turn from point accuracy to distributional calibration. Figure 28 shows the PIT Q-Q plots without and with the post-hoc widening transform.

Figure 28a shows the un-widened PIT. The B-R curve sits below the diagonal across most of the range, most strongly in the middle, with two implications. At high percentiles too many observations land in the upper tail, so the forecast upper bound is too narrow and B-R is overconfident there. At low percentiles too few land in the lower tail, so the lower bound is mildly conservative. The asymmetric widening transform of Section 1.2.5 stretches the upper tail more than the lower, correcting the upper-tail overconfidence while keeping the lower bound conservative. After widening, the B-R curve tracks the diagonal better at every forecast origin and diffusion target (Fig. 28b). We calibrate the anchors in Section 4.1.4 and validate them by leave-one-technology-out cross-validation (Table 28).

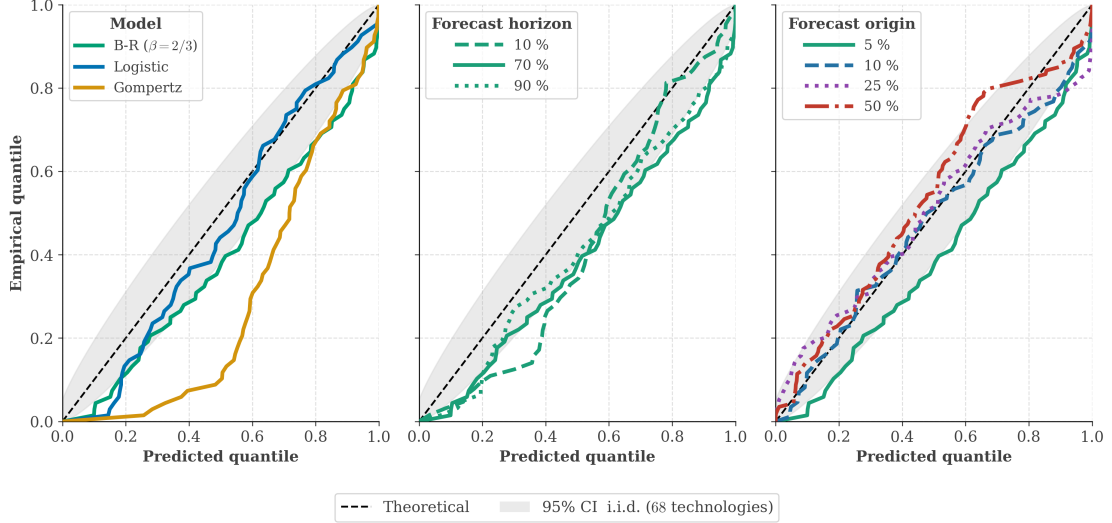
B-R. The Q-Q curve tracks the diagonal at low percentiles and falls below it at high percentiles, mostly within the confidence band. Its baseline KS against the uniform averages 0.23 across origins, and the widening reduces this to a mean of 0.13 (Table 28).

Logistic ($\beta = 1$). Logistic’s pooled bias at the 5% origin is slightly smaller than B-R’s (-0.02 vs -0.07 , Table 23). After widening it is only marginally better calibrated than B-R at the 5%, 10%, and 25% origins and slightly worse at the 50% origin (Table 27). It has worse MedALE than B-R at every origin (Table 23), and needs more widening to achieve the same calibration. The widening exponent needed for the Logistic model is not robust across the 120-technology and the less-filtered 154-technology sets (Table 30), so it is not a reliable choice for unseen technologies.

Gompertz. Gompertz has the largest pooled bias, negative at every origin below 50% (Table 23). Its widening exponent exceeds B-R’s at every origin (Table 27), most at low diffusion where the bias is strongest. Note that a median-preserving widening cannot move the biased median, so Gompertz stays miscalibrated.



(a) Baseline (un-widened)



(b) Widened (asymmetric)

Figure 28: Forecast calibration under the universal prior on the 120-technology set, with FP-shifted forecasts. PIT values are evaluated at the fractional year where each technology’s diffusion crosses the target. (a) Baseline calibration, without the post-hoc widening of Section 1.2.5. (b) The same comparisons after the asymmetric widening of Eq. 33, with anchors calibrated in Section 4.1.4. Within each row, the left panel compares the B-R, Logistic and Gompertz models at a 5% forecast origin and a 70% forecast horizon. The middle panel shows B-R coverage at a 5% forecast origin across three forecast horizons (10%, 70% and 90% diffusion). The right panel shows B-R coverage at a 70% forecast horizon across four forecast origins (5%, 10%, 25% and 50% diffusion). The forecast origin is the diffusion level at which a forecast is made and the forecast horizon the diffusion level at which it is evaluated. Same layout as main paper Figure 3(d-f). The grey band is the 95% pointwise i.i.d. confidence envelope for a uniform draw of $K = 68$ technologies.

4.1.2.3 Interpreting the results

The B-R MedALE rises from the early-diffusion noise floor, plateaus between $d^* \approx 30\%$ and $d^* \approx 80\%$, and declines toward saturation (main paper Fig. 3a, b). We explain this shape, link it to the PIT calibration, and end with our choice of model.

Downward bias in $\hat{k}$. The posterior median $\hat{k}$ sits below the full-sample maximum-likelihood estimate of k on the calibration set. On the 68-technology subset whose data crosses $d = 5\%$ (the

set used in main paper Fig. 3a), the cross-technology median ratio $\hat{k}/k_{\text{full}}$ is approximately 0.75 at $d_{\text{train}} = 5\%$. The bias decreases with more training data (i.e. later forecast origins). The bias is much smaller on the empirical surrogate of Section 1.4 ($\hat{k}/k_{\text{true}} \approx 0.94$ at $d_{\text{train}} = 5\%$). A plausible source is misspecification of the noise model at early diffusion. The surrogate is generated under the model’s noise structure by construction, so this source is largely absent there.

Heterogeneity in the true asymmetry β_i . The B-R model is fit with a universal $\beta = 2/3$. Technology-specific β_i are observed in the data (Section 3.1, Fig. 18b). Modest cross-technology departures from $\beta = 2/3$ introduce a parameter-fit mismatch between the universal model and the technology-specific truth, contributing to the empirical MedALE shape on top of the $\hat{k}$ bias above.

Reproducing the shape. A deterministic simulation isolates the contribution of each candidate mechanism. We draw $N = 120$ technologies with heterogeneous k_i , make shifted forecasts that pass through the training endpoint at $d_{\text{train}} = 5\%$, and take L as known. A 25% downward shift in $\hat{k}$ alone (Table 25, column B) reproduces the plateau-and-decay shape. Adding cross-technology heterogeneity in the true β_i around $2/3$ while the model assumes $\hat{\beta} = 2/3$ (column C) slightly widens the plateau. The empirical surrogate of Section 1.4, where neither condition applies, shows no plateau-and-decay (main paper Fig. 3a band).

Table 25: Cross-technology MedALE at $d_{\text{train}} = 5\%$ on $N = 120$ synthetic technologies, L known, FP-shifted forecasts. Column B: downward $\hat{k}$ bias ($\hat{k} = 0.75k$, $\hat{\beta} = \beta = 2/3$). Column C: B with heterogeneous true $\beta_i \sim \mathcal{N}(2/3, 0.08^2)$ and model $\hat{\beta} = 2/3$.

d^*	B: $\hat{k}$ bias	C: B + β_i het
0.10	0.167	0.168
0.30	0.371	0.375
0.50	0.395	0.401
0.70	0.341	0.347
0.95	0.129	0.133

Implication for PIT calibration. The downward $\hat{k}$ bias also explains the upper-tail overconfidence of the un-widened PIT seen above. Adoption grows faster than the median forecast expects, so too many observations fall in the upper tail, which the asymmetric widening then corrects more aggressively than the lower tail.

Saturation forecasts from $d_{\text{train}} = 5\%$ versus $d_{\text{train}} = 10\%, 25\%$. B-R forecasts made from the 5% origin are more accurate near saturation than those made from the 10% or 25% origins (main paper Fig. 3b). For diffusion targets $d^* \leq 70\%$ the usual ordering holds, where more training data gives smaller error. Near saturation the error is dominated by the asymptote L rather than by the amount of training data. The wide universal L prior is one likely factor, and the mechanisms above may also contribute. The diffusion-aware prior of Section 4.1.7 tightens the L prior by diffusion stage and materially lowers the high-diffusion MedALE. This points to prior calibration rather than a limit of the model.

Model selection. Gompertz is severely biased. At the 5% origin its median forecast is roughly half the observed value (Table 23). It needs the largest widening of the three models across origins and stays badly calibrated even after it (Table 27). It has lower MedALE at some origins and forecast horizons, but only because that bias cancels noise in the median forecast. Logistic is nearly unbiased and calibrates well, but it is less accurate than B-R at every origin (Table 23). B-R is the only model that combines accurate point forecasts, near-unbiasedness, and well-calibrated prediction intervals after only modest widening. It has the lowest averaged MedALE at the 5% origin (Table 23) and the lowest error at every diffusion target above 50% (main paper Fig. 3a). We therefore select B-R as our primary model.

4.1.3 The Bass model

We benchmark a fourth model class against B-R: the Bass diffusion model [51], which explicitly separates innovation (p) and imitation (q) effects. The Bass model is often presented as a differential equation, but it has a closed-form solution for the cumulative diffusion $Y(t)$, given by

$$Y(t) = L \frac{1 - e^{-k(t-t_0)}}{1 + r e^{-k(t-t_0)}}, \quad k = p + q, \quad r = q/p, \quad (58)$$

where L is the market potential, r is the imitation-to-innovation ratio, and $Y(t_0) = 0$. Setting $r = 14$ places the inflection at $Y^* = L(r-1)/(2r) = 46.4\%$ of L , virtually identical to the B-R model with $\beta = \frac{2}{3}$ (46.5%), enabling a like-for-like comparison. We apply the same MA(1) noise structure in FP-transformed space with $\rho = 0.80$ and $\epsilon_{i,T} = 0$, and the same universal prior (Eq. (32)).

Applying this framework to the 120-technology dataset gives substantially worse backtests than B-R. The median $\log(\hat{Y}/Y)$ is strongly negative (i.e. forecasts are biased downward), against near-zero bias for B-R. The distributional forecasts are very poorly calibrated. We did not recalibrate the prior for Bass, so the comparison is not fully controlled, and a dedicated recalibration might narrow the gap. We therefore draw no strong conclusion about the Bass model class, beyond noting that under our prior and noise assumptions it is clearly less accurate than B-R on this dataset.

4.1.4 Calibrating the widening transform on backtesting data

The adaptive widening transform introduced in Section 1.2.5 requires the anchor values $w^*(d)$ that enter Eq. (35). Here we calibrate those anchors on the 120-technology backtesting set, validate the asymmetric form (Eq. (33)) against a symmetric variant and against the empirical surrogates, and report the resulting calibration improvement.

Fitting the widening exponent. We calibrate one exponent w per diffusion origin $d \in \{0.05, 0.10, 0.25, 0.50\}$. For each technology we compute the predictive distribution at the fractional year where the data cross the target diffusion level $d^* \in \{5, 10, \dots\}\%$. We then pool PIT values across technologies and target levels and choose w to minimise the Kolmogorov–Smirnov (KS) statistic

$$D(F_1, F_2) = \sup_x |F_1(x) - F_2(x)|, \quad (59)$$

where F_1 and F_2 are any two CDFs and x ranges over their common support. Here we use the one-sample form with $F_1 = \hat{F}_{\text{PIT}}$ (the empirical CDF of the pooled PIT values) and $F_2(u) = u$ (the $\mathcal{U}(0, 1)$ CDF), so $x = u \in [0, 1]$. We minimise using a bounded scalar search on $w \in [0.5, 8]$.

Table 26 reports w^* at each forecast origin for the asymmetric form (Eq. (33), used throughout the paper) and the symmetric variant in which the same exponent is applied above and below the median (without the $\sqrt{w}$ in the lower-tail branch). The symmetric exponents are smaller because they would over-widen the already-calibrated lower tail. The asymmetric form yields a lower widened KS at $d = 0.05$ (0.14 vs 0.16), $d = 0.10$ (0.10 vs 0.13), and $d = 0.25$ (0.14 vs 0.15). The two forms are essentially tied at $d = 0.50$ (0.12 vs 0.12). The square-root lower-tail exponent was itself chosen from a small sweep across $w^{1/2}$, $w^{1/3}$, $w^{1/4}$, $w^{1/5}$, and no lower-tail widening. $\sqrt{w}$ gave the best average calibration across the forecast origins.

Per-model anchors. The widening exponents above are fitted on the B-R PIT distribution. To check the other models, we re-fit w^* independently on the Logistic and Gompertz distributions at every origin (Table 27). B-R needs the least widening overall. Logistic needs more, especially at the 5% origin, but reaches comparable calibration once widened. Gompertz needs the most at every origin and stays the worst calibrated, because a median-preserving widening cannot remove its negative bias. Figure 29 shows each model widened with its own exponent: B-R reaches the diagonal at every origin, Logistic as well (slight exception at 25% forecast origin), and Gompertz stays off the diagonal below the 50% origin. As more of the series is observed, the data dominate the prior and the models’ growth descriptions converge, so the calibration gap between them narrows (as observed in Figure 29 panel d, where all three models lie nearer the diagonal post widening.).

Table 26: Widening exponent w^* by forecast origin d on the 120-technology backtesting set, fitted under the KS objective. **Asymmetric** applies w above the median and $\sqrt{w}$ below (Eq. 33). **Symmetric** applies w on both sides. The asymmetric form is used throughout the paper. The widened KS is the deployed value, after applying the anchors adaptively per technology as $w^*(1/\tilde{L}^{\text{med}})$.

Forecast origin d	Asymmetric		Symmetric	
	w^*	widened KS	w^*	widened KS
0.05	1.93	0.142	1.76	0.157
0.10	2.72	0.101	2.03	0.126
0.25	3.58	0.137	2.84	0.151
0.50	1.73	0.122	1.34	0.119

Table 27: Per-model widening on the 120-technology backtesting set. For each model we fit the anchor exponent w^* per forecast origin on the model own un-widened PIT distribution, with the asymmetric widening of Eq. 33. Baseline KS measures the un-widened forecasts against uniformity. Widened KS measures the same after the adaptive widening.

Model	Forecast origin d	Baseline KS	w^*	Widened KS
B-R	0.05	0.296	1.93	0.142
	0.10	0.216	2.72	0.101
	0.25	0.215	3.58	0.137
	0.50	0.173	1.73	0.122
Logistic	0.05	0.252	3.37	0.095
	0.10	0.177	2.69	0.091
	0.25	0.240	3.40	0.119
	0.50	0.246	3.82	0.143
Gompertz	0.05	0.599	5.14	0.378
	0.10	0.488	4.28	0.310
	0.25	0.373	6.02	0.198
	0.50	0.176	2.86	0.144

Held-out cross-validation. We fit the widening anchors of Table 26 on the backtesting set, so the reported widened KS is an in-sample quantity. We therefore cross-validate by leaving one technology out. We refit the anchors on the remaining technologies, then apply the adaptive widening $w^*(1/\tilde{L}^{\text{med}})$ to the held-out one. We repeat this for every technology and pool the held-out PIT values. The held-out KS matches the in-sample KS to within 0.005 at every origin, and w^* is stable across folds (Table 28). The widening is therefore a genuine out-of-sample recalibration of interval width, not an artefact of in-sample fitting. We use the 120-technology in-sample anchors for all forecasts in the paper.

Table 28: Leave-one-technology-out cross-validation of the adaptive widening on the 120-technology backtesting set. The model is B-R, the objective is KS, and the transform is the asymmetric sqrt-lower form of Eq. 33. KS (raw) is the un-widened PIT KS against the uniform. KS (in-sample) is the adaptive widened KS under the deployed anchors. KS (LOO-CV) refits the anchors on the other $N - 1$ technologies, applies the adaptive exponent to the held-out technology, and pools the held-out PITs. The w^* LOO range is the refit anchor range across folds.

Forecast origin d	w^*	KS (raw)	KS (in-sample)	KS (LOO-CV)	w^* LOO range
0.05	1.93	0.296	0.142	0.145	[1.91, 2.68]
0.10	2.72	0.216	0.101	0.102	[2.27, 2.89]
0.25	3.58	0.215	0.137	0.139	[3.24, 3.65]
0.50	1.73	0.173	0.122	0.121	[1.60, 1.96]

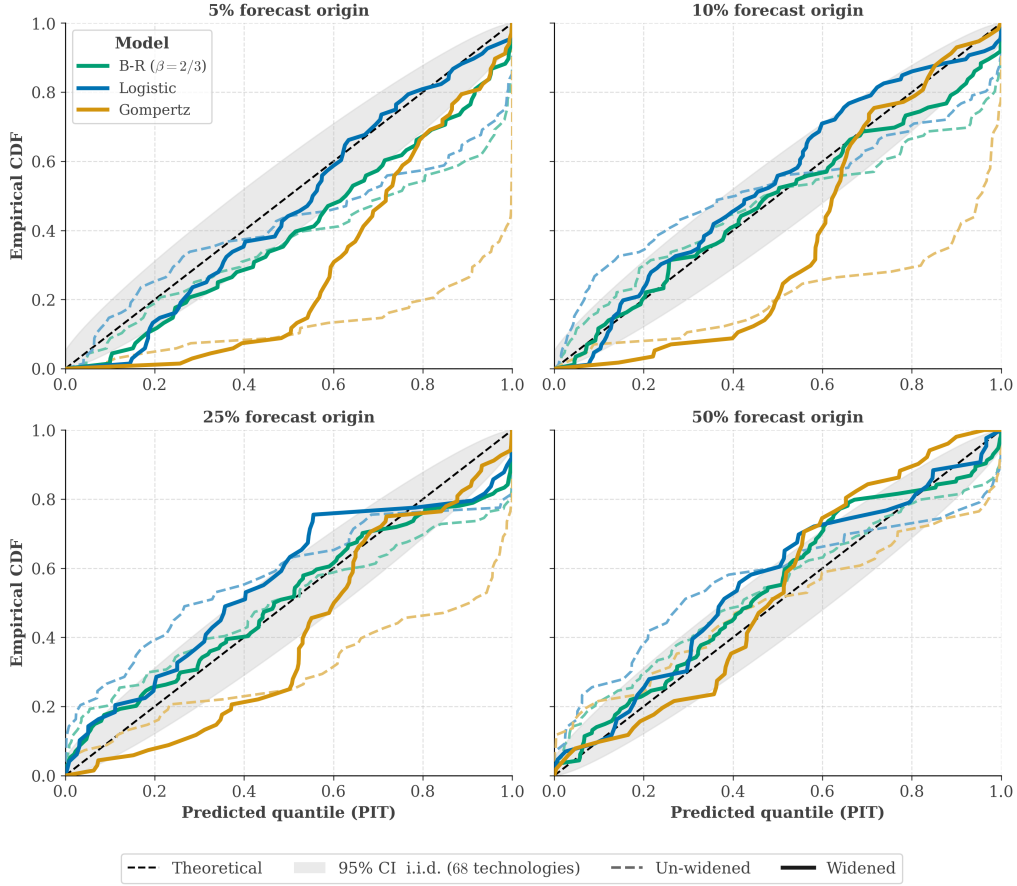


Figure 29: Un-widened and Widened PIT calibration of the three models at the four forecast origins, at the 70% diffusion target. PIT computation and layout follow Section 4.1.1.4 and main paper Fig. 3(d). Each model’s un-widened curve is the lighter dashed line of the same colour, and its widened curve is the solid line. B-R and Logistic reach the diagonal at every origin.

Validation on the empirical surrogates. For a correctly-specified model the PIT is approximately uniform by construction, so re-fitting w^* on surrogate data should return values close to 1. We tested this on the empirical surrogate at the 5% forecast origin, using the same universal prior as the main analysis. The fitted exponent is $w^* \approx 1.1$, far below the empirical $w^* \approx 1.9$ at the same origin (Table 26). It stays near one and is leave-one-out stable whether we estimate the per-technology noise scale $\hat{\sigma}_i$ on the primary 1–95% diffusion window or on the wider 0–95% window that keeps all early-diffusion data (both defined in Section 1.4; Table 29), so the 1% early-diffusion cut does not drive the result. The small residual above 1 likely reflects mild prior misspecification, which can bias the posterior [151, 152], rather than miscalibration of the model form.

