Forecasting technological progress

forthcoming in "The Economy as an Evolving Complex System IV", SFI Press

François Lafond*[a,b]

[a]Institute for New Economic Thinking, University of Oxford
[b]Smith School of Enterprise and the Environment, University of Oxford.

July 7, 2026

Abstract

After a brief history of technological forecasting, I synthesize our work at the Institute for New Economic Thinking over the last decade developing time series models for performance curves. I conclude with ongoing efforts and a research agenda.

Keywords: Performance curves; Experience curves; Diffusion curves; Patent networks.

1 Introduction

Technology forecasting is essential for solving global problems and limiting exposure to catastrophic risks. Humanity's fate often depends on what are, essentially, technological bets. Can we risk emitting more carbon dioxide, because we think that carbon capture technologies will become scalable and cheap? Should we pause the development of generative Artificial Intelligence (AI) and its promises of great productivity gains, because we worry about catastrophic risks associated with artificial general intelligence, as advocated by Bengio et al. (2023)? Should we spend billions of dollars changing our encryption methods, because we think that quantum technologies will make progress quickly and make it possible to break current cryptographic protocols? Should we strictly regulate synthetic biology, stifling medical innovation, because it can create serious biohazard and terrorist threats?

In this chapter, I review recent attempts at predicting technological progress, focusing on predictions of technological performance and unit costs, exposing some key technical issues. I will also briefly discuss other approaches, including predictions of technology diffusion and predictions of patenting rates. Before diving into these, I provide a brief motivation and history of technology forecasting.

*This working paper is the full version of the invited book chapter for "The Economy as an Evolving Complex Systems IV", SFI press, forthcoming. I would like to thank all my collaborators on this research programme, and in particular Doyne Farmer, as well as Lorenzo Agnelli, Lennart Baumgärtner, Joel Dyer, José Moran, Anton Pichler, Benjamin Wagenvoort, Rupert Way and Owen Westfold for comments and stimulating discussions. The usual disclaimer applies. This research was funded by Baillie Gifford and the UKRI through ESRC grant PRINZ (ES/W010356/1). This version correct a typo to the formula for normalized errors p.7 – I am grateful to Wensupu Yang for pointing it out.

1.1 A brief history of technology forecasting

Why is technology forecasting done? Historically it has been done for (i) war, as the military "played a disproportionate role in the emergence of technological forecasting as a serious professional activity" (Ayres, 1969), (ii) industrial policy, as evidenced by the popularity of national foresights exercises (Harper, 2013); (iii) financial gain, as there is a clear business advantage to knowing the future of technology better than competitors, (iv) survival - I have already hinted at biohazards and rogue AI, but to give another example, nuclear technologies made their fair contribution to that motivation; and last but not least (v) fun - this should be evident from the popularity of the science fiction genre.

But while fears, hopes and fantasies about technology are indeed entertaining, technology forecasting is now a very serious activity. It developed as such in the 20th century. Before World War II, technology forecasting methods remained qualitative. In his chapter on "intuitive methods of forecasting", discussing forecasting by experts, Ayres (1969, p. 144) quipped "The first major forecasting effort by a panel, the 1937 study by the Natural Resources Committee of the National Research Council, was a very sober and responsible document which missed virtually all of the major developments of its decade, including antibiotics, radar, jet engines, and atomic energy."

After World War II, "technological forecasting emerged as a recognised management discipline" (Jantsch, 1967). This makes sense, as missed opportunities may cost companies dearly. In two famous examples, the inventor of xerography "pounded the pavement for years in a fruitless search for a company that would develop his invention into a useful product"[1], and the CEO of IBM is said to have declared in 1943 that there was a market for only a few computers[2].

Beyond these eye-catching examples of rather spectacular failure, however, early attempts at technological foresights appear to have been pretty decent. An early study, Gilfillan (1937), found that 60% of 75-year-ahead forecasts from a 1920 *Scientific American* were realized or almost so 16-years ahead. Ayres (1969, p.9-14) reports broadly similar success rates for a number of other public predictions of future inventions. More recently, and again similarly, Albright (2002) found that half of Kahn and Wiener (1967)'s "One hundred technical innovations very likely in the last third of the twentieth century" were realized.

