

Probabilistic methods

How to create empirically validated probabilistic forecasts?
With applications to technology forecasting

François Lafond

PRISMA summer school on uncertainty in energy and IAMs

June 2026



Learning objectives

How can one give the future probability of various outcomes, by studying past data?

- Basic idea: check past forecast errors to inform prediction intervals.
- Stronger case: check whether past forecast errors are compatible with a clear parametric model

We will work through one simple model in detail: a (geometric) random walk with drift.

- Define the model and derive expected forecast errors
- Use backtesting to see if past forecast errors are as expected
- (Standardize forecast errors so different series and horizons can be pooled);
- Construct surrogate datasets to perform testing that accounts for correlations and sample sizes
- If errors are larger than expected, modify the model to allow for larger errors

Then we will look at extensions and other examples.

- Adding an exogenous variable: Wright's law / "Learning" curve
- Non-linear time series: Diffusion curves

The forecasting object

A probabilistic forecast is a full distribution over possible outcomes

$$F_{t,\tau}(y) = \mathbb{P}(Y_{t+\tau} \leq y \mid \mathcal{I}_t)$$

$Y_{t+\tau}$ target variable observed at future time $t + \tau$.

τ forecast horizon.

$\mathcal{I}_t$ information available when the forecast is made.

$F_{t,\tau}$ predictive distribution, from which quantiles, intervals, probabilities, and expected losses can be computed.

- **Unconditional forecast:** What will happen to variable y ?
- **Conditional forecast:** What will happen to variable y if x takes a particular path?
- **Multivariate:** What will happen to variables y and x (unconditional)? / to variables y and x knowing what will happen to z (conditional)?

Scenarios = conditional forecasts?

Example: Unconditional forecast of technological progress

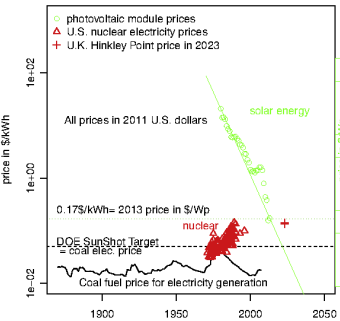


Fig. 1. Different technologies can have very different long-run cost trends.

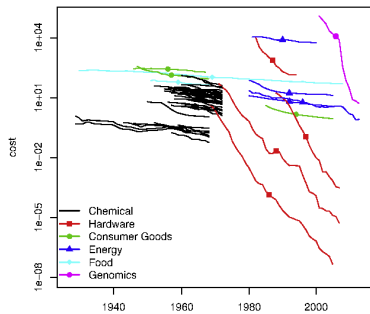


Fig. 2. The dataset contains many short, heterogeneous cost series.

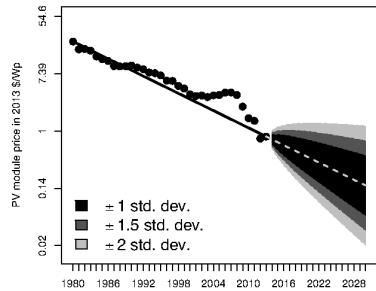


Fig. 10. The goal is a distributional forecast, not just a fitted trend.

Source: Farmer, J.D. and Lafond, F. (2016), How predictable is technological progress?, *Research Policy*

Calibration and sharpness

Calibration. Events assigned probability p occur about fraction p of the time, under repeated comparable forecasts.

$$\mathbb{P}(Y_{t+\tau} \leq q_{\alpha,t,\tau}) \approx \alpha$$

Sharpness. Conditional on calibration, narrower predictive distributions are more useful.

$$\text{width}_{0.90} = q_{0.95,t,\tau} - q_{0.05,t,\tau}$$

A narrow interval that misses too often is overconfident. A very wide interval that always covers is usually uninformative.