Table 29: Widening exponent w^* on the empirical surrogate at the 5% forecast origin, fitted under the KS objective, for the primary 1–95% and the wider 0–95% noise windows (Section 1.4). The last column gives the leave-one-technology-out range of w^* .

Surrogate	Widening	w^*	w^* (LOO range)
$\hat{\sigma}_i^{1-95} (\approx 0.28)$	Asymmetric	1.13	[1.10, 1.14]
$\hat{\sigma}_i^{1-95} (\approx 0.28)$	Symmetric	1.10	[1.08, 1.10]
$\hat{\sigma}_i^{0-95} (\approx 0.33)$	Asymmetric	1.19	[1.17, 1.22]
$\hat{\sigma}_i^{0-95} (\approx 0.33)$	Symmetric	1.13	[1.11, 1.14]

Section 1.3 shows that under a correctly specified model the PIT calibration depends on

whether the prior matches the true parameter distribution. The surrogate is correctly specified by construction, so its slightly off-diagonal PITs reflect prior misspecification rather than model form. On real data, prior misspecification may offset model misspecification, consistent with the approximately well-calibrated real-data PITs.

4.1.5 Robustness to subjective data cleaning

Step 4(b) of the data cleaning procedure (Section 2.0.1) removes 34 time-series through partly subjective judgement. Step 6 removes one further series for high correlation with another. To check that the main results do not depend on these removals, we repeat the full backtesting on a set that retains all 35 of these series.²⁸ The pipeline is identical to the 120-technology analysis. We use the same universal prior of Eq. (32). We fit the widening exponent w^* on this set independently. We look up the per-technology widening through the same adaptive map $\hat{d}_i = 1/\tilde{L}_i^{\text{med}}$ as in the main analysis, with no future knowledge of the saturation level.

Figure 30 shows the MedALE on the 154-technology set. The model ranking matches the 120-technology set. B-R has generally lower MedALE and the peak in MedALE remains. Gompertz has the largest negative bias (median ≈ -0.54). We do not overlay an empirical surrogate band here, because the surrogate is calibrated on the 120-technology parameter set and would mix datasets.

Baseline calibration is shown in Figure 31a and widened calibration in Figure 31b. Both visually follow the 120-technology version.

Table 30 compares the widening exponent w^* re-fitted on each set, per model. B-R is stable: its exponents change by at most 0.32 across origins (13% in relative terms), and its baseline KS, the distance of the un-widened PIT from the uniform, changes by at most 0.03. The benchmark models move more. Logistic is far more sensitive at the 5% origin, where its exponent falls from 3.37 to 2.47 (difference of $\approx 27\%$) while B-R’s barely moves (1.93 to 1.85). The subjective cleaning step therefore does not materially change B-R’s calibration, point-forecast accuracy, or widening.

Table 30: Widening exponent w^* and baseline KS for the 120-technology and 154-technology backtesting sets, per model, under the KS objective with the asymmetric widening of Eq. 33. The 154-technology set retains the 34 subjectively excluded series from Step 4(b). Baseline KS is the distance of the un-widened PIT from the uniform. B-R is shown at every forecast origin; Logistic and Gompertz at the 5% origin only.

d_{train}	w^*			Baseline KS		
	120	154	Δw^*	120	154	ΔKS
<i>B-R</i> ($\beta = 2/3$)						
0.05	1.93	1.85	-0.08	0.295	0.278	-0.017
0.10	2.72	3.04	+0.32	0.217	0.236	+0.019
0.25	3.58	3.57	-0.01	0.216	0.211	-0.005
0.50	1.73	1.50	-0.23	0.174	0.145	-0.029
<i>Logistic</i>						
0.05	3.37	2.47	-0.89	0.251	0.243	-0.007
<i>Gompertz</i>						
0.05	5.14	5.45	+0.31	0.601	0.564	-0.037

The leave-one-technology-out cross-validation of Table 28 also holds on the 154-technology set (Table 31).

²⁸This gives 155 series in total. One of them, annual space launches, declines sharply in its final years. All of its observations then fall within the training window at every backtest origin. It yields no out-of-sample forecast point, so it contributes to no calibration curve. This leaves the 154 technologies analysed here.

Table 31: Leave-one-technology-out cross-validation of the adaptive widening on the 154-technology set. This set adds the 34 subjectively excluded series to the 120 (cf. Table 30). Columns are as in Table 28. The held-out and in-sample KS again agree to within 0.01 at every origin. The widening calibration is robust to the data-selection step.

Forecast origin d	w^*	KS (raw)	KS (in-sample)	KS (LOO-CV)	w^* LOO range
0.05	1.85	0.278	0.156	0.156	[1.71, 1.88]
0.10	3.04	0.236	0.123	0.129	[2.88, 3.37]
0.25	3.57	0.212	0.143	0.144	[3.43, 3.88]
0.50	1.50	0.144	0.125	0.123	[1.42, 1.52]

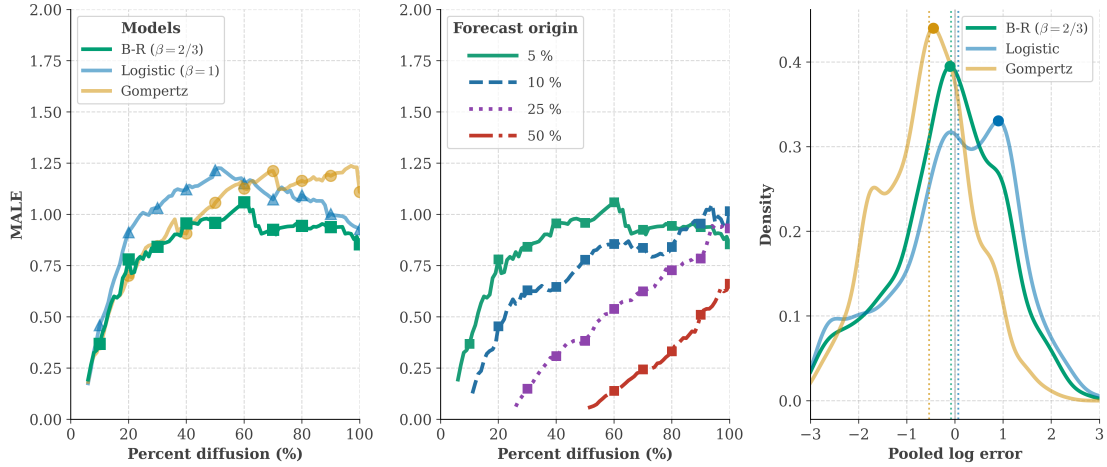
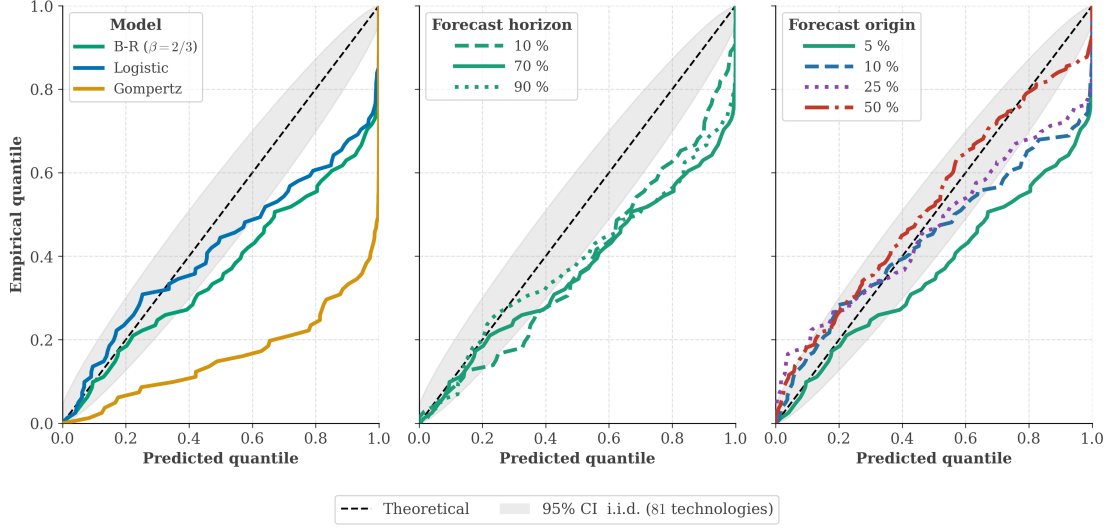
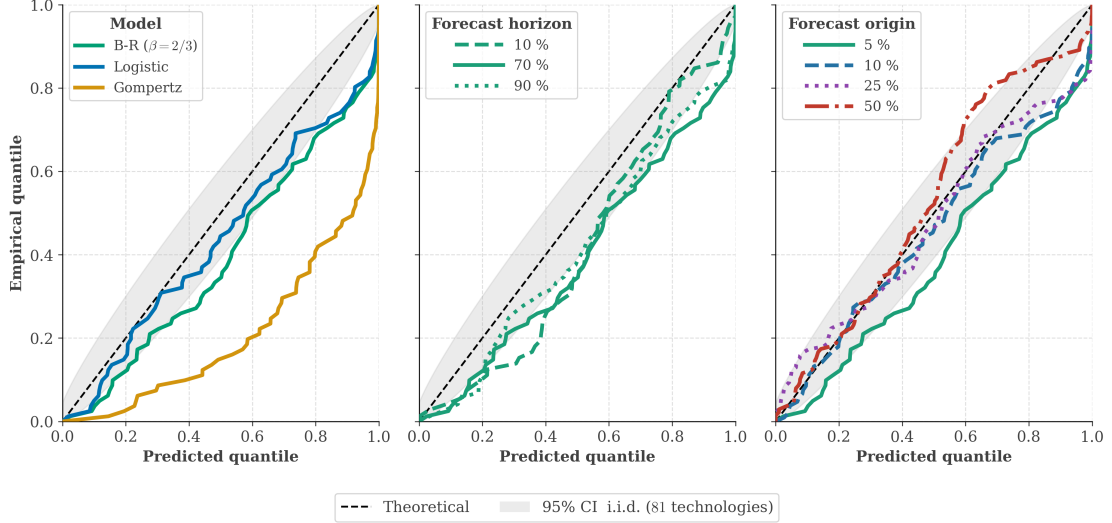


Figure 30: Forecast accuracy on the 154-technology dataset using FP-shifted forecasts. Same layout as main paper Figure 3a, b, c. As there, the y -axis median _{i} is the median across technologies. **(a)** MedALE at each diffusion target for B-R, Logistic, Gompertz at the 5% forecast origin. **(b)** The same for B-R at four forecast origins (5%, 10%, 25%, 50%). **(c)** Pooled per-(technology, horizon) log error $\ln(\hat{Y}/Y)$ density at the 5% origin.



(a) Baseline (un-widened)



(b) Widened (asymmetric)

Figure 31: Calibration on the 154-technology dataset. **Top:** baseline (un-widened) PIT. **Bottom:** after asymmetric widening of Eq. 33 with anchors re-calibrated on this set (Table 30). Same layout as main paper Figure 3(d-f) and Figure 28a: within each row the left panel compares the B-R, Logistic and Gompertz models at a 5% forecast origin and 70% forecast horizon, the middle panel shows B-R coverage at a 5% forecast origin across three forecast horizons (10%, 70% and 90% diffusion), and the right panel shows B-R coverage at a 70% forecast horizon across four forecast origins (5%, 10%, 25% and 50% diffusion).

4.1.6 Further robustness checks

Here we present a brief overview of robustness and sensitivity analyses not included in the main text. With the exception of the prior sensitivity checks, we keep the same priors as in main paper Fig. 3. We show sensitivity analyses for prior specification, for varying β and ρ , and for estimating technology-specific β_i and ρ_i , and discuss how forecast performance improves for more *well-behaved* technologies (i.e. those more similar to our null model in terms of R^2).

Sensitivity to priors. Our results are robust to moderate changes in the prior hyperparameters, though extreme misspecification does degrade performance. A narrow prior centred near the true L and t_0 improves forecasts. But narrow priors may understate parameter uncertainty and become

overconfident, and technologies that fit the narrow prior poorly forecast substantially worse. We therefore keep the prior as uninformative as possible, while using information from data available to us.

We also varied the t_0 offset, using $2/\hat{k}_{\text{ssr}}$ or $4/\hat{k}_{\text{ssr}}$ in place of $3/\hat{k}_{\text{ssr}}$, and found the results not hugely sensitive. A tighter offset ($\ell_{t_0} > T_{\text{max}} - 2/\hat{k}_{\text{ssr}}$) helps marginally at low diffusion but performs poorly at high diffusion. A fixed universal location $\ell_{t_0} = 5$ still works reasonably but is less accurate at low diffusion and worse calibrated than the data-informed location. Setting the prior by diffusion stage helps more, which we examine in Section 4.1.7. Under the tighter offset, the MedALE rankings (Table 23) are preserved, with similar widening and calibration, further evidence that B-R is a reasonable default.

Sensitivity to constant parameters. We analyze sensitivity to the choice of technology-invariant β and ρ , varying each independently while keeping the other fixed at the baseline value. Varying ρ has a minimal effect on MedALE and calibration. Varying β has a small effect on MedALE but can strongly affect calibration. For example, $\beta = 1/3$ leads to underdispersed forecasts that underestimate upward deviations, while $\beta = 3/2$ leads to overdispersed forecasts at the lower quantiles of the forecast distribution. For $\beta = 1/2$, forecast performance is similar to the baseline, albeit marginally worse, indicating that the choice of β is not critical as long as it remains within a reasonable range of around $[0.5, 1]$.

We also estimate technology-specific β_i and ρ_i , and find that the resulting forecasts have similar point-forecast accuracy but worse probabilistic calibration than the baseline with constant β and ρ (less uniform PIT distributions and visible deviations from the 45° line in the Q-Q plot). This suggests that technology-specific β_i and ρ_i are not well identified, and that estimating them leads to overfitting and worse out-of-sample performance.

Forecast performance as a function of in-sample fit. Forecast performance is better for technologies with better in-sample fits, as measured by the R^2 of the full-sample fit. Technologies that behave more similarly to the B-R model in-sample are forecast more accurately out-of-sample. This is expected: technologies better described by the assumed DGP in-sample are likely to yield better forecasts out-of-sample.

4.1.7 The role of the prior: bounding forecast quality

The universal empirical-Bayes prior (Eq. (32)) occupies a middle ground between two extremes. To put its performance in perspective, we evaluate the B-R model under two bracketing configurations:

- **Wide (agnostic) prior, floor.** $\tilde{L}_i \sim \Gamma(2, 1, 100)$ with a fixed, non-data-informed t_0 location ($\ell_{t_0} = 5, s_{t_0} = 30$). This prior encodes almost no information about the technology’s saturation level or timing.
- **Diffusion-aware prior, ceiling.** Priors for L and t_0 are set conditional on the observed diffusion band at the forecast origin, using the five-band specification in Table 32. This encodes the qualitative diffusion-stage knowledge that a practitioner would typically infer from visual inspection of the adoption curve. It does not require knowledge of the true asymptote.

The three configurations thereby span the range from prior ignorance (wide prior), through our default cross-sectionally calibrated prior (universal, $s_L = 35$), to the best case available with observable domain knowledge (diffusion-aware).

Wide prior. Under a wide prior ($s_L = 100$), point accuracy deteriorates. The MedALE is higher than under the universal prior across the whole diffusion range (main paper Fig. 3a, b), most at mid-to-high diffusion where the L posterior is more diffuse. Calibration worsens only slightly. The ordering of B-R, Logistic, and Gompertz is unchanged.

Diffusion-aware prior. The diffusion-aware prior conditions on the approximate diffusion band observable at the time of forecasting. An analyst can distinguish near-exponential early growth from a trajectory with visible curvature without knowing the true asymptote, using only information available at forecast time. In that sense, diffusion-stage awareness is the kind of qualitative prior knowledge a practitioner would typically have.

The prior specification is

$$\pi(\theta_i \mid \text{diffusion band}) = \begin{cases} \tilde{L}_i \sim \Gamma(\alpha_L, \ell_L, s_L), \\ t_{0,i} \sim \Gamma(\alpha_{t_0}, \ell_{t_0}, s_{t_0}), \\ k_i \sim \Gamma(1, 0, 1), \\ \sigma_i \sim \Gamma(1, 0, 1), \end{cases}$$

with hyperparameters varying by diffusion band as in Table 32. The state-dependence reflects structural identification differences. At early diffusion the asymptote L_i and inflection timing $t_{0,i}$ are weakly identified from near-exponential growth, so the prior must remain wide. As curvature becomes visible the likelihood becomes more informative about both parameters, and tighter priors are warranted. Because L_i is normalised by the last observed value, the same numerical scale implies very different absolute asymptotes at 5% versus 30% diffusion, so both the scale and location of the prior are adjusted across bands.

Table 32: Hyperparameters of the diffusion-aware prior by forecast-origin band. All bands: $k_i \sim \Gamma(1, 0, 1)$, $\sigma_i \sim \Gamma(1, 0, 1)$. Here $\tilde{t}_T$ denotes the index of the last observation (first observation at $t = 0$).

Diffusion band	α_L	ℓ_L	s_L	α_{t_0}	ℓ_{t_0}	s_{t_0}
<2%	4	5	30	3	$\tilde{t}_T$	35
2–9%	2	1	40	2	$\max(5, \tilde{t}_T - 5)$	20
9–20%	2	1	25	1.5	$\max(5, \tilde{t}_T - 5)$	25
20–45%	2	1	5	1.5	$\tilde{t}_T/2$	15
>45%	1.5	1	3	1.5	$\tilde{t}_T/4$	20

Under the diffusion-aware prior, point-forecast accuracy improves substantially at a 5% forecast origin. The B-R MedALE is materially lower than under the universal prior across the diffusion range, with the largest gains at intermediate diffusion. Probabilistic calibration is also preserved. The PIT Q-Q plot remains well within the confidence bands, so the tighter diffusion-stage priors do not introduce overconfidence.

Implication. Across the three prior configurations, the ordering of forecast error is consistent. The wide prior is worst, the universal prior is in the middle, and the diffusion-aware prior is best. Our universal prior already achieves a substantial improvement over the wide baseline by exploiting two observable quantities. These are the series-specific growth rate $\hat{k}_{\text{ssr},i}$ and the last observed value $y_{T_{\max}}$. The diffusion-aware results show that analysts who can judge a technology’s diffusion stage from the shape of the adoption curve can expect considerably better forecast quality. We therefore encourage practitioners to exploit available prior information about the technology when applying the B-R model.

4.2 Backtest on energy technologies

Primary energy demand increased by about 2% annually over 75 years, increasing useful energy generation in TWh despite constant or decreasing percentage shares for many technologies.

In this section, we compare backtesting performance of three approaches: the **B-R model** (standard S-curve applied to TWh), the **growing-market B-R model** (S-curve with an exponentially growing asymptote, applied to TWh), and the **shares B-R model** (standard S-curve applied to market shares). The motivation is that many energy technologies have grown in absolute terms alongside total energy demand even after reaching maturity in share terms. Ignoring this can cause the fixed-asymptote model to dramatically underestimate future levels.

The three cases examined here are deliberately adversarial and should not be interpreted as representative of typical model performance. They are chosen precisely because they expose specific failure modes: (1) gas and hydropower reached share-saturation but continued growing in TWh as aggregate energy demand grew; (2) oil experienced a large, unpredictable production shock (a sharp swing in oil production around the 1979 peak) that altered its near-term trajectory; (3) nuclear energy saw rapid growth in the 1970s that was abruptly slowed by a sequence of major accidents and regulatory responses — not a *failed technology* (reactors continue to operate and new ones are being built), but a case in which the realised diffusion path deviated sharply from what data available at an early forecast origin would have predicted. Typical forecast performance across the full 120-technology ensemble is reported in Section 4.1 (and summarised in main paper Figure 3). The modeling choice underlying our main solar and wind forecasts (main paper Figure 4) is discussed in Section 5.2.

Backtests of useful energy generation with the standard B-R model use the universal prior with the model-direct adaptive widening calibrated in Section 4.1. This is the same method as our main-text forecasts. Backtests of market shares and of the growing-market model use no post-hoc widening.

4.2.1 Historical energy demand growth

To better understand the growth dynamics of energy technologies, we first examine historical trends in primary energy demand and useful energy generation for individual technologies. We calculate the growth in primary energy as log-differences, defined in Eq. 44. For convenience, we repeat the formula with new notation here,

$$g(t) = \frac{\log E(t) - \log E(t - h)}{h},$$

where h is the number of years between two consecutive data points, and $E(t)$ is the primary energy generated at time t .