In the 1970s and 1980s, technology forecasting became increasingly integrated with planning, leading to the development of foresights and scenarios for strategic planning (Martino, 1993), sometimes using semi-quantitative approaches such as Delphi (see below). In the 1990s and 2000s, the term "tech mining" (Porter and Cunningham, 2005) appeared, shifting the emphasis to data-based prediction. In the 2010's, larger datasets (web-based data, patents and publications, etc.) triggered a new era of technology forecasting, with new analytical techniques.

In the early 2020s, the growing availability of large language models makes it possible to deal with text data in a much easier and much more effective way, opening new avenues for rigorous or automated approaches to issues that have hitherto been seen as requiring more qualitative or expert-based insights, ranging from data collection to direct predictions of technological milestones. The next section provides a very succinct overview of currently used methods.

[1] According to Xerox's own account, https://www.xerox.com/downloads/usa/en/innovation/innovation_storyofxerography.pdf

[2] The Wikipedia page of Thomas J. Watson, section "Famous attribution" (accessed 1 August 2024) fails to track down a reliable source. Peter Schwartz, a well-known scenario planner, does document a 1981 IBM forecast that shipments of the X1000 would peak in 1983, calling it "the costliest slide in business history" (https://longnow.org/ideas/scenario-planning-for-the-long-term/).

1.2 Methods of technology forecasting

While this chapter focuses primarily on my and my collegues' work at the Complexity Economics group at INET Oxford, I first offer a list of the key methods in technology forecasting (for more complete reviews, see Ayres (1969), Jantsch (1967), Martino (1993) and Ciarli et al. (2016)).

Progress curves and learning curves. This method is based on the idea that a key feature of a technology, such as its unit cost or performance along a key metric, follows trends either as a function of time or as a function of "experience." I will discuss this extensively in Section 2.

Growth curves: diffusion and substitution. This method starts from the observation that there are predictable patterns in which technology diffuses, specifically, "S-shaped" patterns. I will discuss this in Section 3.2.

Patent networks Since the early 2000s we have seen the development of very large, easy to use databases of patents and scientific publications that include metadata that are relational in nature, making it possible to construct networks of technologies, such as citation networks between patents, or networks of similarities between patent categories. I will discuss this in Section 3.3.

Expert elicitation. These methods, such as the well-known Delphi method, do not just to query "experts", but try to avoid or even correct known expert biases: overconfidence, anchoring, unwillingness to back down from the publicly announced positions, peer effects, bandwagon effects, overcompensate past errors, vested interest, etc. Recent expert elicitation studies include an explicit process of expert debiasing (e.g. Verdolini et al. (2018)). Tetlock and Gardner (2016) documented that experts often do not perform better than laypeople, except for a few "superforecasters." Expert elicitations remains popular - see Grace et al. (2024) for an interesting example documenting predictions of AI experts' predictions about future AI capabilities.

Prediction markets. These are closely related to expert elicitation, but with a stronger focus on quantitatively leveraging the "wisdom of crowds." While some platforms have developed, they are not as popular or successful as was hoped a couple of decades ago (Arrow et al., 2008). A number of recent studies have appeared that use large language models as "experts" (or aides to experts), and evaluate performance against human-level performance on prediction platforms. Early results are very encouraging (Halawi et al., 2024; Schoenegger et al., 2024), and make it possible to address technological forecasts that are more qualitative in nature than is possible with the time-series methods reviewed here, while still using a rigorous evaluation framework, including in evaluating errors quantitatively.

Foresight, scenarios, and road maps. This refers to more general exercises than a single expert elicitation study. Foresight exercises constitute "an approach for collectively exploring, anticipating and shaping the future." (Harper, 2013). Most advanced economies regularly produce technology foresights and road maps. These are generally produced by groups of civil servants aided by experts and consultants.

Evolutionary dynamics and taxonomies. The literature on cultural evolution is starting to accumulate considerable data on technology phylogenies (Solé et al., 2013), typically constructed by carefully mapping how one design evolved from a previous one (in this volume, see Duran-Nebreda et al.