Moore's law for one technology

For now, focus on a single technology and suppress the technology index. Let c_t be unit cost and $y_t = \ln c_t$. Deterministic exponential cost decline implies

$$y_t = y_0 + \mu t.$$

A stochastic version is a random walk with drift:

$$\Delta y_t \equiv y_t - y_{t-1} = \mu + \epsilon_t, \quad \epsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

With a training window of m growth rates ending at $t = m + 1$,

$$\hat{\mu} = \frac{1}{m} \sum_{s=2}^t \Delta y_s,$$

$$\hat{y}_{t+\tau} = y_t + \hat{\mu}\tau.$$

Deriving the forecast-error variance

$$\text{True: } y_{t+\tau} = y_t + \mu\tau + \sum_{j=1}^{\tau} \epsilon_{t+j}. \quad (1)$$

$$\text{Predicted: } \hat{y}_{t+\tau} = y_t + \hat{\mu}\tau. \quad (2)$$

The forecast error is

$$\mathcal{E}_{t,\tau} \equiv y_{t+\tau} - \hat{y}_{t+\tau} = (\mu - \hat{\mu})\tau + \sum_{j=1}^{\tau} \epsilon_{t+j}.$$

$$(\hat{\mu} - \mu) \sim \mathcal{N}\left(0, \frac{\sigma^2}{m}\right) \quad (3)$$

$$\sum_{j=1}^{\tau} \epsilon_{t+j} \sim \mathcal{N}\left(0, \sigma^2\tau\right) \quad (4)$$

and past and future shocks are independent, we have

$$\text{Var}(\mathcal{E}_{t,\tau}) = \sigma^2 \frac{\tau^2}{m} + \sigma^2\tau = \sigma^2 \left(\tau + \frac{\tau^2}{m} \right).$$

The probabilistic forecast

If σ^2 is known and the model is correct,

$$y_{t+\tau} \mid \mathcal{I}_t \sim \mathcal{N} \left(y_t + \hat{\mu}\tau, \sigma^2 \left(\tau + \frac{\tau^2}{m} \right) \right).$$

Hence a central $1 - \alpha$ prediction interval for log cost is

$$y_t + \tau \hat{\mu} \pm z_{1-\alpha/2} \sigma \sqrt{\tau + \frac{\tau^2}{m}},$$

where $z_{1-\alpha/2} = 1.96$ for a 95% interval.

In practice, σ^2 is not known so could use $\hat{\sigma}^2$ and the t distribution.

But we have an entire panel of technologies

Now let i index technologies:

$$\Delta y_{i,t} = \mu_i + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim \mathcal{N}(0, \sigma_i^2).$$

The forecast-error law is

$$\mathcal{E}_{i,t,\tau} \sim \mathcal{N}(0, \sigma_i^2 A_{\tau,m}), \quad A_{\tau,m} = \tau + \frac{\tau^2}{m}.$$

The surprise is what *does not* carry an i index:

$A_{\tau,m}$ depends only on horizon and window length.

Technologies have different drifts and volatilities, but the same normalized forecast-error scale under the baseline model.

Standardize, then pool

Standardize forecast errors by the scale predicted by the model:

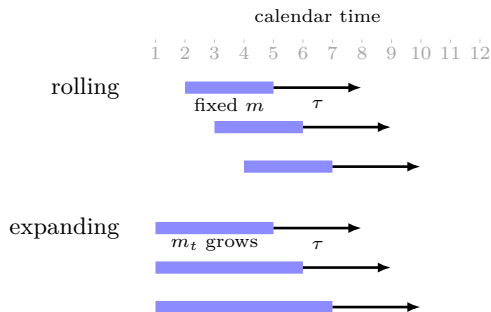
$$\frac{\mathcal{E}_{i,t,\tau}}{\sigma_i \sqrt{A_{\tau,m}}} \sim \mathcal{N}(0, 1).$$

With estimated volatility,

$$\frac{\mathcal{E}_{i,t,\tau}}{\hat{\sigma}_{i,t} \sqrt{A_{\tau,m}}} \approx t_{m-1}.$$

- Different technologies have different μ_i and σ_i .
- Different horizons have different raw-error scales.
- After normalization, short heterogeneous series can be pooled to test one common forecast-error law.

Backtesting: Why rolling (non-expanding) windows?



Why choosing a non-expanding scheme?