Steady growth in primary energy consumption. Figure 32 shows the historical primary energy consumption in terawatt-hours (TWh) and the corresponding year-on-year increase in percentages, using data from [101], showing that primary energy has increased by approximately 2% per year for the past 100 years. Early growth hovers around 1% per year. From 1920 onwards, the average growth rate hovers around 2% per year, with some notable deviations. The period from 1960 to 1970 saw sustained growth above 2%, sometimes reaching 5%. The financial crisis in 2008 caused a temporary dip, and the COVID-19 pandemic in 2020 caused a significant drop — the only two years in the past 70 years where energy consumption decreased year-on-year. Despite these deviations, the long-term 2% trend has remained remarkably stable.

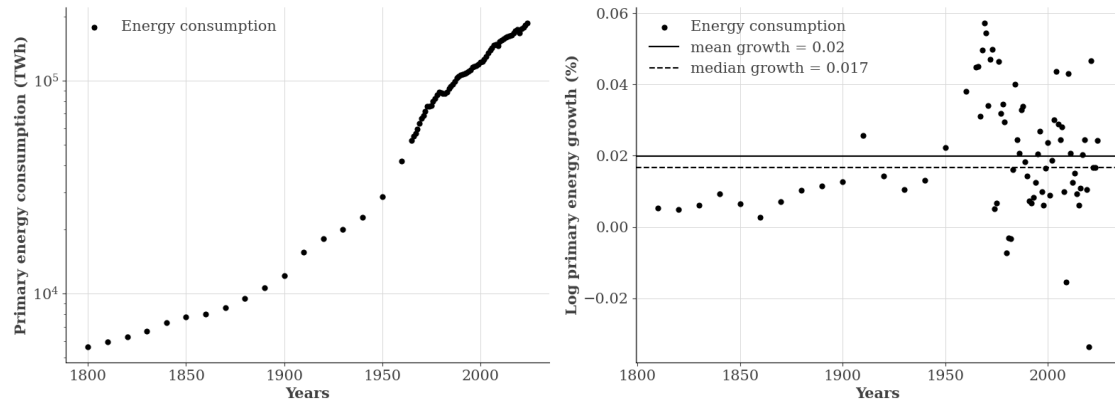


Figure 32: Historical primary energy consumption in TWh and the year-on-year growth rate. Primary energy has increased by approximately 2% per year for the past 100 years.

Growth in useful energy generation for individual technologies. Figure 33 shows the historical useful energy generation for selected energy technologies (in shares and TWh).

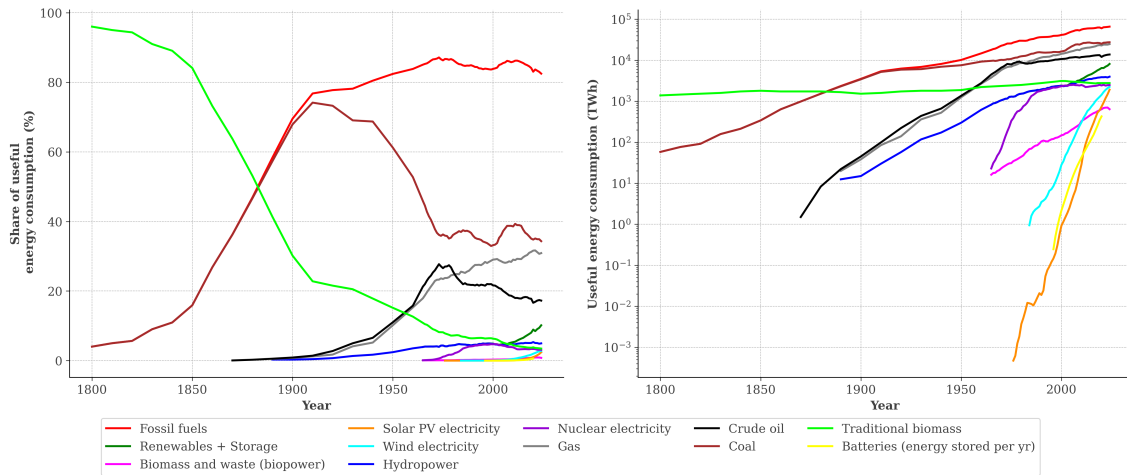


Figure 33: Historical useful energy generation for selected technologies in (a) shares and (b) TWh (in log scale). Useful energy generation for most fossil-fuel technologies (red) has increased in TWh over the past 20-30 years, while their share has remained approximately constant or even decreased. Examples shown are oil (black) and gas (gray).

Fossil fuels (sum of gas, oil, and coal, in red) have hovered around 80% of useful energy generation for the past 30 years, while their generation in TWh has steadily increased, evidenced by the constant line in panel (a) versus the red curve in panel (b). Oil (black) decreased in share while increasing in TWh of production. Gas (gray) increased in both.

Nuclear (purple) and traditional biomass (lime) are the only technologies shown that decreased in both share and total energy generation over the past decades. Nuclear fell slightly in share (from a plateau of $\approx 4.5\%$ to $\approx 3\%$) and marginally in TWh [101]. Traditional biomass decreased significantly in share and marginally in TWh [101, 153].

4.2.2 Comparing backtest performance of three approaches

Here we present the backtest results using three different approaches. (1) Forecasting in shares using the shares B-R model. (2) Forecasting in TWh using the fixed-asymptote B-R model. (3) Forecasting in TWh using the growing-market B-R model. We start by introducing the growing market model.

4.2.2.1 Modeling market growth explicitly

To improve forecast performance in growing markets, we extend the B-R model by replacing the fixed asymptote L_i with a time-varying carrying capacity $L_i \exp(gt)$, where g is an exogenous annual growth rate. This is equivalent to assuming the market ceiling itself grows exponentially, a good approximation when aggregate energy demand grows at a roughly constant rate. Concretely, we model technology diffusion using the adjusted B-R curve

$$Y_{i,t} = \frac{L_i \exp(gt)}{(1 + \exp(-\beta k_i(t - t_{0,i}) + \beta \sigma_i(\rho \epsilon_{i,t-1} + \epsilon_{i,t})))^{1/\beta}} \quad (60)$$

Estimating g from technology-specific time-series is difficult, especially for dominant technologies at early stages of diffusion where g and k are hard to separate. We set $g = 0.02$ based on Figure 32.

Limitations of the growing market assumption. In practice, as discussed in Section 2, technologies may decline rather than keep growing at the market rate after reaching maturity, so the 2% p.a. assumption does not hold uniformly. Oil useful energy consumption grew from approximately 8,500 TWh in 1983 to 13,500 TWh in 2022 ($\sim 75\%$ over 40 years, or $\sim 1.3\%$ p.a.). Nuclear and traditional biomass actually *decreased*. Only hydropower and gas match the $\sim 2\%$ p.a. assumption.

Energy substitution models capture these competitive dynamics [58, 154–156]: oil substituted wood during the early 20th century, and gas has substituted coal in heating and electricity generation [157]. Despite these limitations, the 2% p.a. growing-asymptote assumption yields tractable Bayesian inference and reasonable backtest performance for technologies whose growth closely tracks aggregate energy demand.

The growing market S-curve model We use model (60) with $\beta = \frac{2}{3}$, $\rho = 0.8$, and $g = 0.02$ as universal parameters. To estimate the remaining parameters, we adjust the transformation function that linearizes the S-curve to account for market growth:

$$f(Y_{i,t}; L_i, \beta) + k_i(t - t_{0,i}) = \frac{1}{\beta} \log \left(\frac{(L_i \exp(gt))^\beta}{Y_{i,t}^\beta} - 1 \right) + k_i(t - t_{0,i}).$$

The (t, t') element of the Jacobian matrix is

$$\left(\frac{\partial f(Y_{i,t})}{\partial Y_{i,t}} \right)_{t,t'} = \frac{\partial f(Y_{i,t})}{\partial Y_{i,t'}} = \delta_{t,t'} \frac{-(L_i \exp(gt))^\beta}{Y_{i,t}((L_i \exp(gt))^\beta - Y_{i,t}^\beta)} \quad (61)$$

and the log-determinant is

$$\log \det \left(\frac{\partial f(Y_{i,t})}{\partial Y_{i,t}} \right) = \sum_{t=1}^T \left[\beta g t + \beta \log(L_i) - \log \left(Y_{i,t}((L_i \exp(gt))^\beta - Y_{i,t}^\beta) \right) \right].$$

We then perform Bayesian inference as before using this adjusted transformation function and Jacobian.

Priors and assumptions. For models without asymptotic growth we require $L_i > \max_t(Y_{i,t})$ for $t \in [0, T]$, where T is the last time-point in the training data. As in Section 1.2, we place priors on scaled parameters $\tilde{L}_i = L_i / \max_t(Y_{i,t})$ and $\tilde{t}_{0,i} = t_{0,i} - t_1$, where t_1 is the first time-point in the training data. Figures 34b and 34e use the fixed-asymptote B-R model with the same universal prior as in the main analysis (Section 1.2).

In the growing-market model (60), the asymptote increases over time, so we require $L_i \exp(gt) > Y_{i,t}$ rather than $L_i > \max_t(Y_{i,t})$. In particular, $\tilde{L}_i$ can be less than 1 because market growth compensates. We therefore adjust the priors for $\tilde{L}_i$ and $\tilde{t}_{0,i}$, while keeping k_i and σ_i unchanged. Furthermore, we set a smaller lower bound on $\tilde{t}_{0,i}$ to ensure that the inflection point is not too late, which would lead to implausible forecasts. We adjust the scales of $\tilde{L}_i$ and $\tilde{t}_{0,i}$ (and the shape of $\tilde{t}_{0,i}$)

so that the prior predictive distribution remains reasonable. The exogenous growth assumption changes the relationship between these parameters and the data, particularly for L .

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2, 0, 5) & \text{(lower bound: } \tilde{L}_i \geq 0) \\ \log k_i \sim \log \Gamma(1, 0, 1) & \text{(unchanged)} \\ \log \tilde{t}_{0,i} \sim \log \Gamma(2, \max(5, T_{\max} - 6/\hat{k}_{\text{ssr},i}), 15) & \text{(lower bound: tech-specific)} \\ \log \sigma_i \sim \log \Gamma(1, 0, 1) & \text{(unchanged)} \end{cases} \quad (62)$$

The prior for $\tilde{L}_i$ now has a lower bound of 0 rather than 1.

For shares, L_i is bounded between 0 and 1. Hence, we keep all priors unchanged except for L_i , for which we use a uniform distribution between 0 and 1 as a non-informative prior. That is, the prior for shares is

$$\pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \mathcal{U}(0, 1), \\ t_{0,i} \sim \Gamma(3, \max(5, T_{\max} - 3/\hat{k}_{\text{ssr},i}), 30), \\ k_i \sim \Gamma(1, 0, 1), \\ \sigma_i \sim \Gamma(1, 0, 1). \end{cases} \quad (63)$$

Prior adjustment for near-saturated shares. For the shares B-R model applied to gas and hydropower, both technologies are already near or at saturation in shares at the $T - 30$ training-test split. In this regime, the data-informed lower bound $\ell_{t_{0,i}} = \max(5, T_{\max} - 3/\hat{k}_{\text{ssr},i})$ can inadvertently push the prior support for $t_{0,i}$ to lie entirely *after* the last observed data point, resulting in degenerate posteriors. We therefore set a smaller offset ($2/\hat{k}_{\text{ssr},i}$ instead of $3/\hat{k}_{\text{ssr},i}$) for these two cases, ensuring the prior places the inflection point within the observed data range.

4.2.2.2 Comparing the three approaches.

We show results for two technologies where we expect the growing market model to work well, hydropower and gas. We then discuss the intrinsically more difficult example of oil.

Technologies that grow asymptotically with the market: Hydropower and gas. Figure 34 compares backtest results for gas and hydropower using each of the three methods for a 30-year forecast horizon. In the test period, both technologies' useful energy shares change only modestly (gas increases slightly; hydropower is flat), so the shares model is a reasonable approximation. Results would naturally vary under different training-test splits or prior configurations. Figure 34a shows that the shares forecast produces reasonable results for this training-test split. The posterior places both technologies near saturation in shares (the posterior median of $\tilde{L}_i$ is close to the last observed value), which is consistent with the data. The forecasts appear slightly underdispersed: realized data points occasionally fall outside the 95% credible interval, likely because the B-R model's noise variance shrinks to zero as diffusion approaches the asymptote, understating residual uncertainty near maturity.

Figures 34b and 34e show that the fixed-asymptote B-R model applied to TWh is overconfident even with adaptive widening. Realized generation rises above the prediction interval because gas and hydropower keep growing with the market. Their useful energy shares stay flat while their absolute generation grows with overall demand, so generation exceeds the model's fixed ceiling. This is a structural failure of the model.

Figures 34c and 34f show that the growing market S-curve model captures future data points well, with the median forecast close to realized values. Both technologies grew at approximately 2% p.a. in this period, so the model is expected to perform well. Note that adjusting the priors for L_i and $t_{0,i}$ to account for the growing market size can lead to better performance.

These results inform our approach for solar and wind. For mature technologies with flat shares (gas, hydropower), the fixed-asymptote model fails in TWh because the growing market dominates. For solar and wind, however, we are in the opposite regime: both technologies are at early diffusion stages where total energy generation is still far below any plausible asymptote. In this regime, the fixed-asymptote model remains valid because the S-curve dynamics (not market growth) drive the forecast, and we use it for our main solar and wind projections in Section 5.2.

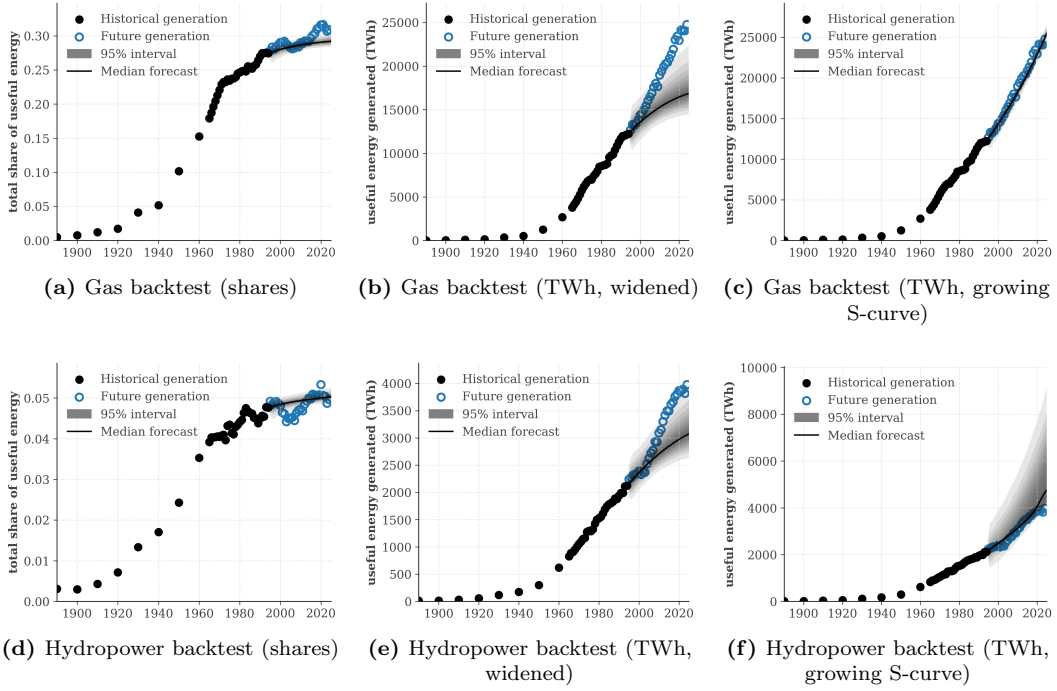


Figure 34: Example backtest comparison for three forecasting methods on gas and hydropower useful energy generation shares and absolute generation. (a-c) Backtest on gas shares, useful energy with no growth assumptions, and useful energy with growth assumptions. (d-f) Backtest on hydropower shares, useful energy with no growth assumptions, and useful energy with growth assumptions.

Predicting through an oil production shock. Figure 35 shows the distributional forecast and log-error for oil useful energy generation, using the fixed-asymptote B-R model, when training before (1870–1972), during (1870–1979), and after the oil production shock (1870–1985). The figure illustrates how training-period timing relative to structural breaks affects forecast accuracy. Training pre-shock (1870–1972) or post-shock (1870–1985) yields reasonable performance: the pre-shock forecast captures future deviations within 95% prediction intervals, while the post-shock forecast achieves log-errors below 15% (in value) with all points inside prediction intervals. Yet post-shock forecasts systematically underestimate future data points. All points lie in the upper part of the distribution, possibly because the energy market is growing.

Training through the oil production shock (1870–1979) produces systematic overestimation in the median, with log-errors increasing to $\sim 20\%$ as the model fails to capture subsequent drops. The general pattern is robust across prior configurations. Even in this adversarial case the realized data remain within the prediction interval. The structural break is a low-probability event under the model, not an impossible one. A purely data-driven model has no mechanism to anticipate such production shocks, and the adaptive widening assigns them non-negligible probability.

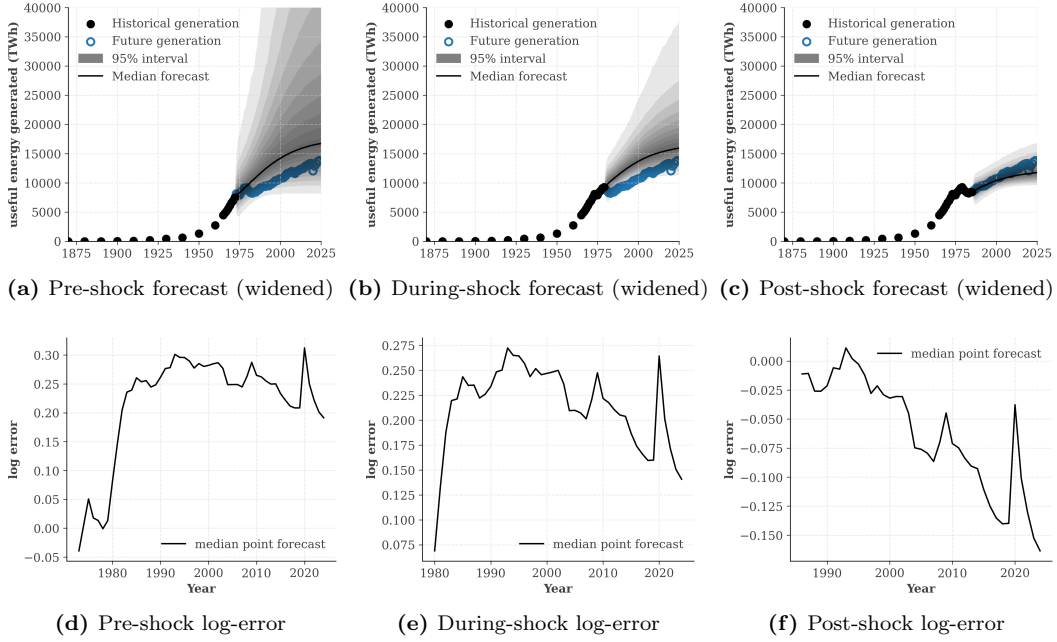


Figure 35: Backtest for oil useful energy generation using the fixed-asymptote B-R model, with distributional forecasts and log-errors for pre-shock, during-shock, and post-shock training sets. (a) Pre-shock forecast captures future deviations well. (b) During-shock forecast. Most data points fall within the lower part of the prediction interval, with a few below it. (c) Post-shock forecast captures realized data points, which lie in the upper part of the distribution. (d–f) Corresponding log-errors. The median forecast for all three scenarios is reasonable, with log-errors generally within $\pm 25\%$.

4.2.3 Exogenous disruptions: nuclear energy

Nuclear energy is an instructive case of a technology whose diffusion trajectory was repeatedly disrupted by exogenous safety events. Unlike the technologies in our main cross-section dataset, nuclear has experienced multiple abrupt deceleration episodes and is therefore a stress-test for any smooth S-curve model.