(2025, Fig.3) for an example in software). I am not aware that these techniques have been used for forecasting, but this would be an interesting avenue of research, as it is one of the rare approaches that has a shot at predicting qualitative change. Lafond and Kim (2019) studied the evolution of the US patent classification system, arguing that radical innovations are by definition *sui generis* (of their own kind), and should thus be reflected by the creation of new categories. Patent reclassification indeed appears related to quality, as measured by future citations. There have been some attempts at predicting the emergence of new patent categories (Érdi et al., 2013). While the work on economic complexity, reviewed in this volume by Neffke et al. (2025), Frenken and Neffke (2025) and Coyle (2025), does predict the probability that an entity adopts a new technology, it does so under a fixed classification scheme, making clear the need for techniques that predict the *expansion* of the technological space. Arthur (2025), in this volume, proposes combinatorial evolution as a mechanism through which the technological space expands.

Total factor productivity. In economics, time series forecasting is highly developed, and it is well acknowledged that long-run economic growth originates mostly in technological change. Thus we can think of most long-term forecasts of GDP as technology forecasts. However, there seems to be almost no work on directly forecasting long-term total factor productivity growth. An exception is Philippon (2023), who finds that factor productivity growth is better predicted using linear than exponential growth. Goldin et al. (2024) discuss how predictions of future of productivity growth revolve around "optimists" and "pessimists", with few quantitative forecasts being made.

So, what works best? I have not seen many studies comparing methods, as they usually predict different things. An area that has attracted attention, however, is the costs of energy technologies. This is important: The costs of the "backstop" technology in Integrated Assessment Models is key to determine the "optimal choice of abatement." Unfortunately, Way et al. (2022) documented that Integrated Assessment Models have systematically used too pessimistic forecasts of the costs of some renewables such as solar and wind. Meng et al. (2021) concluded that for key energy technologies, time series models outperformed experts.

I will come back to multi-methods approaches and forecast combination in the conclusion; for now, let us look in detail at time series models.

2 Performance curves and learning curves

Here I review recent work on "performance" or "cost curves" (as a function of time) and "learning curve" or "experience curves" (as a function of "experience").

2.1 Variables and datasets

Different studies use different concepts of performance, and when explaining performance using a learning curve approach, different studies use difference concepts for experience.

For performance, some authors use technical performance records along one metric (Koh and Magee, 2006, 2008), and study different metrics separately; others try to construct "scores", merging different metrics, based on engineering/expert knowledge (Martino, 1993; Benson et al., 2018). Statistical offices producing price indices try in principle to produce quality-adjusted prices, often using hedonic regressions; that is, they construct scores by inferring the relative importance of various performance characteristics directly from the data by observing which ones contribute most to

price differences (Gordon, 1990). Other studies use productivity metrics, either labor productivity (Lafond et al., 2022) or total factor productivity (Nordhaus, 2014).

In the case of learning curves, for "experience", some studies use installed capacity (that is, excluding retired capacity), or, similarly, account for obsolescence (Argote and Epple, 1990). Typically the variable under consideration is production, but sometimes, as in Arrow (1962)'s original paper, it is investment.

Finally, defining scope is a real problem. Even if we can find a good definition for what technology is - simplifying Arthur (2010), a technology is a specific set of means (typically from a particular techno-scientific domains) for a specific end - because technologies are made of other technologies recursively one needs to decide on a "level" or "boundary", such as energy technologies/solar energy/photovoltaic/thin-film, and in parallel, technology components (e.g. solar PV module cost vs installed system cost vs LCOE). This matters because neither Wright's law nor Moore's law have nice aggregative properties. If they are true at one level, they would in general not hold true at another (see e.g. Nordhaus (2014)). In practice, intuition appears to dictate primary data collection, and data availability dictates data analysis.

Here our discussion is based on thinking about performance as unit costs, and about experience as cumulative production, as in Nagy et al. (2013).