- Rolling keeps sample size m fixed, so $A_{\tau,m}$ is comparable across origins.
- Expanding uses all available data, but m_t changes.

Note also

- Large τ leaves fewer origins with $t + \tau \leq T$: Long horizons are harder to assess.
- Overlap and reused samples make errors serially dependent: Econometric tests must address this: Diebold–Mariano, West, Clark–McCracken, Giacomini–White, Clements–Hendry.

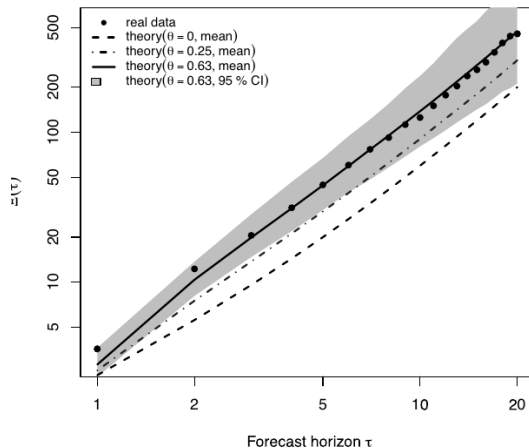
Surrogate datasets

A surrogate dataset is a model-based synthetic history, generated under a stated null model and evaluated with exactly the same procedure as the real data.

1. Fix the null model, for example the random walk with drift.
2. Generate B synthetic datasets $D^{(1)}, \dots, D^{(B)}$ with the same technologies (using full sample $\hat{\mu}_i$ and $\hat{\sigma}_i$), time coverage, training windows, and forecast horizons as the real dataset.
3. For each surrogate dataset, rerun the full backtesting algorithm and compute a diagnostic statistic of interest, $S^{(b)}$.
4. Compare the observed statistic S_{obs} with the surrogate distribution.

The surrogate distribution tells us what deviations are expected, assuming the model is correct, from finite samples, overlapping forecast errors, etc.

Empirical check: error growth



Source: Farmer and Lafond (2016).

Recall

$$E\left[\frac{\mathcal{E}_{i,t_0,\tau}^2}{\sigma_i^2}\right] = \tau + \frac{\tau^2}{m}$$

Diagnostic. Does the mean squared normalized forecast error $\Xi(\tau)$ grow with horizon as the model predicts?

- Black dots are empirical backtest errors.
- Lowest (dashed) line is $\tau + \tau^2/m$

Autocorrelation is not a detail

The i.i.d. random-walk model underpredicts empirical errors. A simple extension uses moving-average noise:

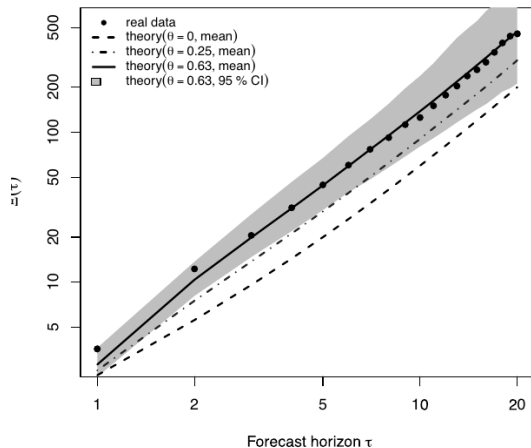
$$\eta_{i,t} = u_{i,t} + \rho u_{i,t-1}, \quad u_{i,t} \sim \mathcal{N}(0, \sigma_{u,i}^2), \quad \sigma_{\eta,i}^2 = (1 + \rho^2) \sigma_{u,i}^2.$$

If $\sigma_{\eta,i}^2$ denotes the variance of one-period observed noise. For large m and τ :

$$\text{Var}(\mathcal{E}_{i,t,\tau}) \approx \sigma_{\eta,i}^2 \frac{(1 + \rho)^2}{1 + \rho^2} \left(\tau + \frac{\tau^2}{m} \right).$$

Positive autocorrelation raises the long-run variance, so intervals widen even if the median forecast is unchanged.