Backtest results. Figure 36 shows backtests using the fixed-asymptote B-R model with forecast origins in 1974 (before the major growth decline) and 1984 (after the Three Mile Island accident), with the same universal prior used for solar backtests in Section 5.1. Current nuclear useful energy generation is approximately 2,500 TWh. From the 1974 forecast origin (panel a), the model predicts a 2024 median of $\sim 11,000$ TWh — over $4\times$ the realized value — yet realized data still lie within the lower tails of the 95% prediction interval. This is expected: a data-driven model has no mechanism to anticipate the sequence of accidents and regulatory changes that followed, but the wide prior on L ensures the structural break remains within the distribution’s support.

By 1984, the data already show a clear deceleration following Three Mile Island, and the model is able to reasonably estimate the trend going forward (though it cannot anticipate Chernobyl in 1986 or Fukushima in 2011).

Historical context. Nuclear power was promoted as a peaceful use of atomic energy after WWII [158], and in the 1970s was projected to become a dominant energy source [159]. Today it accounts for approximately 3% of useful energy consumption. Figure 37 shows the historical useful energy share and log-growth. Three major accidents (Three Mile Island (1979) [160], Chernobyl (1986), and Fukushima (2011)) each produced sharp growth declines visible in Figure 37. No new US reactor has been ordered since 1978 [161], and public support shifted predominantly negative after these events [162, 163]. Nuclear electricity reached an approximate plateau around 1995 at $\approx 4.5\%$ of useful energy consumption; since 2011, log-growth has hovered near -2% . These repeated

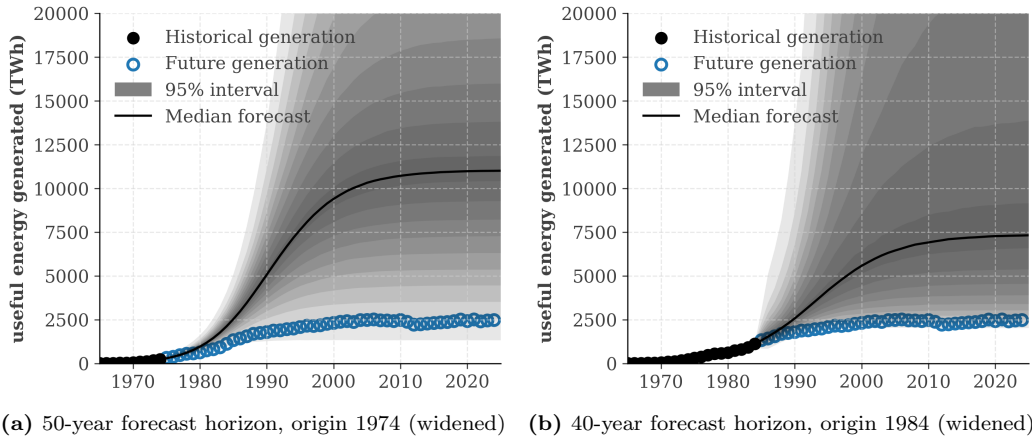


Figure 36: Backtest for nuclear energy useful energy generation using the fixed-asymptote B-R model. (a) 50-year horizon from 1974 (before the major accidents): the model overestimates by $\sim 4\times$, but realized data still lie within the tails of the prediction interval. (b) 40-year horizon from 1984 (after Three Mile Island; pre-Chernobyl): the post-TMI deceleration is already visible in the training data, and the model captures the trend reasonably well.

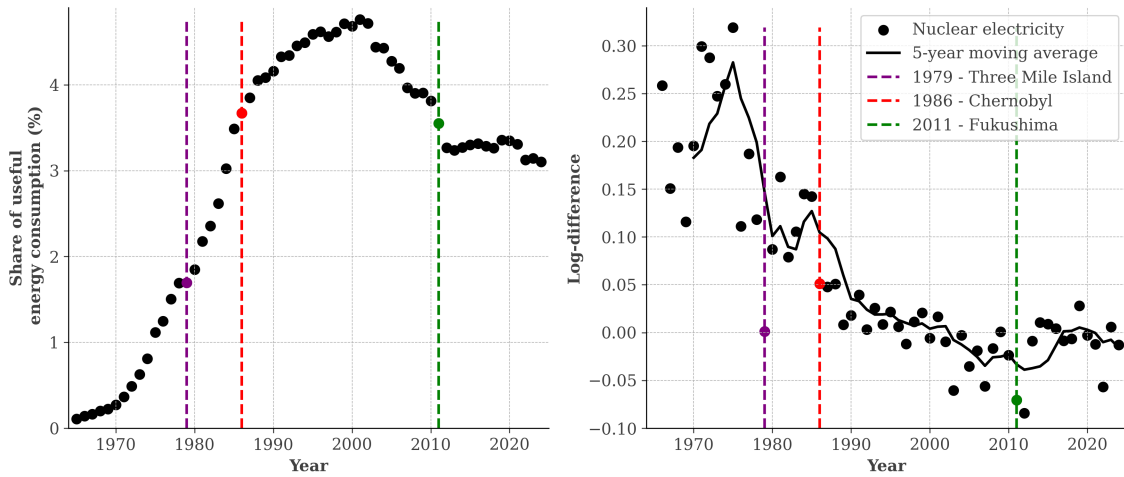


Figure 37: Useful energy consumption share and log-growth of nuclear electricity. Growth was approximately 30% p.a. during the rapid expansion phase. After Three Mile Island (1979), growth decreased; after Chernobyl (1986), log-growth fell from $\sim 15\%$ to $\sim 5\%$; after Fukushima (2011), log-growth turned negative before recovering to $\approx 2\%$ p.a.

disruptions explain the forecast failures above: each accident reset the diffusion trajectory in a way that a smooth S-curve cannot capture.

The nuclear case illustrates a general limitation. The S-curve model cannot always capture sudden structural breaks in adoption driven by exogenous events. Such disruptions are rare across the technology ensemble, although this is partly due to selection bias.

5

Application to solar and wind

5.1 Backtesting solar and wind time-series

To confirm the robustness of our forecast results on solar PV and wind (main paper Figure 4), we backtest on historical data for a 15-year forecast horizon. This horizon corresponds to energy-system planning time-scales [164, 165] and matches the forecast horizons used in Section 5.2.

Before interpreting the backtest results, two features of solar PV and wind are important context. Both technologies have followed near-exponential growth through 2024, with no sustained deviation from trend (Section 5.2, Figure 44). When training from a 2010 forecast origin, nearly all historical data therefore lies in the exponential regime. This provides limited information about eventual saturation and makes the prior more influential than in the true out-of-sample exercise. The 2010 origin is early for both technologies, and particularly so for solar: solar PV represented 0.04% of global useful energy in 2010 and wind 0.5%, compared to 2.4% and 2.8% respectively in 2024 (the forecast origin of the main-text predictions). The backtest thus constitutes a conservative stress test.

Backtests of useful energy generation and capacity use the universal prior with the model-direct adaptive widening calibrated in Section 4.1. This is the same method as our main-text forecasts. Backtests of market shares and of the asymptotic-growth model use no post-hoc widening.

Figure 38 shows the median and 95% prediction interval for solar and wind useful energy generation, shares, and capacity. We marginally underpredict solar in the median and overpredict wind, except for wind useful energy generation. Almost all realizations fall within the 95% prediction interval. For solar useful energy generation, all realizations fall within the interval, but they lie systematically in its upper part, and the 2024 point is only just inside. This reflects the median underprediction of solar growth. The remaining exceptions are the 2024 wind useful-energy-share point below the interval and wind capacity below it in 2017 to 2019.

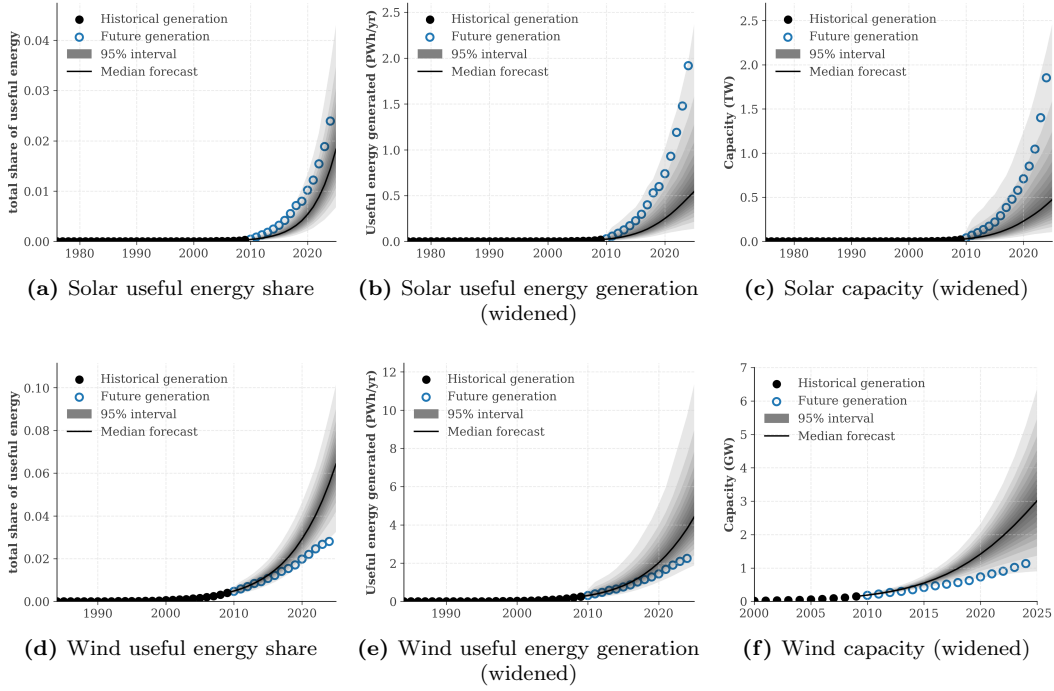


Figure 38: Backtest results for solar and wind useful energy generation, shares, and capacity. Black dots are historical data. Blue circles are future realizations relative to the forecast origin. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. Panels labelled (widened) use the adaptive widening of Section 4.1. (a)–(c) show backtests for solar PV, and (d)–(f) show backtests for wind.

Backtest using the S-curve with asymptotic growth. Figure 39 shows the same backtest for solar PV and wind useful energy for a 15-year horizon, but using the S-curve model with asymptotic market growth. If technologies are far from saturation, the forecast should be similar under both models. The results are indeed almost the same. When diffusion is far from saturation, $Y(t)/[L \exp(gt)] \ll 1$, so the S-curve trajectory is governed mostly by the exponential-growth phase of the B-R model. In this regime, the additional $\exp(gt)$ factor plays no large role in the near-term forecasted trajectory. The two model variants diverge materially only as the technology approaches its effective market ceiling.

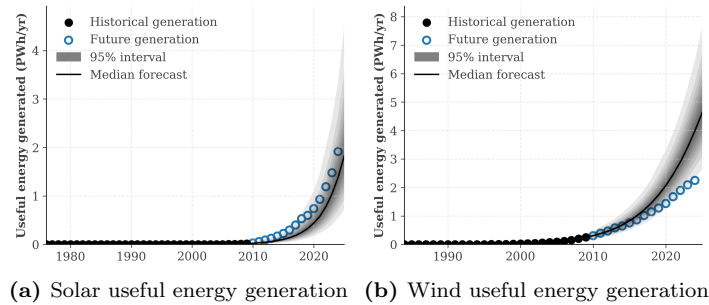


Figure 39: Backtest results for solar and wind useful energy generation using the asymptotic growth model. Black dots are historical data. Blue circles are future realizations. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. (a) shows the backtest for solar PV useful energy, and (b) shows the backtest for wind useful energy.

Additional solar and wind time-series Solar PV and wind time-series are also commonly reported as shares of total electricity generation. Figure 40 presents the backtest for electricity shares using data from OWID [102], providing an independent robustness check that the model generalises to this alternative diffusion measure. Prediction intervals are wider here because the electricity-shares series begin in 1985 rather than 1975, shortening the training window and reducing the information available about early diffusion dynamics. In-sample noise is also higher relative to the useful energy series, further contributing to the wider prediction intervals.

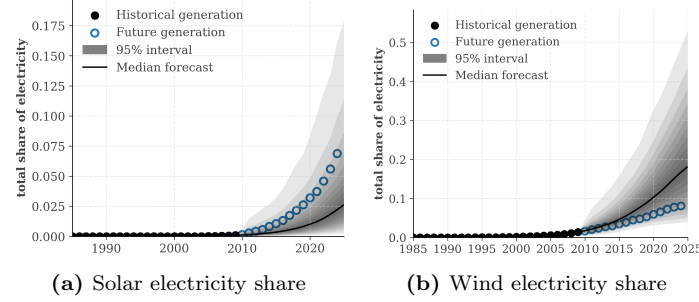


Figure 40: Backtest results for solar PV and wind electricity shares. Black dots are historical data. Blue circles are future realizations. The black line is the median forecast. The shaded region is the 95% prediction interval in 5% steps. (a) shows the backtest for solar PV electricity share, and (b) shows the backtest for wind electricity share.

5.2 Forecasting solar and wind useful energy

We present distributional forecasts for solar PV and wind useful energy generation, shares, and capacity using the Bertalanffy-Richards curve with technology-invariant parameters $\rho = 0.80$, $\beta = 2/3$, $\epsilon_{i,T} = 0$. The historical data sources are described in Section 2, with records through 2024. Farmer and Lafond [141] provide the method for the exponential-growth forecasts. We compare our forecasts to IEA NZE [166], IPCC AR6 [167], and Way et al. [49].²⁹

Throughout this section we keep two conceptually distinct issues separate. First, *installed capacity* (in TW) measures the maximum potential output of a technology, whereas the *energy actually delivered* depends additionally on the capacity factor and downstream conversion losses. Second, we distinguish *secondary energy* (raw electricity generated), *final energy* (electricity delivered to end-users, after transmission and distribution losses), and *useful energy* (the energy content of services provided, after end-use losses). We prefer useful energy because it best represents the energy services delivered by solar and wind and enables consistent comparison via shares. Where the IEA reports milestones in capacity (TW), we also forecast capacity directly. Section 5.2.2 below details the capacity-to-useful-energy conversion and shows that fitting capacity and converting yields similar forecast distributions to fitting useful energy directly.

5.2.1 IEA net-zero milestones for solar capacity

The IEA publishes Net-Zero Emissions (NZE) 2050 milestones for solar PV installed capacity [166, 170]: deployment targets at 2030, 2035, 2040, and 2050 consistent with limiting warming to 1.5°C. We address two questions. First, given an assumed asymptote L , when does the B-R model predict each milestone is reached, and how does that change with L ? Second, what does our distributional forecast for solar capacity imply for whether the milestones will be met?

5.2.1.1 Conditional milestone-year analysis

L is hard to identify from early data, and experts may disagree on its plausible range. We complement the distributional forecast with a conditional exercise. We treat L as a known input, fit $(\hat{k}, \hat{t}_0)$ by OLS on the GFP-linearised model (Eq. (13)), and compute the year at which diffusion reaches each IEA milestone as a function of L .

To estimate the year t^* such that diffusion $Y(t^*)$ equals milestone M_{t^*} , we invert the B-R model (Eq. (13)) to solve for t^* ,

$$t^*(L, M_{t^*}) = \hat{t}_0 - \frac{1}{\beta \hat{k}} \log \left(\left(\frac{L}{M_{t^*}} \right)^\beta - 1 \right) + \frac{\eta_{t^*}}{\beta \hat{k}}, \quad (64)$$

where $\eta_{t^*} = \hat{\sigma}(\epsilon_{t^*} + \rho \epsilon_{t^*-1})$. The noise term is additive in the time domain, so $\pm 1\sigma$ error bands are

$$t^*(L, M_{t^*}) = \hat{t}_0 - \frac{1}{\beta \hat{k}} \log \left(\left(\frac{L}{M_{t^*}} \right)^\beta - 1 \right) \pm \frac{\sigma_\eta}{\beta \hat{k}}, \quad (65)$$

where $\sigma_\eta = \hat{\sigma} \sqrt{1 + \rho^2}$. These OLS bands reflect only the stochastic noise η_{t^*} and do not account for posterior uncertainty in $\hat{k}$, $\hat{t}_0$, or $\hat{\sigma}$. They therefore understate the total uncertainty. We also perform a full Bayesian fit with the same universal priors as the unconditional forecasts, treating L as fixed. Figure 41 compares the OLS, the Bayesian, and the Bayesian widened 68% bands.

The two methods agree closely on the median. The OLS and Bayesian median milestone years differ by less than 0.1 year at every L for the IEA solar capacity milestones. The widening transform preserves the median by construction, so panel (c) shares the median of panel (b). The year each milestone is reached is therefore robust at the median. It does not depend on the method or on the widening, and only the width of the bands changes. This is what matters for the application to solar PV below, which reports median years (Table 34).

²⁹See [41, 55, 144, 168, 169] for alternative approaches.

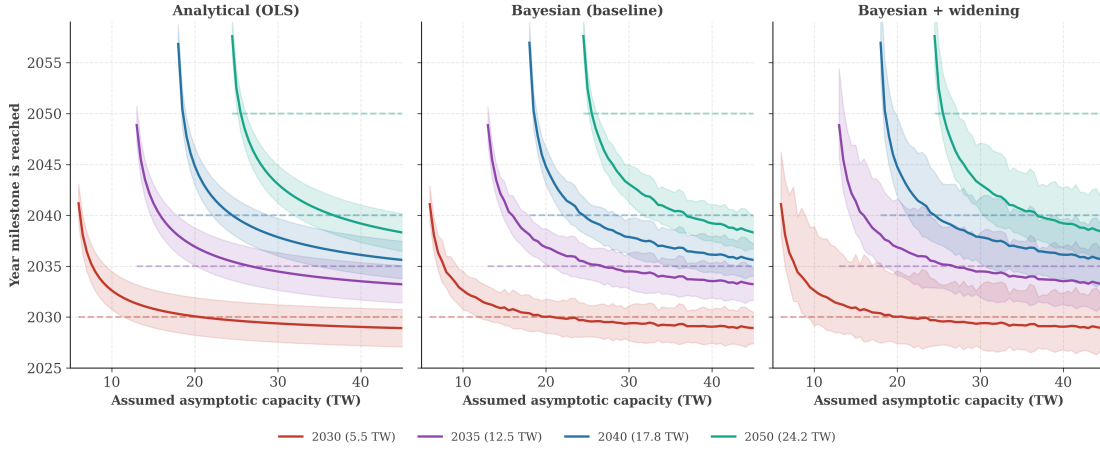


Figure 41: Year at which each IEA NZE milestone is reached as a function of the assumed asymptote L . **(a):** OLS analytical median with $\pm 1\sigma$ bands as per main paper Figure 4(b). **(b)** Bayesian posterior keeping L fixed, with posterior uncertainty in $\hat{k}$, $\hat{t}_0$, and $\hat{\sigma}$. **(c)** The same Bayesian posterior after the adaptive widening transform of Section 1.2.5. Dashed horizontal lines mark IEA target years. The widening is calibrated on the unconditional 120-technology forecasts, so panel (c) is an approximation in this fixed- L case.

Validation on the 120-technology dataset. We test the conditional method on the 120 technologies of Section 2. For each technology we take the asymptote L from the full-sample OLS fit on the GFP-linearised B-R model with free β (the same fit used for prior calibration elsewhere in the pipeline). Diffusion is defined as $d = Y/L$, and milestone M corresponds to $Y \geq M \cdot L$. For each series we then fit $(\hat{k}, \hat{t}_0)$ by OLS on the GFP-linearised model with fixed $\beta = 2/3$ using observations with $d < d_{\text{fit}}$. We predict the year diffusion reaches each milestone fraction M and compare to the realised year.

Table 33 reports the median prediction error (predicted minus realised milestone year) at two values of d_{fit} that bracket solar PV’s 2024 situation. The lower diffusion ($d_{\text{fit}} = 2.5\%$) is close to solar PV’s posterior-median current diffusion ($\approx 2.2\%$). The higher diffusion ($d_{\text{fit}} = 10\%$) matches the diffusion implied by the IEA NZE 2050 asymptote $L = 26$ TW. At the lower diffusion only 56 of the 120 historical analogues have enough early data to fit the method. Median errors grow from +4.7 years at $M = 0.25$ to +12.1 years at $M = 0.90$. At the higher diffusion the sample is larger ($N = 86$) and errors are smaller, from +1.1 years at $M = 0.25$ to +6.0 years at $M = 0.90$.