2.2 Moore's law, Wright's law, and Sahal's identity

Many studies have observed that technical performance or unit cost tend to follow an exponential trend[3]. Denoting unit cost of technology i by c, we have

$$c_t = c_0 \exp(\mu t), \tag{1}$$

where μ is the rate of exponential growth, c_0 is a constant, and t is time. Because Gordon Moore's predicted the number of transistors on integrated circuits would double every two years, Nagy et al. (2013) called this Moore's law. Similarly, since Wright's study of airplane manufacturing, many studies have observed unit costs falling with cumulative production. Denoting production by q, cumulative production z is

$$z_t = \sum_{j=t_0}^{t} q_j, \tag{2}$$

where t_0 is the time at which production started.[4] Then, "Wright's law" is

$$c_t = c_0 z_t^{\omega}, \tag{3}$$

where ω is the experience exponent[5]. Sahal (1979) noted that Moore's law and Wright's law are highly related, in the sense that if experience grows exponentially at rate r,

$$z_t = z_0 \exp(rt), \tag{4}$$

[3] See Nagy et al. (2011) for a study a detection of super-exponential trends in information technologies.

[4] An important issue empirically is that one typically observes production only several years after t_0. It is then necessary to estimate initial cumulative production to construct experience. When production grows exponentially at rate g, one may use the formula $z_{\text{init}} = q_{\text{init}}/\hat{g}$ (Lafond et al., 2018). Otherwise, one does need to try to carefully reconstruct the initial stock; see Lafond et al. (2022) for a practical example using WWII military equipment.

[5] The parameter ω can be thought of as an elasticity. The literature also often refers to the "learning rate" (LR), which represents the percentage decline on cost for a doubling a experience. If experience doubles, the new cost c' is $c' = c_0(2z)^{\omega} = 2^{\omega}c$, so the learning rate is usually defined as $LR \equiv 1 - 2^{\omega}$. For instance, a rule of thumb is that for $\omega \approx -1/3$ (roughly what has been found in the classic cases of WWII airplanes, or solar PV), costs fall by around 20% for each doubling of experience.

substituting Eq. 4 into Eq. 3 gives Eq. 1 with $\mu = \omega r$. The growth rate of cost is, trivially, the elasticity of cost to experience times the growth rate of experience.

Nagy et al. (2013) pioneered a systematic study of Moore's and Wright's laws, building the Santa Fe Performance Curve Database, demonstrating that the relationship $\mu = \omega r$ holds well when the three parameters are estimated independently. They showed that in practice, Moore's law and Wright's laws tend to perform relatively similarly out-of-sample, and using an encompassing model $c_t = c_0 z_t^{\omega} \exp(\mu t)$ does not work very well. The next section will explain why, providing considerable details on when *stochastic* Wright's law and Moore's law are observationally equivalent, what we can and cannot deduce from this, and why it is so hard to test which one is best.

2.3 Statistical structure and distributional forecasts

So far, we have assumed that the relationships described are deterministic. It is crucial to determine a sound statistical basis for these models, not only to get better parameter estimates, but also because the statistical structure strongly determines the properties of the forecast errors, and therefore, of prediction intervals. Farmer and Lafond (2016) and Lafond et al. (2018) addressed this problem for Moore's and Wright's law respectively. Appendix A provides an in-depth analysis of the distribution of forecast errors, including proofs of all the results stated here.

2.3.1 Moore's law

For Moore's law, an option would be a "deterministic trend" or "mean reversion" model such as $\ln c_t = \alpha + \mu t + \eta_t$, perhaps with autocorrelated η_t. Instead, Farmer and Lafond (2016) propose to use a geometric random walk with drift, where the growth rate at each period is drawn from a distribution centered on μ,

$$\Delta \ln c_t \equiv \ln c_t - \ln c_{t-1} = \mu + \eta_t, \tag{5}$$

where for simplicity $\eta_t \sim$ i.i.d.$\mathcal{N}(0, \sigma_\eta^2)$. The parameter estimates $\hat{\mu}$ and $\hat{\sigma}_\eta^2$ are simply the sample mean and variance of the growth rates. I denote the number of time periods by $m+1$, and thus the sample size (the number of growth rates) by m.