Empirical check: error growth



Recall

$$E\left[\frac{\mathcal{E}_{i,t_0,\tau}^2}{\sigma_i^2}\right] = \tau + \frac{\tau^2}{m}$$

Diagnostic. Does the mean squared normalized forecast error $\Xi(\tau)$ grow with horizon as the model predicts?

- Black dots are empirical backtest errors.
- Different lines show predicted growth of forecast errors from the surrogate datasets under different autocorrelation parameter values.
- The grey band is the surrogate-dataset 95% region.

Source: Farmer and Lafond (2016).

Empirical check: distributional collapse

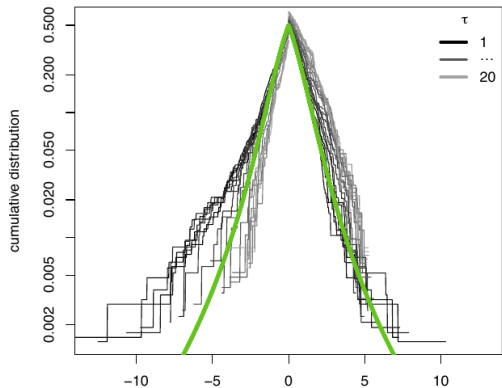


Fig. 7: rescaled errors by forecast horizon.

Source: Farmer and Lafond (2016). Pooling after standardization tests whether many technologies can be described by one common forecast-error law.

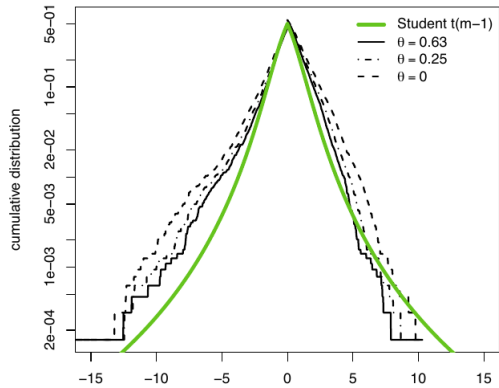
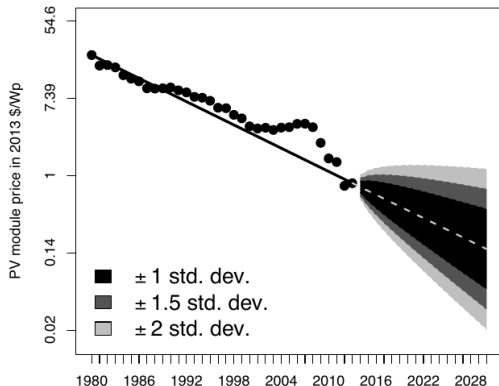


Fig. 8: rescaled errors pooled across horizons.

Assuming this validates the model, we are ready to make predictions



Farmer and Lafond (2016), Fig. 10: photovoltaic module price forecast from a calibrated log-cost model.

Let $t_0 = 2013$, $y_t = \ln c_t$, and use the observed PV history up to t_0 .

- Estimate the drift and noise scale from the PV series.
- Take autocorrelation $\rho = 0.63$.
- For each horizon τ , compute the predictive law for

$$y_{t_0+\tau} \sim \mathcal{N}\left(y_{t_0} + \hat{\mu}\tau, \sigma_\eta^2 \frac{(1+\rho)^2}{1+\rho^2} \left(\tau + \frac{\tau^2}{m}\right)\right)$$

- Transform the log-price quantiles back to 2013 \$/Wp.