Table 33: Median prediction error (predicted minus realised milestone year, in years) on the 120-technology dataset. We take each technology’s asymptote L from the full-sample free- β OLS fit on the GFP-linearised B-R model. Diffusion is $d = Y/L$. We fit $(\hat{k}, \hat{t}_0)$ by OLS with fixed $\beta = 2/3$ on observations with $d < d_{\text{fit}}$, predict the year diffusion reaches each milestone fraction M , and compare to the realised year. N is the number of technologies with a successful fit and a realised milestone. The two rows bracket solar PV’s situation in 2024: $d_{\text{fit}} = 2.5\%$ approximates our posterior-median current diffusion ($\approx 2.1\%$), while $d_{\text{fit}} = 10\%$ reflects the diffusion implied by the IEA NZE 2050 asymptote $L = 26$ TW ($2.2/26 \approx 8.5\%$).

d_{fit}	N	$M=0.25$	$M=0.50$	$M=0.75$	$M=0.90$
2.5%	56	+4.7	+7.9	+12.0	+12.1
10%	86	+1.1	+1.6	+3.3	+6.0

How much can we trust the method? The signed median error is positive in every cell of Table 33. That is, the conditional method predicts milestones *later* than technologies realised them. The milestone-year predictions of Figure 41 are therefore conservative. Under the IEA NZE asymptote assumption ($L = 26$ TW, $d_{\text{fit}} = 10\%$), each IEA capacity target divided by L gives its milestone fraction M : 0.21 for 2030, 0.48 for 2035, 0.68 for 2040, and 0.93 for 2050. The corresponding median biases are approximately +1, +2, +3, and +6 years. Under our posterior on

L , where solar’s current diffusion is about 2.1%, the same milestones lie at smaller M but require a much longer extrapolation, and the corresponding median biases are approximately +5, +8, +12, and +12 years.

The median is robust, but are the prediction bands the right width? The widening transform was calibrated on the unconditional forecasts, so applying it to these fixed- L bands is an approximation. We test it by backtesting coverage on the 120 technologies, one milestone at a time. Take the 50% milestone, the year a technology first reaches half of its asymptote. For each technology the fixed- L Bayesian fit gives a predictive distribution over that year, and we know the year it was actually reached. The 68% coverage is the fraction of technologies whose actual year falls inside the central 68% of their own predictive distribution. A well-calibrated band has coverage equal to its nominal level. We compute this for each milestone (25%, 50%, 75%, 90%) and for several values of d_{fit} , the diffusion up to which we fit each technology. Without widening the bands are too narrow, just as in the unconditional forecasts. Widening moves coverage closer to nominal. Fitting up to 10% diffusion, for example, raises the 68% coverage of the 50% milestone from 33% to 55%, and its 90% coverage from 45% to 76%. Coverage is better when we fit up to a larger diffusion, where more data is available.

Application to solar PV. Table 34 reports the year reached for each milestone at three representative asymptotes, with Bayesian widened 68% bounds. At $L = 26$ TW the model meets the 2030 milestone (5.5 TW) and the 2035 milestone (12.5 TW) in their target years, and reaches the 2040 and 2050 milestones roughly one year early. At $L = 40$ TW the model reaches the final 2050 milestone (24.2 TW) in 2039, approximately a decade ahead of target.

Table 34: Estimated year of reaching IEA net-zero-by-2050 milestones for solar PV capacity, conditional on asymptote L . The median year t^* is computed analytically from the Bertalanffy-Richards curve with $\beta = \frac{2}{3}$. The lo/hi columns are the 68% band after applying the adaptive widening transform of Section 1.2.5 (Eq. 33), with w^* looked up at $d_{\text{eff}} = Y_{2024}/L$ from the same anchors as the unconditional capacity forecasts. A fixed- L Bayesian fit with the universal priors of Section 1.2 yields a posterior median for t^* within 0.1 year of the OLS value for every L tabulated. A dash indicates the milestone exceeds the assumed asymptote.

L (TW)	IEA milestone	Median year	68% lo	68% hi	Years early/late
10	2030 (5.5 TW)	2033	2029	2039	3 late
10	2035 (12.5 TW)	–	–	–	–
10	2040 (17.8 TW)	–	–	–	–
10	2050 (24.2 TW)	–	–	–	–
26	2030 (5.5 TW)	2030	2027	2034	on time
26	2035 (12.5 TW)	2035	2032	2039	on time
26	2040 (17.8 TW)	2039	2036	2043	1 early
26	2050 (24.2 TW)	2049	2046	2053	1 early
40	2030 (5.5 TW)	2029	2026	2033	1 early
40	2035 (12.5 TW)	2034	2031	2037	1 early
40	2040 (17.8 TW)	2036	2034	2040	4 early
40	2050 (24.2 TW)	2039	2037	2043	11 early

5.2.1.2 Distributional forecast for solar PV capacity

The conditional analysis above assumes L is known. Figure 42 shows the distributional forecast for solar PV capacity before and after the widening transform of Section 1.2.5. The 24.2 TW by 2050 IEA NZE target lies below the 15th percentile of the widened forecast, indicating that even under pessimistic diffusion scenarios solar PV is likely to meet all IEA net-zero capacity milestones. The unconditional median reaches approximately 78 TW in 2050, well above the highest IEA target.

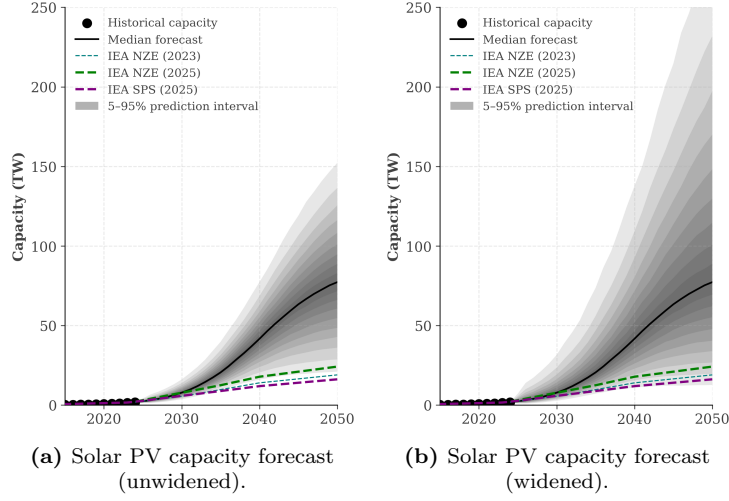


Figure 42: (a) Distributional B-R forecast for solar PV installed capacity under the universal prior, unwidened. (b) Same forecast after the adaptive widening transform of Section 1.2.5. Dashed lines show the IEA NZE and Stated Policies scenarios from the 2023 [166] and 2025 [170] editions. The most demanding IEA milestone (24.2 TW by 2050) lies below the 15th percentile of the widened forecast.

5.2.2 From capacity to useful energy

Solar useful energy forecasts can be built from two alternative data sources: (i) **useful energy (direct)**: historical useful energy actually delivered (in PWh/yr) from Way et al. [49]; or (ii) **installed capacity (converted)**: historical installed capacity (in TW) from IRENA [139], converted to useful energy via

$$E_{\text{useful}} = C \times 8760 \times \text{CF} \times 0.87 \times 0.9, \quad (66)$$

where C is installed capacity in TW and 8760 is the number of hours per year. The capacity factor CF (typically 0.10–0.25 for solar PV globally [171]) converts installed capacity to secondary energy (electricity generated). The factor 0.87 is the median secondary-to-final energy conversion ratio computed from 1,167 vetted IPCC AR6 scenarios [167], accounting for transmission and distribution losses estimated at 4–15% globally [172]. The factor 0.9 converts from final to useful energy, following Way et al. [49] and accounting for end-use losses [173].

Using $\text{CF} = 0.17$ (the historical median global value [171]), 78 TW (approximately the median solar capacity forecast in 2050) yields ≈ 91 PWh of useful energy per year via Eq. (66), broadly consistent with the median predicted useful energy of ≈ 85 PWh in main text Figure 4 in 2050. The conversion is sensitive to the assumed capacity factor. A slightly lower value within the historical range ($\text{CF} \approx 0.15$) gives a capacity-route median of ≈ 80 PWh, almost identical to the direct estimate. Even at $\text{CF} = 0.17$, the two routes produce very similar distributional forecasts, as quantified below (Fig. 43). The unwidened 5th–95th percentile capacity forecast at 2050 is ≈ 21 –152 TW, translating to ≈ 25 –177 PWh of useful energy and consistent with the corresponding unwidened useful-energy direct interval. After applying the adaptive widening transform of Section 1.2.5, these intervals widen to ≈ 13 –280 TW (≈ 15 –330 PWh) for the capacity route and ≈ 18 –290 PWh for the useful-energy direct route. The widening is applied identically to both routes, so the similarity between them is preserved (panel c of Fig. 43 below). The widened lower bound is conservative since the asymmetric transform widens the upper tail more strongly than the lower (Section 1.2.5).

To confirm that the choice of data source does not materially affect the forecast, we compare the two distributional forecasts at each year using the two-sample KS statistic (Eq. (59)) with $F_1 = \hat{F}_{\text{useful}}$ and $F_2 = \hat{F}_{\text{cap}}$, the empirical CDFs of the useful-energy and capacity-converted forecast samples respectively, with x in PWh/yr, and the overlap coefficient

$$\text{O} = \int \min(f_{\text{useful}}(x), f_{\text{cap}}(x)) dx, \quad (67)$$

where $D = 0$ (resp. $O = 1$) indicates identical distributions.

Figure 43 presents the comparison in three panels. On a log scale (panel a), the two historical series show a slight divergence up to around 2000, after which they align closely through 2024, confirming that the constant-CF conversion is a reasonable approximation over the observed period. The early divergence may reflect a capacity factor that differed from the fixed value of 0.17 during the early phase of solar PV deployment. The widened forecast distributions at 2030, 2040, and 2050 are broadly overlapping (panel b), with comparable medians and interquartile ranges across the forecast horizon, though the useful-energy (directly predicted) median is marginally lower in 2050 under a 0.17 capacity factor. The KS D statistic remains below 0.10 and the overlap above 0.9 throughout (panel c). The same conclusion holds for the unwidened forecasts. This agreement is not an artefact of the particular posterior draws used. Repeating the comparison with independent MCMC chains pooled for each route leaves the two distributions similarly matched across the forecast horizon. We conclude that forecasting capacity and converting to useful energy, or forecasting useful energy directly, generates similar distributional forecasts for solar PV.

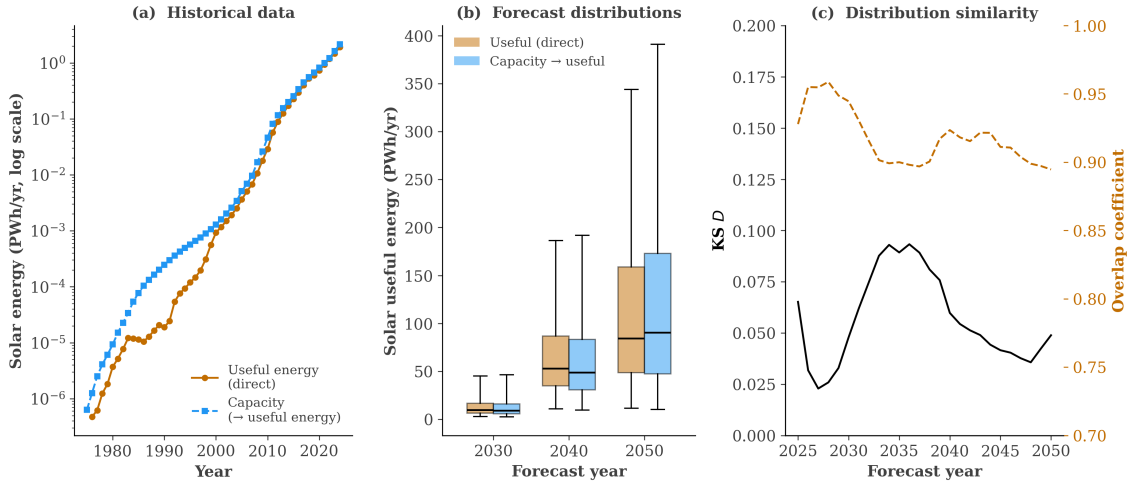


Figure 43: (a) Historical useful energy (orange) and capacity converted to useful energy via Eq. (66) (blue), on a log scale. The two series show a slight divergence up to around 2000, after which they align closely. (b) Side-by-side boxplots of the forecast distributions at 2030, 2040, and 2050, both widened via the adaptive sqrt-lower transform of Section 1.2.5 (box = IQR, whiskers = $1.5 \times$ IQR). (c) Kolmogorov–Smirnov D statistic (black, left axis) and overlap coefficient (orange dashed, right axis) from 2025 to 2050.

5.2.3 Forecasts for useful energy

5.2.3.1 Setup

We now turn to distributional forecasts for useful energy generation and useful energy shares.

Converting between the energy-accounting conventions defined above (secondary, final, useful) requires two steps. Sadovskaia et al. [172] estimate global transmission and distribution losses at 4–15% (secondary to final energy). Way et al. [49] assume end-use losses of 10% for solar and wind electricity (final to useful energy). We apply the two factors sequentially.

We compute the median secondary-to-final energy ratio across vetted IPCC AR6 scenarios [167] for 2030, 2040, and 2050, obtaining a conversion factor of 0.87. We then multiply by 0.9 to convert from final to useful energy, following Way et al. [49] and accounting for end-use losses [173]. To compare with IEA electricity generation scenarios, we multiply the reported final energy values by 0.9 to obtain useful energy estimates [166, 170]. The derivation of these conversion factors and a validation that this approach yields similar forecast distributions to fitting useful energy directly are provided in Section 5.2.2.

Before comparing our forecasts to scenarios, we examine how our model can predict departure from exponential behavior even though, for early solar PV and wind data, observed growth is close to exponential.

5.2.3.2 Analyzing the growth of solar and wind

This subsection examines how our B-R forecasts incorporate information about when solar and wind will decelerate from their current near-exponential growth. We establish formally that the near-exponential growth observed so far is exactly what the B-R model predicts for early-stage technologies. We then identify the point at which the median B-R forecast diverges from a pure exponential extrapolation. This is the earliest horizon at which the S-curve dynamics become detectable in the data.

So far, solar PV and wind have exhibited growth patterns very close to exponential [40, 49, 141]. The B-R model predicts this. For a technology in its early growth phase ($t \ll t_0$) we have $\exp(-\beta k(t - t_0)) \gg 1$, so the B-R model reduces to

$$y_t \equiv \log Y_t \approx (\log L - kt_0) + kt - \sigma(\epsilon_t + \rho\epsilon_{t-1}), \quad \epsilon_t \sim \mathcal{N}(0, 1), \quad (68)$$

i.e. a deterministic linear trend at rate k with MA(1) noise. Our ability to predict future deployment hinges upon deviations from this exponential trend. Following Farmer and Lafond [141], we analyze when the median B-R forecast departs significantly from such an exponential null model.

The deterministic trend implied by the model is $y_1 + \hat{\mu}(t - t_1)$, where $y_1 = \log Y(t_1)$ is the log of the first observation and $\hat{\mu} = \overline{\Delta y}$ is the sample mean of the log-differences. The OLS residuals from this trend are $\hat{r}_i = y(t_i) - \hat{y}(t_i)$, with residual standard error

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n \hat{r}_i^2}{n - 2}}, \quad (69)$$

where the denominator $n - 2$ reflects one degree of freedom lost to the estimated slope $\hat{\mu}$ and one to the constraint that the trend passes through (t_1, y_1) , which forces $\hat{r}_1 = 0$. The in-sample prediction interval is

$$\hat{y}(t) \pm 2\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(t - \bar{t})^2}{S_{xx}}}, \quad (70)$$

where $\bar{t}$ is the sample mean of the time index and $S_{xx} = \sum_i (t_i - \bar{t})^2$. The factor of 2 approximates the exact critical value $t_{n-2, 0.025}$; for our sample sizes ($n \geq 42$) the approximation error is less than 1%. This is the standard OLS prediction interval for a new observation at time t within the sample period [174]. We define the normalized deviation from exponential growth as

$$d(t) = \frac{\hat{r}_t}{\hat{\sigma}}. \quad (71)$$

A technology deviates significantly from exponential growth at time t when $|d(t)|$ exceeds the bound implied by Eq. 70.

For out-of-sample forecasting of a random walk with drift, we follow Farmer and Lafond [141] and use a τ -step-ahead forecast variance that accounts for the uncertainty in the estimated drift,

$$V(\tau) = \frac{\hat{s}^2}{1 + \hat{\phi}^2} \left[-2\hat{\phi} + \left(1 + \frac{2(m-1)}{m}\hat{\phi} + \hat{\phi}^2 \right) \left(\tau + \frac{\tau^2}{m} \right) \right], \quad (72)$$

where $\hat{s}^2 = \frac{1}{m-1} \sum_{t=2}^T (\Delta y_t - \hat{\mu})^2$ is the sample variance of the log-differences and $m = T - 1$ is the number of log-differences. The τ^2/m term is due to the fact that the drift is estimated. Then we use

$$\hat{y}(T + \tau) = y_T + \hat{\mu}\tau \pm 2\sqrt{V(\tau)} \quad (73)$$

as a 95% prediction interval for the τ -step-ahead forecast.

For solar PV, $\hat{\mu} = 0.32$ (37% annual growth, doubling every 2.2 years), $\hat{\phi} = 0.41$, $\hat{s}^2 = 0.035$, and $\hat{\sigma} = 0.50$. For wind, $\hat{\mu} = 0.19$ (21% annual growth, doubling every 3.6 years), $\hat{\phi} = 0.32$, $\hat{s}^2 = 0.009$, and $\hat{\sigma} = 0.45$. Figure 44 presents the results.

Figure 44a shows that both technologies have followed approximate exponential growth over the past several decades. Figure 44b shows that historical deviations exhibit a wave-like pattern consistent with the positively correlated errors of the MA(1) process, but no sustained systematic

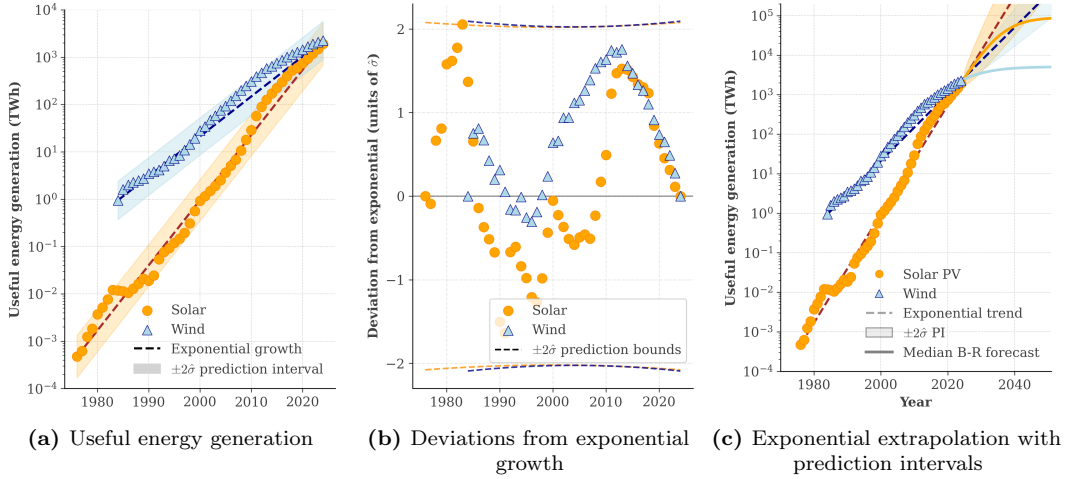


Figure 44: (a) Historical useful energy generation on a logarithmic scale, with exponential trend (dashed) and $\pm 2\sigma$ prediction bands (shaded). (b) Normalized deviations $d(t)$. The wave-like pattern is consistent with the positively correlated errors implied by the MA(1) process. (c) Exponential extrapolation with prediction intervals (shaded), exponential trend line (dashed), and median B-R forecasts (solid).

departure from the trend within the historical sample. This is exactly what the B-R model predicts for technologies still in their early exponential phase: finite- L S-curves look indistinguishable from pure exponential growth until diffusion rises well above a few percent of the asymptote. The finite estimate of L is therefore not identified from observed saturation in the data. It is instead informed by our empirical Bayes prior on L , which encodes evidence from the cross-technology distribution of asymptotes (Section 1.2). Figure 44c shows when the median B-R forecast is expected to begin departing from exponential growth.

5.2.3.3 Comparing forecasts to scenarios

Figure 45 shows forecasts for shares and generation of useful energy from solar PV and wind, compared to the IEA NZE scenario [166,170] and the Way et al. fast-transition trajectory [49]. The IEA NZE scenario for 1.5°C warming [166] and the Way et al. fast-transition trajectory [49] both lie below our median solar forecast. For wind, IEA NZE and Way et al. project higher deployment than our unconditional median forecast. Both sit on the upper end of the 90% prediction interval (5th–95th percentile). Our median forecast projects total combined solar and wind generation of approximately 90 PWh by 2050, compared with approximately 53 PWh in the IEA net-zero scenario [166,170]. Expressed as a share of total electricity generation, solar and wind reach 73% in the IEA net-zero scenario and 54% under Stated Policies. Our forecasts indicate that solar PV is likely to be the dominant source of renewable energy by 2050, while wind plays a much smaller role than in the IEA NZE scenario, albeit with a fat tail that includes higher wind deployment. These widened bounds are probably conservative for solar. In backtesting, fewer than 5% of technologies fall below their 5th-percentile forecast at early diffusion, so the lower bound is cautious (the Q-Q curve in main paper Figure 3 sits below the diagonal at low percentiles). Because the widening is calibrated across technologies, we are cognizant that it may also over-widen the upper tail for an early-stage technology like solar.

Figure 46 shows the same forecast without the widening transform. That is the unwidened, raw forecast of the one shown above and in Figure 4(d–f) of the main paper. The median is unchanged and only the prediction intervals narrow. For the combined solar and wind forecast the 90% prediction interval (5–95%) in 2050 is 32–169 PWh/yr compared to 22–305 PWh/yr widened. The qualitative result is robust to the widening: solar grows well above its current level and above the IEA and AR6 scenarios, while wind stays small. Recalibrating the widening on the 154-technology set leaves the forecast essentially unchanged. The 2050 solar interval becomes approximately 19–280 PWh/yr, against approximately 18–290 PWh/yr under the 120-technology

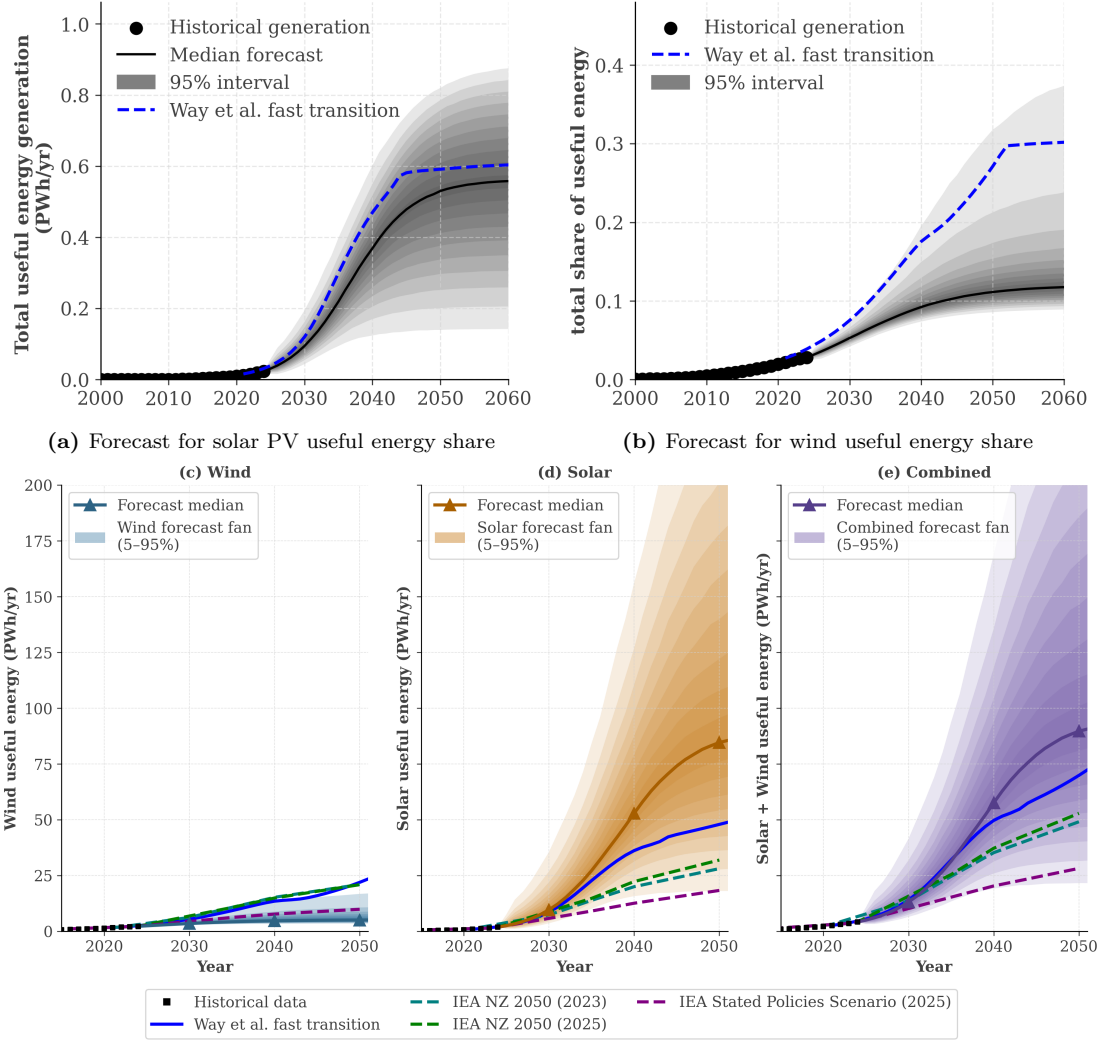


Figure 45: Same Bertalanffy-Richards distributional forecast as main-text Figure 4(d-f) for (a) solar useful-energy share and (b) wind useful-energy share. (c) - (e) are identical forecasts to main-text Figure 4(d-f), here compared to the IEA NZE scenario [166, 170] (green dashed; final electricity multiplied by 0.9 to obtain useful energy following Way et al. [49]) and the Way et al. fast-transition trajectory [49], rather than the AR6 ensemble.

anchors, and the wind interval 4–15 PWh/yr against 4–17 PWh/yr.

By 2050, our median forecast implies that solar will have an over 5-fold higher useful energy share than wind (over 50% versus $\approx 10\%$). The implied ratio in absolute useful energy generation is substantially larger. The median solar forecast reaches approximately 85 PWh/yr versus approximately 5 PWh/yr for wind, a roughly 17-fold difference. The share and generation ratios diverge because total useful energy demand is not constant. When both technologies are measured relative to the same growing total, the gap in forecast seems to be smaller. This motivates the exogenous growth model presented below, which explicitly accounts for demand growth and yields shares and absolute generation forecasts that are more consistent with each other. Combining the two forecasts makes the implied demand path explicit: dividing the median solar generation (≈ 85 PWh/yr) by the median solar share ($\approx 55\%$) implies total useful energy demand of roughly 155 PWh/yr by 2050, growth of about 2.4% per year from today’s 80–85 PWh/yr. This is close to the historical rate of about 2% per year [49] used for the demand-consistency check in the main text, under which demand reaches roughly 140 PWh/yr and our median shares forecasts (Figure 45) as well as the median useful energy forecasts of main text Figure 4(d-f) have solar and wind supplying

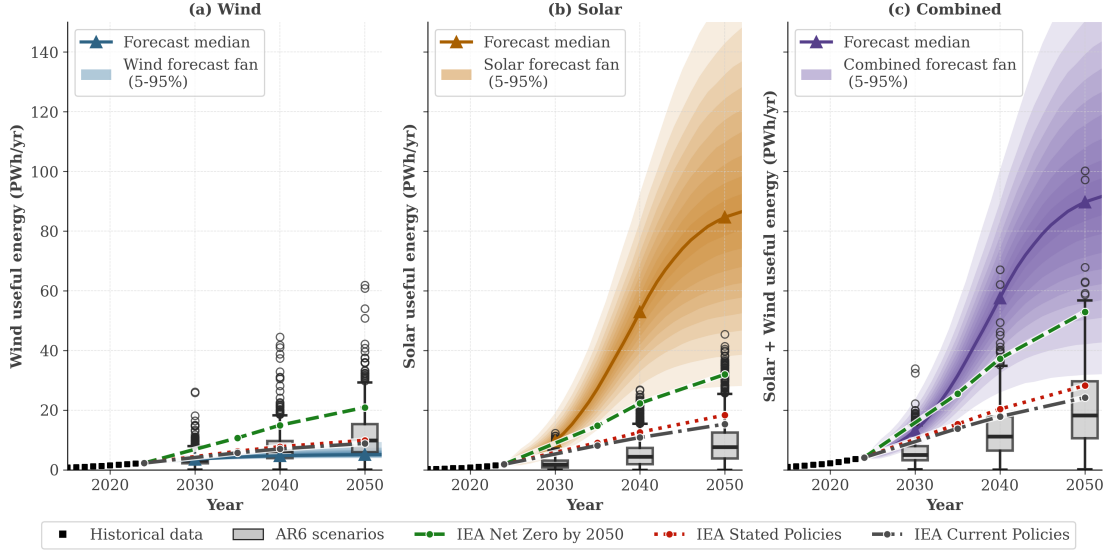


Figure 46: Unwidened predictive distribution of solar, wind, and combined. The same B-R distributional forecast as main-text Figure 4(d–f), for (a) wind, (b) solar PV, and (c) the two combined useful-energy generation. The medians match main-text Figure 4(d–f). Dashed lines show the IEA World Energy Outlook 2025 scenarios (Net Zero Emissions by 2050, Stated Policies, Current Policies) [170], converted to useful energy by multiplying by 0.9 following Way et al. [49].

about two-thirds of it in 2050.

Comparison with the NGFS scenarios. Figure 47 compares our forecast for solar and wind useful energy generation, as shown in main text Figure 4(d–f), with the NGFS scenarios [175], produced for central banks and financial supervisors. Our combined solar and wind median lies above all NGFS pathways. The most ambitious NGFS pathway reaches 70 PWh/yr in 2050, against our median of nearly 90 PWh/yr. Similar to other ensembles, the more ambitious NGFS pathways place wind above the upper bound of our wind forecast while placing solar well below our median.

Inferred percentage of adoption. The posterior distribution for the asymptote implies a distribution over the current diffusion level. For a technology still growing, $\max_t Y_{i,t} = Y_{i,T}$, so the normalized asymptote is $\tilde{L}_i = L_i/Y_{i,T}$. The implied current adoption percentage is therefore

$$D_{i,T} = \frac{Y_{i,T}}{L_i} \times 100 = \frac{100}{\tilde{L}_i} \%. \quad (74)$$

Because $\tilde{L}_i$ has a posterior distribution, so does $D_{i,T}$. We apply the asymmetric widening transform of Section 1.2.5 to the posterior on L before computing the implied diffusion. The widening transform preserves the median by construction and stretches the tails. The widened posterior is the one used in the main paper because it is calibrated to match the empirical PIT distribution observed in backtesting (Section 4.1). Figure 48 reports the implied diffusion distribution. The top row (a, b) shows the un-widened and the bottom row (c, d) shows the widened implied diffusion of the 2024 generation. The median implied diffusion is identical in both rows by construction. The 5–95% interval is wider in the bottom row. The median implied diffusion is approximately 2.1% for solar PV and 43.4% for wind. The solar posterior places solar firmly in early diffusion. The wind posterior places wind at a more advanced stage but with greater uncertainty about how advanced. Re-fitting the wind-time-series from forecast origins one to five years earlier yields a posterior on L that narrows toward the same low value, indicating the data is becoming increasingly informative about the asymptote.

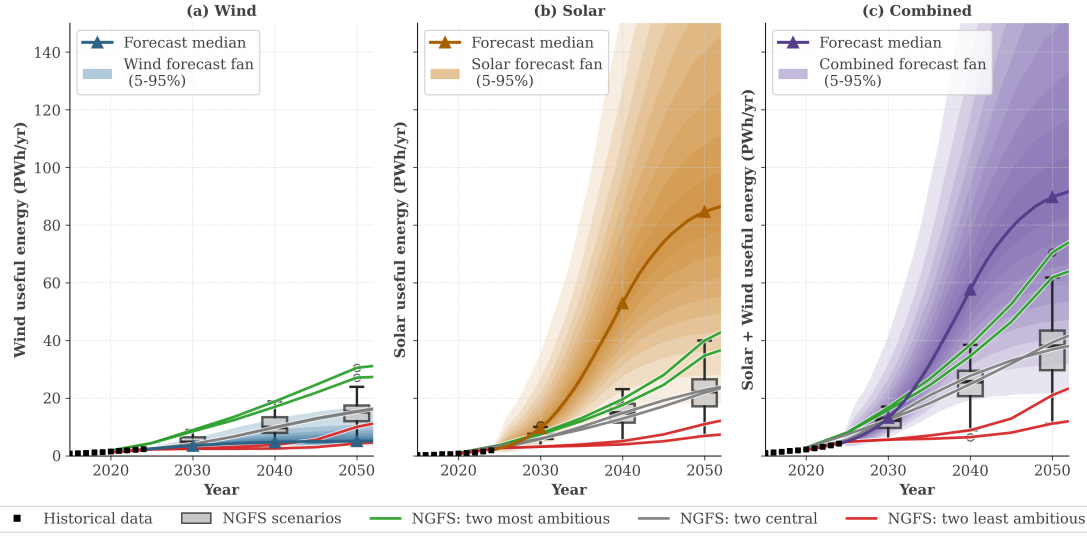


Figure 47: Our distributional forecast (shaded fan, 5–95% prediction interval) compared with the NGFS Phase 5 ensemble, summarised as box-and-whisker plots at each decade (box = interquartile range, whiskers = $1.5 \times \text{IQR}$, outliers as circles). Coloured lines show six representative pathways: the two most ambitious (green), two central (grey) and two least ambitious (red) by 2050 value. **(a)** wind, **(b)** solar PV, **(c)** the two combined. Same layout as main text Figure 4(d–f).

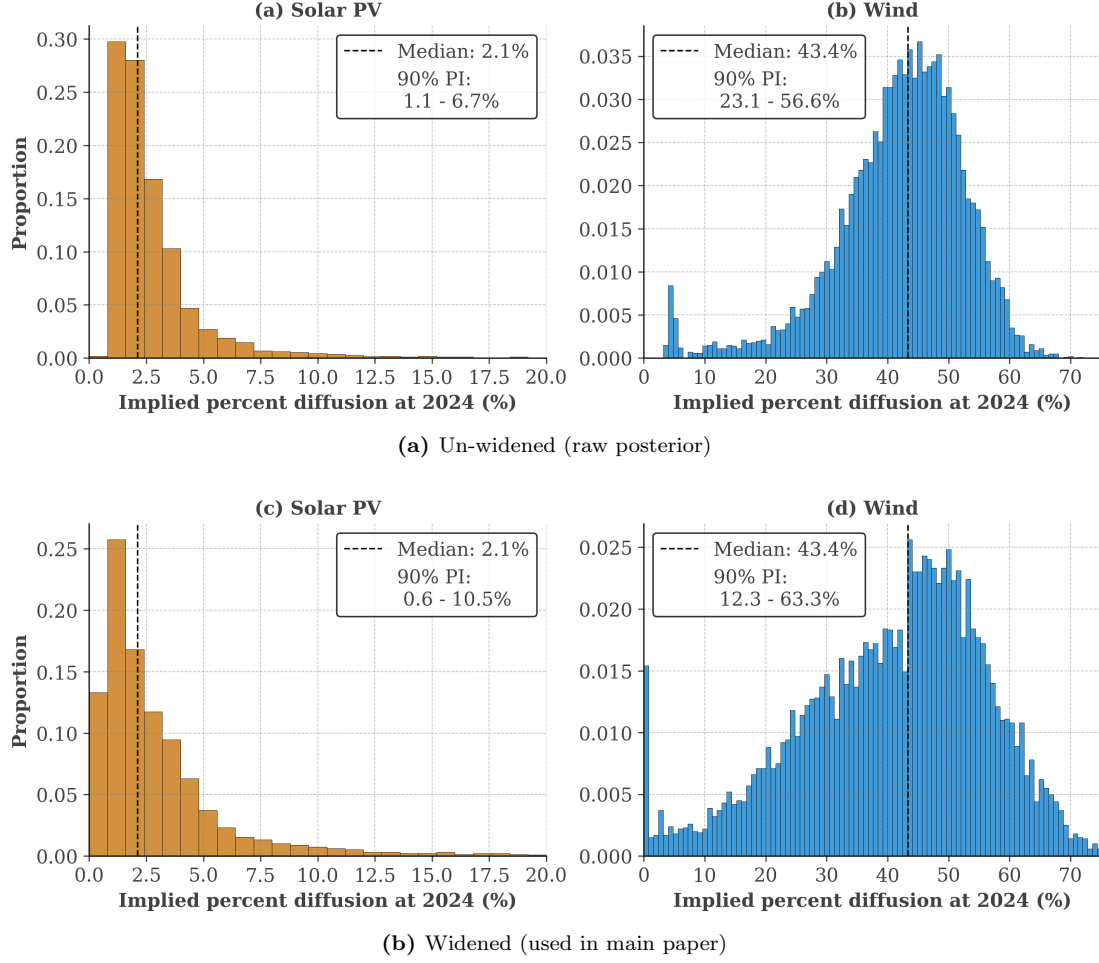


Figure 48: Posterior distribution of the current diffusion level $D_{i,T} = 100/\tilde{L}_i$ for solar PV and wind useful energy generation. **Top row (un-widened):** raw posterior. **Bottom row (widened, used in main paper):** after applying the asymmetric widening transform of Section 1.2.5. Vertical dashed lines mark the median implied diffusion (approximately 2.1% for solar PV and 43.4% for wind).

Exogenous demand growth. Figure 49 shows forecasts as in main-text Figure 4, but assuming 2% annual exogenous demand growth, using the methodology of Section 4.2. Solar reaches approximately 60 PWh/yr by 2050, compared to approximately 85 PWh/yr under the B-R model (without exogenous growth). This may seem counter-intuitive, but reflects a change in the model’s parameters and a corresponding change in the L prior. To accommodate the exponential growth term, we use a lower L prior (since some growth is now absorbed by the exogenous term). This produces a lower median forecast in Solar’s case. That is, the result is sensitive to the prior choice. Nonetheless, the qualitative implications of the forecast remain similar: the median solar forecast lies well above the IEA NZE scenario, and the AR6 scenarios lie well below the median. The main difference is that, under the exogenous growth model, the median forecast up to 2040 lies slightly below the Way et al. [49] fast transition. Under the B-R model (without exogenous growth), the median lies marginally above it. For wind, the median forecast lies below the IEA NZE scenario under both B-R variants, but the exogenous growth model indicates higher upper bounds of the distributional forecast for wind. In both cases, the forecasts indicate solar is still in the early diffusion stage, while wind is in a more advanced stage of diffusion. We extend these forecasts to 2060 rather than 2050, because the exogenous demand growth keeps the trajectories rising as the technologies approach saturation, and the extra decade shows how they separate from the scenarios. The widening transform is not calibrated for this model, so we show the un-widened forecast. Applying it would only widen the bounds and would not change the conclusion. Under the exogenous model the median wind forecast tracks the median AR6 scenario, whereas in the main paper it lies below.