Bayesian reinterpretation

An improper prior on μ would give the Bayesian posterior for μ as

$$\tilde{\mu} \mid \mathcal{I}_t \sim \mathcal{N}\left(\hat{\mu}, \frac{\sigma^2}{m}\right),$$

We can simulate future log cost by drawing from the distribution of the drift, and drawing future values the noise $\eta_{t+1}, \dots, \eta_{t+\tau} \sim \mathcal{N}(0, \sigma^2)$.

$$\tilde{y}_{t+\tau} = y_t + \tau \tilde{\mu} + \sum_{s=1}^{\tau} \eta_{t+s}.$$

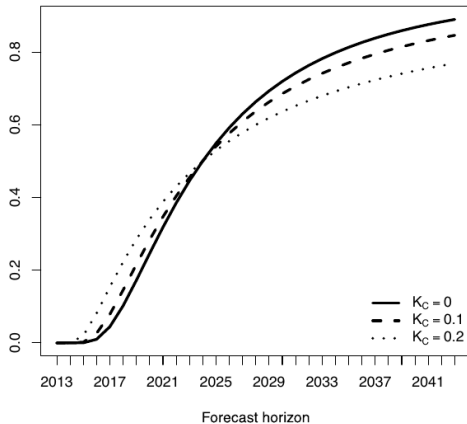
Integrating out $\tilde{\mu}$ gives

$$\tilde{y}_{t+\tau} \mid \mathcal{I}_t \sim \mathcal{N}\left(y_t + \tau \hat{\mu}, \sigma^2 \left(\tau + \frac{\tau^2}{m}\right)\right).$$

Decision use: probabilities, not only lines

Once we have a predictive distribution, we can ask:

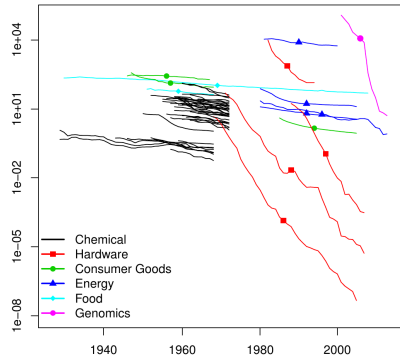
- What is $\mathbb{P}(c_{2035} < x)$ for a cost threshold x ?
- What is $\mathbb{P}(c_{A,2035} < c_{B,2035})$ for competing technologies?
- How much portfolio diversity is valuable when forecasts are uncertain?



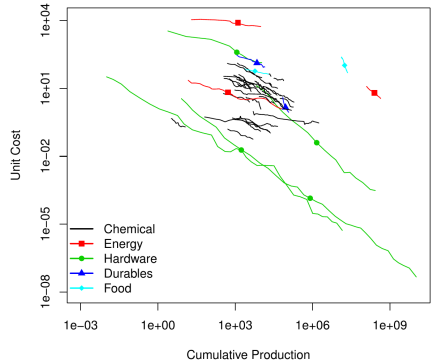
Farmer and Lafond (2016), Fig. 11: forecast probability that PV becomes cheaper than a competing technology.

Adding exogenous variables: Wright's law

Moore view: costs fall $\mu\%$ every year



Wright view: costs fall $\omega\%$ for every 1% of experience growth



cost
experience

$c_t = c_0 e^{\mu t}$
 $z_t = z_0 e^{r t}$

$\left. \vphantom{\begin{matrix} c_t = c_0 e^{\mu t} \\ z_t = z_0 e^{r t} \end{matrix}} \right\} \begin{matrix} c_t \propto z_t^\omega \\ \omega = \mu / r \end{matrix}$

If experience grows roughly exponentially, time trends and experience curves are hard to separate.

Stochastic Wright's law

Let $X_t = \Delta \ln z_t$ and $Y_t = \Delta \ln c_t$.

A differenced Wright model is:

$$Y_t = \omega X_t + \eta_t.$$

Forecasts are conditional on future experience:

$$\hat{y}_{t+\tau} = y_t + \hat{\omega} \sum_{s=1}^{\tau} X_{t+s}.$$

Moore forecasts costs unconditionally from time.

Wright forecasts costs conditional on a deployment path, so it is more relevant for policy counterfactuals.

Forecast uncertainty under Wright's law

A useful approximation is:

$$\text{Var}(\mathcal{E}_\tau) = \sigma_\eta^2 \left(\tau + \frac{\tau^2}{m} \frac{\hat{r}_{\text{future}}^2}{\hat{r}_{\text{past}}^2 + \hat{\sigma}_{z,\text{past}}^2} \right).$$

- More variation in past experience growth improves estimation of ω .
- Faster future experience growth amplifies slope-estimation error.
- Third, a high growth rate of experience in the past (high r_{past}) also reduces the forecast errors, as experience spans a larger range

More on Wright's law

Wright's law is appealing...