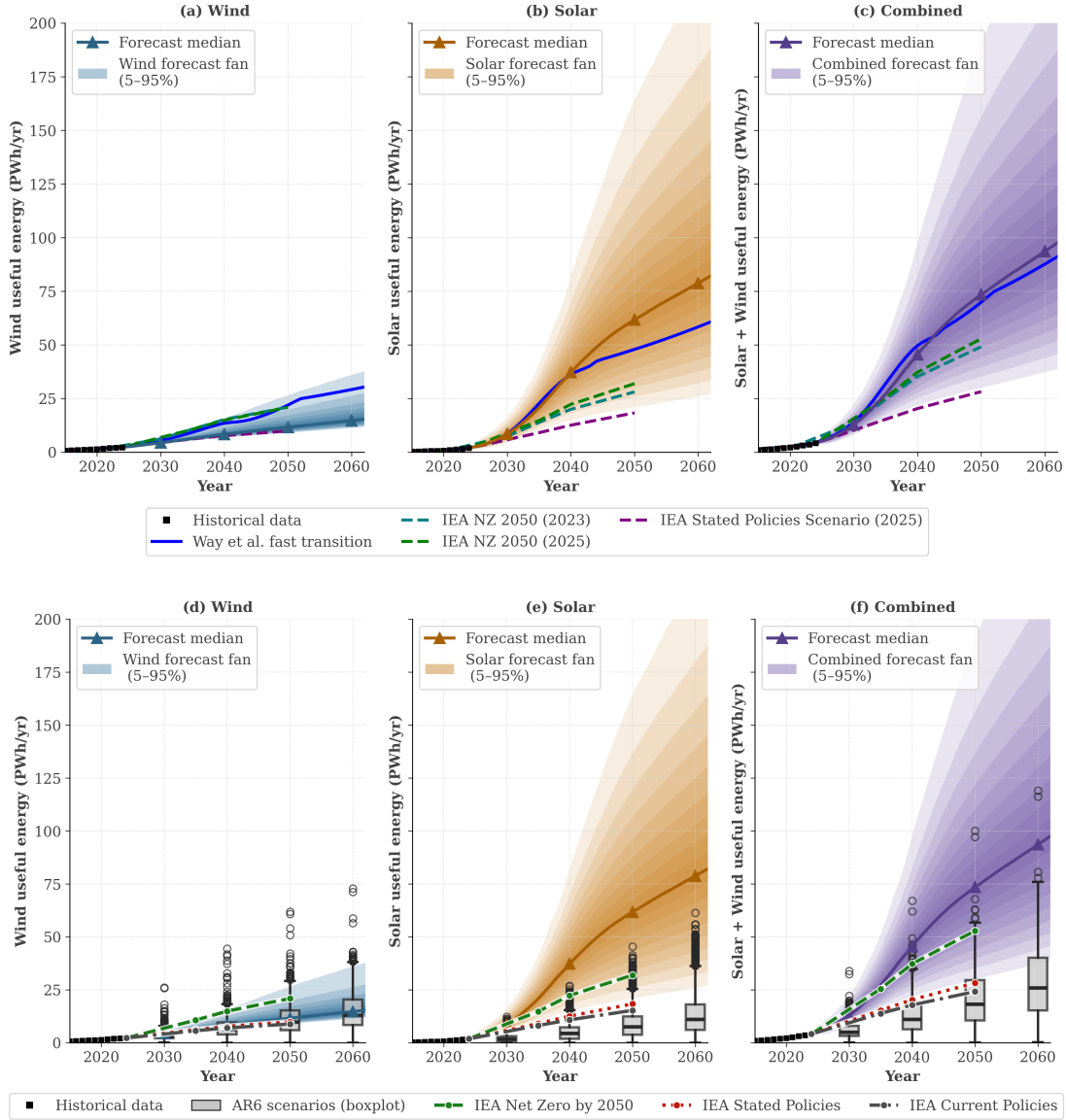


Figure 49: Un-widened forecasts assuming 2% annual exogenous demand growth, for solar PV, wind, and the two combined, shown to 2060. **Top row:** our forecast against the Way et al. [49] fast-transition trajectory and the IEA NZE scenarios [166, 170]. **Bottom row:** the same forecast against the IPCC AR6 scenarios [167], in the layout of main paper Figure 4(d-f), with wind, solar, and the two combined. The shaded band is the 5–95% prediction interval. Dashed lines show the IEA World Energy Outlook 2025 scenarios (Net Zero Emissions by 2050, Stated Policies, Current Policies) [170].

Comparison to literature capacity projections. Several studies project future solar PV capacity: 21.7 TW by 2050 and 84.5 TW by 2100 [176], and 133 TW by 2100 [177]. Assuming a capacity factor of 0.17 and applying the useful-energy conversion factor of $0.9 \times 0.87 \approx 0.78$ as per above, these capacity projections correspond to useful energy levels of approximately 25 PWh, 98 PWh, and 154 PWh, respectively, all falling within the prediction interval for the forecast we have shown for both solar PV useful energy and capacity.

Achieving these higher capacity levels would require continued growth in energy and/or electricity demand potentially beyond the baseline. Surges in electricity demand have occurred, driven by AI and data centres among other factors. Electricity demand grew 4.4% in 2024 and 3.3% in 2025 [178]. Advanced technologies such as space-based solar power [179, 180] or perovskite tandem solar cells [176] could deliver improvements to capacity factors and/or the useful energy conversion

factor, which would allow for higher useful energy generation for a given capacity. However, these technologies are currently in early stages of development, and it is uncertain whether they will be deployed at scale by 2050.

5.2.4 A pessimistic prior

Assume the position of a skeptic who believes that solar PV is likely not to grow much further. For example, they may believe that solar PV is already in a phase of decelerating growth. In such a case one may opt for more pessimistic priors for the asymptote L and the inflection point parameter t_0 . Here we present forecasts for solar PV and wind useful energy assuming the prior

$$\text{backtesting prior: } \pi(\theta_i) = \begin{cases} \tilde{L}_i \sim \Gamma(2, 1, 5), \\ t_{0,i} \sim \Gamma(2, \max(5, T_{\max} - 3/\hat{k}_{\text{ssr},i}), 10), \\ k_i \sim \Gamma(1, 0, 1), \\ \sigma_i \sim \Gamma(1, 0, 1). \end{cases} \quad (75)$$

In this prior, the mode of the asymptote L is 6 times the current level (approximately 11.5 PWh/yr for solar). For solar, where $\hat{k}_{\text{ssr}} \approx 0.3$ and $T_{\max} = 50$, the mode of the inflection point $t_0 = \max(5, 50 - 3/0.3) + 10 \approx 50$ (equivalent to t_0 being at the last observation).

Figure 50 shows the widened forecasts under this pessimistic prior. The widening transform of Section 1.2.5 was calibrated under the universal prior of Eq. (32) and is not re-calibrated here, hence should be treated as a conservative approximation. The median solar forecast in 2050 has decreased to approximately 63 PWh/yr, from approximately 85 PWh/yr under the prior of Eq. (32). Even under a pessimistic prior, the data is informative enough to converge to a posterior that still predicts substantial growth in solar PV useful energy generation. That is, the median posterior asymptote is 66 PWh/yr, compared to the prior median of 18 PWh/yr (and prior mode of 11 – 12 PWh/yr). The median wind forecast in 2050 has remained at approximately 5 PWh/yr compared to the prior used in the main paper, and its posterior is now very narrow. The width of the distributional forecasts has decreased, reflecting the more concentrated prior. Even under this pessimistic prior the conclusions are qualitatively the same as in the main paper. Solar still grows far beyond its current level, while wind stays small.

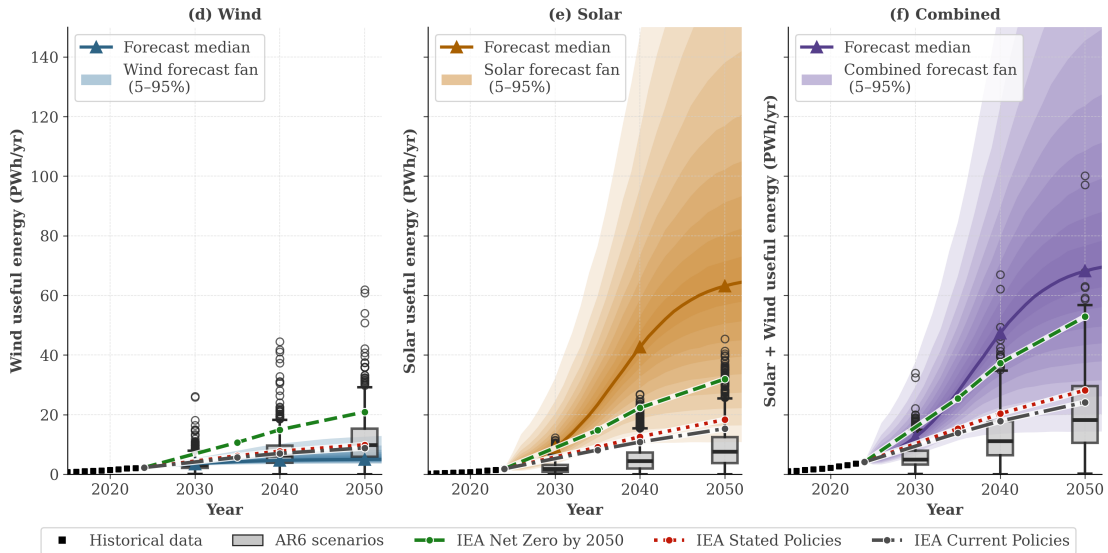


Figure 50: Bertalanffy-Richards forecasts with 90% prediction intervals in 5% steps (shaded), under a pessimistic prior for the asymptote L and the inflection point t_0 . Dashed lines show the IEA World Energy Outlook 2025 scenarios (Net Zero Emissions by 2050, Stated Policies, Current Policies) [170]. The median solar forecast in 2050 is approximately 63 PWh/yr. The median wind forecast in 2050 is approximately 5 PWh/yr.

Figure 51 compares prior predictive distributions to the posterior predictive distribution for solar PV and wind useful energy generation. The data has shifted the solar PV distribution towards much higher values, and the posterior median is substantially higher than the prior median in 2050. The data has shifted the wind distribution to much lower values. This illustrates how the data can overcome a pessimistic or optimistic prior, yielding forecasts of substantial growth (or limited growth, when the optimistic prior is incorrect). Despite this, it also indicates that the prior still has an influence on the forecast, as the median forecast has decreased compared to the prior of Eq. (32).

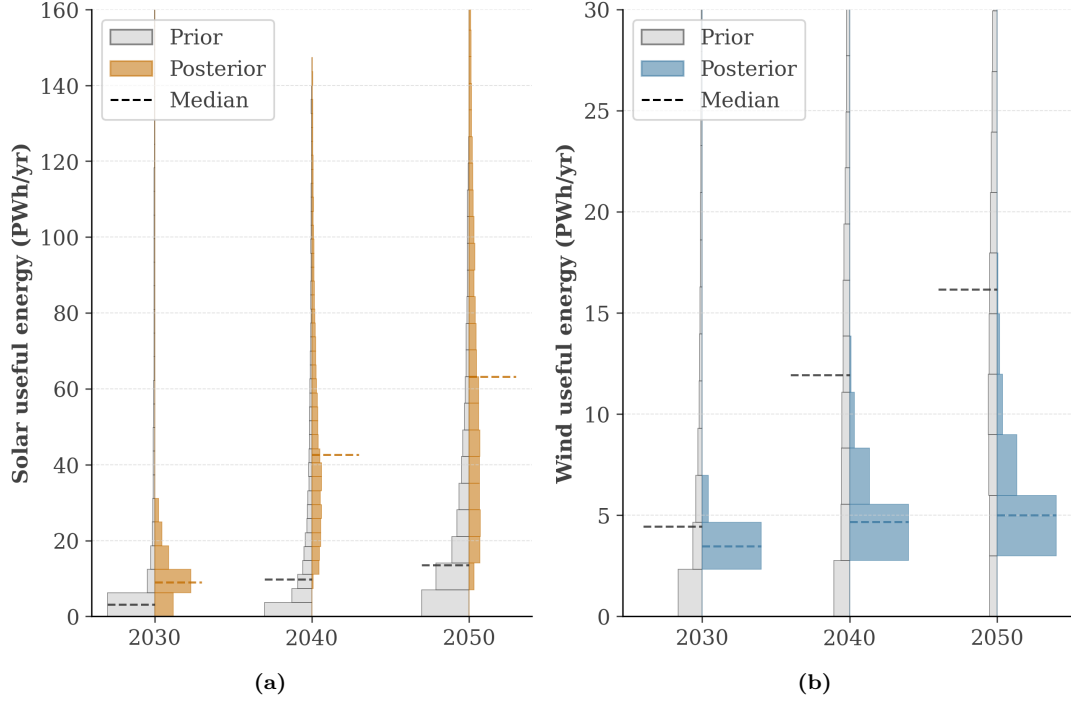


Figure 51: The figure compares the prior predictive distribution (grey) to the posterior predictive distribution under a pessimistic prior for the asymptote L and the inflection point t_0 . **(a)** Solar PV (orange). **(b)** Wind (blue) useful energy generation. The data has shifted the solar PV distribution towards much higher values, and the posterior median is substantially higher than the prior median in 2050. The data has shifted the wind distribution to much lower values.

Across these three variations, the universal prior, exogenous demand growth, and a pessimistic prior, our qualitative conclusion is unchanged.

References

- [1] Diego Comin and Marti Mestieri. Technology diffusion: Measurement, causes, and consequences. In *Handbook of economic growth*, volume 2, pages 565–622. Elsevier, 2014.
- [2] Arnulf Grubler. Energy transitions research: Insights and cautionary tales. *Energy policy*, 50:8–16, 2012.
- [3] Benjamin K. Sovacool. How long will it take? conceptualizing the temporal dynamics of energy transitions. *Energy Research & Social Science*, 13:202–215, 2016.
- [4] Roger Fouquet. The slow search for solutions: Lessons from historical energy transitions by sector and service. *Energy Policy*, 38(11):6586–6596, 2010. Historical analysis of energy transitions showing long timescales and sector-specific dynamics.
- [5] Zvi Griliches. Hybrid corn: An exploration in the economics of technological change. *Econometrica, Journal of the Econometric Society*, pages 501–522, 1957.
- [6] Everett M Rogers. *Diffusion of Innovations*. Free Press, New York, 5th edition, 2003.
- [7] Arnulf Grübler. Diffusion: long-term patterns and discontinuities. In *Diffusion of technologies and social behavior*, pages 451–482. Springer, 1991.
- [8] Arnulf Grübler, Nebojša Nakićenović, and David G Victor. Dynamics of energy technologies and global change. *Energy policy*, 27(5):247–280, 1999.
- [9] Charlie Wilson. *Meta-analysis of unit and industry level scaling dynamics in energy technologies and climate change mitigation scenarios*. JSTOR, 2009.
- [10] Praveen Asthana. Jumping the technology s-curve. *Ieee Spectrum*, 32(6):49–54, 1995.
- [11] K Usha Rao and VVN Kishore. A review of technology diffusion models with special reference to renewable energy technologies. *Renewable and sustainable energy reviews*, 14(3):1070–1078, 2010.
- [12] Nigel Meade. The use of growth curves in forecasting market development—a review and appraisal. *Journal of Forecasting*, 3(4):429–451, 1984.
- [13] Vijay Mahajan, Eitan Muller, and Frank M Bass. New product diffusion models in marketing: A review and directions for research. *Journal of marketing*, 54(1):1–26, 1990.
- [14] Vijay Mahajan, Eitan Muller, and Frank M Bass. New-product diffusion models. *Handbooks in operations research and management science*, 5:349–408, 1993.
- [15] Rui Baptista. Do innovations diffuse faster within geographical clusters? *International Journal of industrial organization*, 18(3):515–535, 2000.
- [16] Bruce Robinson and Chet Lakhani. Dynamic price models for new-product planning. *Management science*, 21(10):1113–1122, 1975.
- [17] Malcolm E Turner Jr, Edwin L Bradley Jr, Katherine A Kirk, and Kenneth M Pruitt. A theory of growth. *Mathematical Biosciences*, 29(3-4):367–373, 1976.
- [18] Malcolm E Turner Jr, Brent A Blumenstein, and Jeanne L Sebaugh. 265 note: A generalization of the logistic law of growth. *Biometrics*, pages 577–580, 1969.
- [19] JA Nelder. The fitting of a generalization of the logistic curve. *Biometrics*, 17(1):89–110, 1961.
- [20] Anastasios Tsoularis and James Wallace. Analysis of logistic growth models. *Mathematical biosciences*, 179(1):21–55, 2002.

- [21] Verhulst Pierre-François. Notice sur la loi que la population poursuit dans son accroissement. *Corresp. Math. Phys.*, 10:113–121, 1838.
- [22] Masayuki Kakehashi and Shoko Kawano. Fundamentals of mathematical models of infectious diseases and their application to data analyses. In *Handbook of statistics*, volume 36, pages 3–45. Elsevier, 2017.
- [23] Angela R McLean. Infectious disease modeling. In *Infectious diseases*, pages 9–20. Springer, 2023.
- [24] John Kruschke. *Doing Bayesian data analysis*. Academic Press, 2014.
- [25] Irina Kareva and Georgy Karev. *Modeling Evolution of Heterogeneous Populations: Theory and Applications*. Academic Press, 2019.
- [26] Dmitry Kucharavy and Roland De Guio. Application of s-shaped curves. *Procedia Engineering*, 9:559–572, 2011.
- [27] Arnulf Grubler. *The rise and fall of infrastructures: dynamics of evolution and technological change in transport*. Physica-Verlag, 1990.
- [28] Benjamin Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies: in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, 115:513–585, 1825.
- [29] Kathleen MC Tjørve and Even Tjørve. The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family. *PloS one*, 12(6):e0178691, 2017.
- [30] CE Timothy Paine, Toby R Marthens, Deborah R Vogt, Drew Purves, Mark Rees, Andy Hector, and Lindsay A Turnbull. How to fit nonlinear plant growth models and calculate growth rates: an update for ecologists. *Methods in Ecology and Evolution*, 3(2):245–256, 2012.
- [31] William S Hoar, David J Randall, and Edward M Donaldson. *Fish physiology*. Academic Press, 1983.
- [32] Anna Kane Laird. Dynamics of tumour growth. *British journal of cancer*, 18(3):490, 1964.
- [33] Marcel H Zwietering, Ida Jongenburger, Frank M Rombouts, and Kjaem Van’t Riet. Modeling of the bacterial growth curve. *Applied and environmental microbiology*, 56(6):1875–1881, 1990.
- [34] MIE Halmi, MS Shukor, WLW Johari, and MY Shukor. Evaluation of several mathematical models for fitting the growth of the algae dunaliella tertiolecta. *Asian Journal of Plant Biology*, 2(1):1–6, 2014.
- [35] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John ML Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8):e1003800, 2014.
- [36] Charles P Winsor. The gompertz curve as a growth curve. *Proceedings of the national academy of sciences*, 18(1):1–8, 1932.
- [37] SE Aggrey. Comparison of three nonlinear and spline regression models for describing chicken growth curves. *Poultry science*, 81(12):1782–1788, 2002.
- [38] J Matthias Starck and Robert E Ricklefs. *Avian growth and development: evolution within the altricial-precocial spectrum*. Oxford University Press, 1998.
- [39] Guy E. Skinner, John W. Larkin, and E. Jeffery Rhodehamel. Mathematical modeling of microbial growth: a review. *Journal of food safety*, 14(3):175–217, 1994.

- [40] C Lennart Baumgärtner, Rupert Way, Matthew C Ives, and J Doyne Farmer. The need for better statistical testing in data-driven energy technology modeling. *Joule*, 8(9):2453–2466, 2024.
- [41] Aleh Cherp, Vadim Vinichenko, Jale Tosun, Joel A Gordon, and Jessica Jewell. National growth dynamics of wind and solar power compared to the growth required for global climate targets. *Nature Energy*, 6(7):742–754, 2021.
- [42] Mohammad Bakher Naseri and Greg Elliott. The diffusion of online shopping in australia: Comparing the bass, logistic and gompertz growth models. *Journal of Marketing Analytics*, 1:49–60, 2013.
- [43] Daisuke Satoh. Model selection among growth curve models that have the same number of parameters. *Cogent Mathematics & Statistics*, 6(1):1660503, 2019.
- [44] S Vieira and R Hoffmann. Comparison of the logistic and the gompertz growth functions considering additive and multiplicative error terms. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 26(2):143–148, 1977.
- [45] M Dhar and P Bhattacharya. Comparison of the logistic and the gompertz curve under different constraints. *Journal of Statistics and Management Systems*, 21(7):1189–1210, 2018.
- [46] Elvan Akin, Neslihan Nesliye Pelen, Ismail Uğur Tiryaki, and Fusun Yalcin. Parameter identification for gompertz and logistic dynamic equations. *PLoS One*, 15(4):e0230582, 2020.
- [47] Tatcha Sudtasan and Hitoshi Mitomo. Comparison of diffusion models for forecasting the growth of broadband markets in thailand. In *14th ITS Asia-Pacific Regional Conference, Kyoto 2017: Mapping ICT into Transformation for the Next Information Society*. International Telecommunications Society (ITS), 2017.
- [48] Manohar Nagula. Forecasting of fuel cell technology in hybrid and electric vehicles using gompertz growth curve. *Journal of Statistics and Management Systems*, 19(1):73–88, 2016.
- [49] Rupert Way, Matthew C Ives, Penny Mealy, and J Doyne Farmer. Empirically grounded technology forecasts and the energy transition. *Joule*, 6(9):2057–2082, 2022.
- [50] H Peter Boswijk and Philip Hans Franses. On the econometrics of the bass diffusion model. *Journal of Business & Economic Statistics*, 23(3):255–268, 2005.
- [51] Frank M Bass. A new product growth for model consumer durables. *Management science*, 15(5):215–227, 1969.
- [52] Paul A Geroski. Models of technology diffusion. *Research policy*, 29(4-5):603–625, 2000.
- [53] Nigel Meade and Towhidul Islam. Technological forecasting—model selection, model stability, and combining models. *Management science*, 44(8):1115–1130, 1998.
- [54] Aslan Lotfi, Ali Lotfi, and William E Halal. Forecasting technology diffusion: a new generalisation of the logistic model. *Technology Analysis & Strategic Management*, 26(8):943–957, 2014.
- [55] Avi Jakhmola, Jessica Jewell, Vadim Vinichenko, and Aleh Cherp. Probabilistic projections of global wind and solar power growth based on historical national experience. *Nature Energy*, 11(5):743–755, 2026.
- [56] Maria Priestley, Timothy J Sluckin, and Thanassis Tiropanis. Innovation on the web: the end of the s-curve? *Internet Histories*, 4(4):390–412, 2020.
- [57] W. Brian Arthur. Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 99(394):116–131, 1989. Foundational work on path dependence and technology lock-in effects.

- [58] John C Fisher and Robert H Pry. A simple substitution model of technological change. *Technological forecasting and social change*, 3:75–88, 1971.
- [59] Subhash C Bhargava. A generalized form of the fisher-pry model of technological substitution. *Technological Forecasting and Social Change*, 49(1):27–33, 1995.
- [60] Kenneth P Burnham, David R Anderson, Kenneth P Burnham, and David R Anderson. *Practical use of the information-theoretic approach*. Springer, 1998.
- [61] Leigh J Halliwell. The log-gamma distribution and non-normal error. *Variance*, 13(2):173–189, 2021.
- [62] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
- [63] Siddhartha Chib and Edward Greenberg. Understanding the metropolis-hastings algorithm. *The american statistician*, 49(4):327–335, 1995.
- [64] Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 2796–2804. PMLR, 2018.
- [65] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268, 2007.
- [66] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [67] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.
- [68] Wes McKinney et al. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445(1):51–56, 2010.
- [69] Skipper Seabold, Josef Perktold, et al. Statsmodels: econometric and statistical modeling with python. *SciPy*, 7(1):92–96, 2010.
- [70] Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, 7(1):110–120, 1997.
- [71] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.
- [72] Mylene Bedard. Optimal acceptance rates for metropolis algorithms: Moving beyond 0.234. *Stochastic Processes and their Applications*, 118(12):2198–2222, 2008.
- [73] Aidan Li, Liyan Wang, Tianye Dou, and Jeffrey S Rosenthal. Exploring the generalizability of the optimal 0.234 acceptance rate in random-walk metropolis and parallel tempering algorithms. *Communications in Statistics-Simulation and Computation*, pages 1–31, 2025.

- [74] Sean Talts, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*, 2018.
- [75] William A Link and Mitchell J Eaton. On thinning of chains in mcmc. *Methods in ecology and evolution*, 3(1):112–115, 2012.
- [76] A. Philip Dawid. Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society: Series A (General)*, 147(2):278–292, 1984.
- [77] Francis X. Diebold, Todd A. Gunther, and Anthony S. Tay. Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883, 1998.
- [78] Murray Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472, 1952.
- [79] Herbert A. David and Haikady N. Nagaraja. *Order Statistics*. Wiley, Hoboken, NJ, 3 edition, 2003.
- [80] Charles J Clopper and Egon S Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.
- [81] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, FL, 3 edition, 2013.
- [82] Christian P Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2nd edition, 2007.
- [83] Carl N Morris. Parametric empirical bayes inference: theory and applications. *Journal of the American statistical Association*, 78(381):47–55, 1983.
- [84] James Berger. The case for objective bayesian analysis. *Bayesian Analysis*, 1(3):385 – 402, 2006.
- [85] Timothy R Konold, Elizabeth A Sanders, and Kelvin Afolabi. From seinfeld to statistics: On the dangers of double dipping with bayesian inference. *Structural Equation Modeling: A Multidisciplinary Journal*, 32(4):647–656, 2025.
- [86] Bradley P Carlin and Thomas A Louis. *Bayesian methods for data analysis*. CRC Press, 3rd edition, 2008.
- [87] Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.
- [88] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.
- [89] Stephen P Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.
- [90] Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pages 473–483, 1992.
- [91] Diego A Comin and Bart Hobijn. The chat dataset. Working Paper 15319, National Bureau of Economic Research, September 2009. data retrieved from: <https://data.nber.org/data-appendix/w15319/>.
- [92] Gregory Nemet, Jenna Greene, Finn Müller-Hansen, and Jan C Minx. Dataset on the adoption of historical technologies informs the scale-up of emerging carbon dioxide removal measures. *Communications Earth & Environment*, 4(1):397, 2023. data retrieved from: <https://www.nature.com/articles/s43247-023-01056-1>.

- [93] The World Bank. Mobile cellular subscriptions (per 100 people), 2023. data retrieved from World Development Indicators, <https://data.worldbank.org/indicator/IT.CEL.SETS.P2>.
- [94] Hannah Ritchie and Max Roser. Technology adoption. *Our World in Data*, 2017. <https://ourworldindata.org/technology-adoption>.
- [95] International Energy Agency. Energy statistics data browser, 2023.
- [96] Arnulf Grübler. Time for a change: on the patterns of diffusion of innovation. *Daedalus*, 125(3):19–42, 1996.
- [97] Arnulf Grübler and Nebojša Nakićenović. Long waves, technology diffusion, and substitution. *Review (Fernand Braudel Center)*, 14(2):313–343, 1991.
- [98] IHS Markit Ltd. Sea-web ships: Maritime fleet register, 2023. Fleet register provider, acquired by S&P Global Inc. Accessed at <https://ihsmarkit.com/>.
- [99] RethinkX. Rethinkx technology production data, 2024.
- [100] Leopold Janauschek. *Originum cisterciensium*, volume 1. Alfred Hoelder, 1877.
- [101] Hannah Ritchie, Pablo Rosado, and Max Roser. Energy production and consumption. *Our World in Data*, 2020. <https://ourworldindata.org/energy-production-consumption>.
- [102] Hannah Ritchie and Pablo Rosado. Electricity mix. *Our World in Data*, 2020. <https://ourworldindata.org/electricity-mix>.
- [103] United Nations Statistics Division and Organisation for Economic Co-operation and Development. International standard industrial classification of all economic activities (isic), revision 4, 2008. Used for Industry, Infrastructure, and Transport sectors classification.
- [104] International Energy Agency. World energy statistics, 2023. Used for Energy sector classification.
- [105] World Health Organization. World health organization classifications, 2023. Used for Health, Waste Management, and Water sectors classification.
- [106] Food and Agriculture Organization. Food and agriculture organization classifications, 2023. Used for Agriculture sector classification.
- [107] International Telecommunication Union. International telecommunication union standards, 2023. Used for ICT sector classification.
- [108] United States Geological Survey. Usgs mineral commodity summaries, 2023. Used for mining-related technologies consolidated into Industry sector classification.
- [109] Vito MR Muggeo. Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19):3055–3071, 2003.
- [110] International Energy Agency. Tracking clean energy progress 2023, 2023.
- [111] IRENA. Sodium-ion batteries: A technology brief, 2025.
- [112] International Energy Agency. Batteries and secure energy transitions, 2024.
- [113] Charlotte Muller. The aluminum monopoly and the war. *Political Science Quarterly*, 60(1):14–43, 1945.
- [114] Sam H. Patterson. *Bauxite reserves and potential aluminum resources of the world*. Number 1228 in Geological Survey Bulletin. U.S. Government Printing Office, Washington, D.C., 1967.
- [115] O Yul Kwon. The korean financial crisis: diagnosis, remedies and prospects. *Journal of the Asia Pacific Economy*, 3(3):331–357, 1998.

- [116] SY Kim. South korea’s policy responses to global economic crisis. In *Proceedings of the First Middle East Conference on Global Business, Economics, Finance and Banking*, pages 1–27, 2014.
- [117] International Monetary Fund. The IMF’s response to the asian crisis, 1997.
- [118] Seung-kyung Kim and John Finch. Living with rhetoric, living against rhetoric: Korean families and the imf economic crisis. *Korean Studies*, pages 120–139, 2002.
- [119] Colin Weiss. De-dollarization? diversification? exploring central bank gold purchases and the dollar’s role in international reserves. International Finance Discussion Papers 1420, Board of Governors of the Federal Reserve System, Washington, 2025.
- [120] Christine T. Reilly. Food fight: Eating and identity in japan during the second world war. *International Journal of Arts and Sciences*, 3(13):410–431, 2010. Documents severe sugar scarcity during WWII in Japan, with Americans blocking shipping from Taiwan causing per capita consumption to decline sharply by 1945.
- [121] L. Halliday Piel. Food rationing and children’s self-reliance in japan, 1942–1952. *The Journal of the History of Childhood and Youth*, 5(3):393–418, 2012. Examines food rationing in Japan during and after WWII, including severe restrictions on sugar from 1942-1952.
- [122] Bruce Foster Johnston. *Food and agriculture in Japan, 1880-1950*. Stanford University, 1953.
- [123] Barak Kushner. Sweetness and empire: Sugar consumption in imperial japan. In Penelope Francks and Janet Hunter, editors, *The Historical Consumer: Consumption and Everyday Life in Japan, 1850-2000*, pages 127–151. Palgrave Macmillan, London, 2012. Discusses sugar rationing and consumption patterns in Japan during the 1930s-1940s wartime period.
- [124] Yasukichi Yasuba. Japan’s post-war growth in historical perspective. *Japan Forum*, 3(1):57–70, 1991.
- [125] Muhammad Fauzi, Indri Kartiko-Sari, and Hemant Poudyal. Trends of dietary intakes and metabolic diseases in japanese adults: assessment of national health promotion policy and national health and nutrition survey 1995–2019. *Journal of Clinical Medicine*, 11(9):2350, 2022.
- [126] Helgi Library. Sugar consumption per capita in japan. <https://www.helgilibrary.com/indicators/sugar-consumption-per-capita/japan/>, 2023. Data sourced from FAOSTAT. Accessed December 2025.
- [127] Yasuhiro Matsumura. Nutrition trends in japan. *Asia Pacific journal of clinical nutrition*, 10:S40–S47, 2001.
- [128] Kevin H. R. Rouwenhorst, Anthony S. Travis, and Leon Lefferts. 1921–2021: A century of renewable ammonia synthesis. *Sustainable Chemistry*, 3(2):149–171, 2022.
- [129] James Bessen. *Learning by Doing: The Real Connection between Innovation, Wages, and Wealth*. Yale University Press, New Haven, 2015.
- [130] Steven Klepper. Industry life cycles. *Industrial and Corporate Change*, 6(1):145–182, 1997.
- [131] David S. Landes. *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge University Press, Cambridge, 1969.
- [132] Gisela Rua. Diffusion of containerization. Finance and Economics Discussion Series 2014-88, Board of Governors of the Federal Reserve System, Washington, DC, 2014.
- [133] Urs Schenker, Martin Scheringer, and Konrad Hungerbühler. Investigating the global fate of DDT: Model evaluation and estimation of future trends. *Environmental Science & Technology*, 42(4):1178–1184, 2008.

- [134] National Research Council. Fostering research on the economic and social impacts of information technology. Technical report, National Academies Press, Washington, DC, 1998.
- [135] Bryce Ryan and Neal C. Gross. The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology*, 8(1):15–24, 1943.
- [136] Harold A. Linstone and Devendra Sahal. *Technological Substitution: Forecasting Techniques and Applications*. Elsevier, New York, 1976.
- [137] David A. Hounshell. *From the American System to Mass Production, 1800–1932: The Development of Manufacturing Technology in the United States*. Johns Hopkins University Press, Baltimore, 1984.
- [138] Suzi Kerr and Richard G. Newell. Policy-induced technology adoption: Evidence from the U.S. lead phasedown. *Journal of Industrial Economics*, 51(3):317–343, 2003.
- [139] IRENA. Total solar capacity, 2025. Retrieved September 29, 2025 from <https://archive.ourworldindata.org/20250909-093708/grapher/installed-solar-pv-capacity.html> (archived on September 9, 2025).
- [140] Gregory F. Nemet. Interim monitoring of cost dynamics for publicly supported energy technologies. *Energy Policy*, 37(3):825–835, 2009. Obtained via the Performance Curve Database.
- [141] J Doyne Farmer and François Lafond. How predictable is technological progress? *Research Policy*, 45(3):647–665, 2016.
- [142] Ludwig von Bertalanffy. Quantitative Laws in Metabolism and Growth. *The Quarterly Review of Biology*, 32(3):217–231, September 1957. Publisher: The University of Chicago Press.
- [143] Geoffrey A Moore. *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers*. HarperBusiness, New York, 3rd edition, 2014.
- [144] Wenxin Ye, Kangxin An, Shihui Zhang, Zihan Zhen, Wenjia Cai, and Can Wang. Probabilistic analysis of tripling global renewables based on bayesian-inferred growth dynamics model. *Renewable Energy*, 256:124496, 2026.
- [145] Béla Nagy, J Doyne Farmer, Quan M Bui, and Jessika E Trancik. Statistical basis for predicting technological progress. *PloS one*, 8(2):e52669, 2013.
- [146] Arnulf Grübler. *Technology and Global Change*. Cambridge University Press, 1998.
- [147] Brendon Tankwa, Lucas Vazquez Bassat, Peter Barbrook-Johnson, and J Doyne Farmer. Technological progress at national level: Increasing diffusion speeds with ever-changing leaders and followers. Working paper no. 2025-01, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, Oxford, UK, 2025.
- [148] Charlie Wilson, Arnulf Grubler, Nuno Bento, Stephen Healey, Simon De Stercke, and Caroline Zimm. Granular technologies to accelerate decarbonization. *Science*, 368(6486):36–39, 2020.
- [149] Diego Comin and Bart Hobijn. An exploration of technology diffusion. *American economic review*, 100(5):2031–2059, 2010.
- [150] Didier Sornette. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Springer, 2004.
- [151] Andrew Gelman, Daniel Simpson, and Michael Betancourt. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017.
- [152] Stefan Van Dongen. Prior specification in bayesian statistics: three cautionary tales. *Journal of theoretical biology*, 242(1):90–100, 2006.

- [153] Hannah Ritchie and Pablo Rosado. Energy mix. *Our World in Data*, 2020. Accessed at <https://ourworldindata.org/energy-mix>.
- [154] Cesare Marchetti. Primary energy substitution models: on the interaction between energy and society. *Technological forecasting and social change*, 10(4):345–356, 1977.
- [155] J-F Mercure, Pablo Salas, Pim Vercoulen, Gregor Semieniuk, Aileen Lam, Hector Pollitt, Philip B Holden, Negar Vakilifard, Unnada Chewpreecha, Neil R Edwards, et al. Reframing incentives for climate policy action. *Nature Energy*, 6(12):1133–1143, 2021.
- [156] Femke JMM Nijse, Jean-Francois Mercure, Nadia Ameli, Francesca Larosa, Sumit Kothari, Jamie Rickman, Pim Vercoulen, and Hector Pollitt. The momentum of the solar energy transition. *Nature Communications*, 14(1):6542, 2023.
- [157] International Energy Agency. Energy efficiency 2021. Technical report, IEA, Paris, 2021.
- [158] William Beaver. The failed promise of nuclear power. *The Independent Review*, 15(3):399–411, 2011.
- [159] Richard Pfau. *No Sacrifice Too Great: The Life of Lewis Strauss*. Charlottesville: University of Virginia Press, 1985.
- [160] Andrew Hopkins. Was three mile island a ‘normal accident’? *Journal of contingencies and crisis management*, 9(2):65–72, 2001.
- [161] Henry W Kendall. The failure of nuclear power. In *Risk, Organizations, and Society*, pages 163–218. Springer, 1991.
- [162] Anne-Sophie Hacquin, Sacha Altay, Lene Aarøe, and Hugo Mercier. Disgust sensitivity and public opinion on nuclear energy. *Journal of Environmental Psychology*, 80:101749, 2022.
- [163] Wolfgang Krohn and Peter Weingart. Commentary: Nuclear power as a social experiment—european political “fall out” from the chernobyl meltdown. *Science, Technology, & Human Values*, 12(2):52–58, 1987.
- [164] IPCC. Climate change 2022: Mitigation of climate change. <https://www.ipcc.ch/report/ar6/wg3/>, 2022. Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
- [165] IPCC. Climate change 2022: Mitigation of climate change. summary for policymakers. <https://www.ipcc.ch/report/ar6/wg3/>, 2022. Summary for Policymakers, Working Group III contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.
- [166] International Energy Agency. *Net Zero Roadmap: A Global Pathway to Keep the 1.5 °C Goal in Reach*. OECD Publishing, 2023.
- [167] Celine Guivarch, Elmar Kriegler, Joana Portugal-Pereira, Valentina Bosetti, James Edmonds, Manfred Fischedick, Petr Havlik, Paulina Jaramillo, Volker Krey, Franck Lecocq, et al. Annex iii: Scenarios and modelling methods. *IPCC 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 1842–1908, 2023.
- [168] Francois Rozon, Craig McGregor, and Michael Owen. Long-term forecasting framework for renewable energy technologies’ installed capacity and costs for 2050. *Energies*, 16(19):6874, 2023.
- [169] Alessandro Martulli, Kristin Brandt, Florian Allroggen, and Robert Malina. The potential scale-up of sustainable aviation fuels production capacity to meet global and eu policy targets. *Nature Communications*, 16(1):11619, 2025.

- [170] International Energy Agency. World energy outlook 2025. Technical report, IEA, Paris, 2025. Licence: CC BY 4.0 (report); CC BY NC SA 4.0 (Annex A).
- [171] IRENA. Renewable power generation costs in 2023. Technical report, International Renewable Energy Agency, Abu Dhabi, 2024.
- [172] Kristina Sadovskaia, Dmitrii Bogdanov, Samuli Honkapuro, and Christian Breyer. Power transmission and distribution losses—a model based on available empirical data and future trends for all countries globally. *International Journal of Electrical Power & Energy Systems*, 107:98–109, 2019.
- [173] Hannah Ritchie. Primary, secondary, final, and useful energy: Why are there different ways of measuring energy? *Our World in Data*, 2022. <https://archive.ourworldindata.org/20251125-173858/energy-definitions.html>.
- [174] George Casella and Roger Berger. *Statistical inference*. Chapman and Hall/CRC, 2024.
- [175] Network for Greening the Financial System. NGFS climate scenarios for central banks and supervisors – phase v. Technical report, Network for Greening the Financial System, 2024. Scenario data accessed via the NGFS–IIASA Scenario Explorer, <https://data.ece.iiasa.ac.at/ngfs>.
- [176] Lukas Wagner, Jiajia Suo, Bowen Yang, Dmitry Bogachuk, Estelle Gervais, Robert Pietzcker, Andrea Gassmann, and Jan Christoph Goldschmidt. The resource demands of multi-terawatt-scale perovskite tandem photovoltaics. *Joule*, 8(4):1142–1160, 2024.
- [177] Keiichi Komoto, Masakuza Ito, Peter van der Vleuten, David Faiman, and Kosuke Kurokawa. *Energy from the desert: very large scale photovoltaic systems: socio-economic, financial, technical and environmental aspects*. Routledge, 2007.
- [178] IEA. Electricity mid-year update 2025. Technical report, International Energy Agency, Paris, 2025. Licence: CC BY 4.0.
- [179] Peter E Glaser. Power from the sun: its future. *Science*, 162(3856):857–861, 1968.
- [180] Robert L Olsen et al. Space solar power to meet future energy needs. In *Space Solar Power Review*, volume 1, pages 49–72, 1979.