- It translates deployment choices into potential cost changes.
- It supports policy counterfactuals and portfolio problems.
- It helps characterise a tradeoff between increasing returns and diversification.

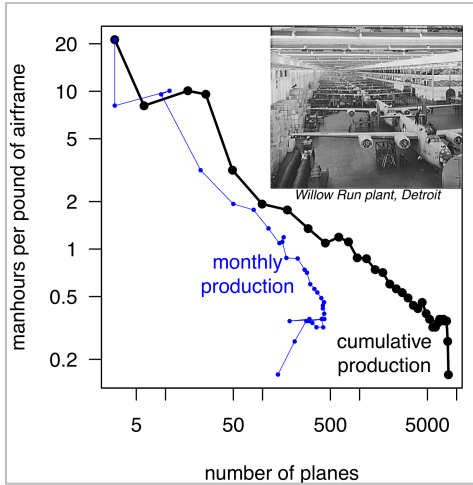
...but hard to test: causality and multi-collinearity

- Costs and production co-evolve: cheap technologies diffuse, and diffusion may make them cheaper.
- Log cost and log cumulative production are often both persistent.
- Level regressions can be spurious, but no co-integration

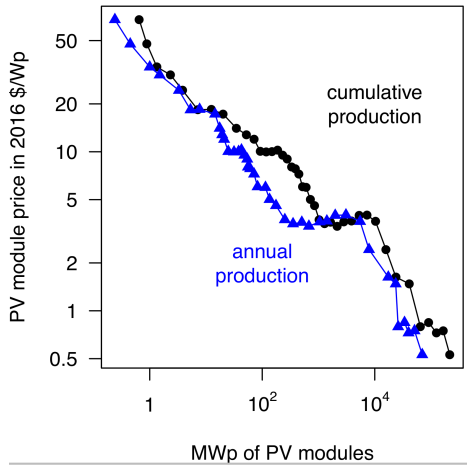
Practical implication. Wright's law is attractive for policy because it is endogenous to deployment, but it is hard to establish that it is better than Moore's law.

A useful natural experiment

WWII

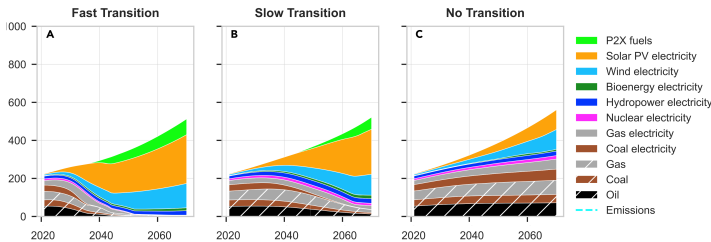


Solar panels

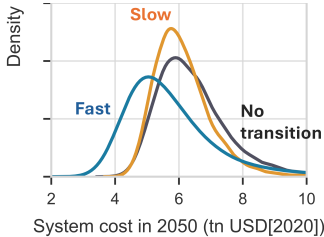


Net-zero transition: a fast transition is likely cheaper

Useful energy by scenario (EJ/yr)



Distribution of 2050 system cost



Faster deployment drives experience-curve cost declines, so the fast transition reaches lower 2050 system cost with high probability.

Source: Way, Ives, Mealy and Farmer (2022), *Joule*.

Non-linearity: diffusion curves

Forecasting technology *diffusion* is harder because:

- growth is nonlinear and eventually saturates;
- the asymptote L is often unknown;
- early data can look exponential even when the process is S-shaped;
- errors can be autocorrelated, heteroscedastic, and biased.

All work presented here derives from Ben Wagenvoort's PhD thesis (Wagenvoort, B., Dyer, J., Lafond F. and Farmer, J.D., Universality and predictability of technology diffusion, in progress)

A useful S-curve family

The Bertalanffy–Richards model is a flexible one-parameter S-curve family:

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

with closed-form solution

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

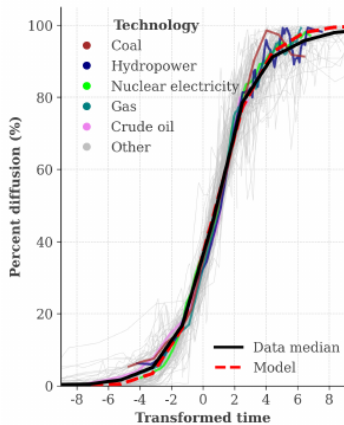
k initial growth rate.

L asymptotic diffusion level.

t_0 locates the growth process in time.

β shape parameter ($\beta = 1$ recovers the logistic).

Common coordinates reveal a universal diffusion shape



(a) Collapse plot

For each technology i , fit the location, scale, and asymptote of its S-curve. Then transform the raw history into common coordinates:

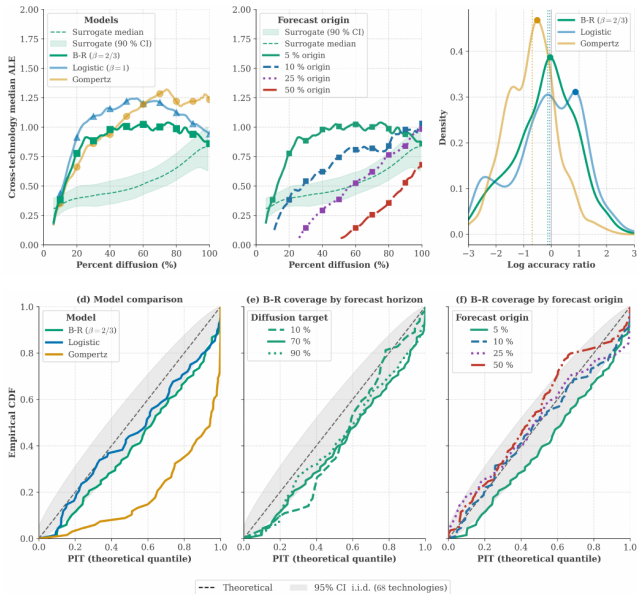
$$\tilde{Y}_{i,t} = \frac{Y_{i,t}}{\hat{L}_i}, \quad \tilde{t}_{i,t} = \hat{k}_i(t - \hat{t}_{0,i}).$$

This sets the fitted asymptote, speed, and location to common values:

$$\tilde{L} = 1, \quad \tilde{k} = 1, \quad \tilde{t}_0 = 0.$$

- Retrospective normalization compares shapes, not forecasts.
- The collapse suggests a common diffusion geometry across mature technologies.

Source: Wagenvoort, Dyer, Lafond and Farmer (2026), **work in progress**.



Backtest approach:

- **Ground truth:** mature technologies only.
- **Origins:** at 5, 10, 25, 50% of $\hat{L}$, refit the Bertalanffy–Richards S-curve on data up to t and forecast forward.
- **Forecast:** Empirical-Bayes posterior predictive; point forecast is its median.
- **Accuracy, bias (a–c):** absolute log error $|\ln(\hat{Y}/Y)|$, cross-technology median by horizon.
- **Calibration (d–f):** coverage after a widening transform, vs. a B-R surrogate ensemble.

Going further

Important topics we have only touched:

- **Conformal prediction:** calibrate residuals $r_j = |y_j - \hat{y}_j|$ and use their $1 - \alpha$ quantile; useful, but exchangeability is a serious assumption for time series and scenarios.
- **Proper scoring rules:** evaluate full distributions with log score, CRPS, interval scores, and Brier scores for events.
- **Forecast combinations:** averages, Bayesian model averaging, stacking, and mixture-of-experts methods can outperform single models.
- **Dependence:** multivariate forecasts need correlated errors, common factors, and redundancy diagnostics.
- **Non-stationarity:** structural breaks, changing policies, and changing measurement systems can dominate elegant formulas.
- **Decision analysis:** the best forecast depends on the loss function and the action it informs.