Having estimated the average growth rate, our best forecast is just that this growth rate will repeat every period, so, for τ-steps ahead,

$$\ln \hat{c}_{m+\tau} = \ln c_m + \hat{\mu}\tau. \tag{6}$$

If the Data Generating Process (DGP) is indeed Eq. 5, and we have made forecasts using (6), we can calculate explicitly the forecast errors

$$\mathcal{E}_\tau \equiv \ln c_{m+\tau} - \ln \hat{c}_{m+\tau} \sim \mathcal{N}\Big(0, \mathrm{Var}(\mathcal{E})\Big),$$

where

$$\mathrm{Var}\Big(\mathcal{E}_\tau\Big) = \sigma_\eta^2 \Big(\tau + \frac{\tau^2}{m}\Big). \tag{7}$$

Knowing the distribution of the errors is useful because we can make a distributional forecast as

$$\Pr(\ln \tilde{c}_{m+\tau}) = \ln \hat{c}_{m+\tau} + \Pr(\mathcal{E}) \sim \mathcal{N}\Big(\ln c_m + \hat{\mu}\tau, \sigma_\eta^2 \Big(\tau + \frac{\tau^2}{m}\Big)\Big). \tag{8}$$

The formula for the variance of forecast errors, Eq. 7, is well worth pausing. It shows that (squared) errors (i) do not depend on the drift, (ii) increase with volatility, (iii) increase with the forecast horizon, and (iv) are composed of two terms. The linear term τ is due to the fact that

unforecastable noise accumulates in the future. The second term is due to error in estimating the drift; while it tends to zero as sample size m increases, for a small sample size and long horizon it is likely to be the dominant source of errors - more precisely, when $\tau^2/m > \tau$, that is, when the horizon is greater than the sample size. Perhaps counter-intuitively, the further ahead in the future we try to predict, the more the error is dominated by mistakes in understanding the past, rather than by unpredictable future shocks.

Eq. 7 is also extremely useful for model validation, because if different technologies have different volatilities $\sigma_{\eta,i}$, we can renormalise the errors as $\mathcal{E}_{i,\tau}/\sigma_{\eta,i}$ to get a distribution that is, in principle, the same for all technologies, so that forecast errors for all technologies can be pooled together for testing the validity of the model. Further, we can also construct renormalised errors $\mathcal{E}_{i,\tau}/\left[\sigma_{\eta,i}\sqrt{\tau+\frac{\tau^2}{m}}\right]$ to pool together errors at various forecast horizons.

In Farmer and Lafond (2016), we computed empirical forecast errors on more than 50 technologies, using various sample sizes and forecast horizons, and used the normalisation above to pool the errors and test whether they followed the expected distribution. We found that they did not; however, if instead of assuming i.i.d. noise, we assumed an MA(1) process $\eta_t = \epsilon_t + \rho\epsilon_{t-1}$ with i.i.d. ϵ, the empirical forecast errors were well aligned with what we expected them to be if the DGP was true. Under that null hypothesis, the formula for the variance of the forecast errors is a little unsightly (Appendix A), but its large τ, large m approximation reads

$$\mathrm{Var}\Big(\mathcal{E}_\tau\Big) \approx \sigma_\eta^2 \frac{(1+\rho)^2}{1+\rho^2}\Big(\tau + \frac{\tau^2}{m}\Big). \tag{9}$$

To test this using empirically constructed forecast errors, in Farmer and Lafond (2016) (and in Lafond et al. (2018) for Wright's law), we faced two additional issues. First, when forecast errors are computed by rolling windows, they are correlated, and this is not taken into account into the formulas. Second, in practice σ_η^2 and ρ need to be estimated. To tackle these two issues, instead of testing whether empirically derived forecast errors have the theoretically predicted distribution, we tested whether empirically derived forecast errors on real data have the same distribution as empirically derived forecast errors on surrogate datasets, that is, datasets with the same unbalanced panel structure and generated using the assumed DGP with parameters as estimated on the real data.

For the parameter ρ, which is very hard to estimate on short time series, we assumed a "universal" value and found that using $\rho = 0.6$ allowed a good match between the distributions of forecast errors on empirical vs surrogate datasets.

While the method for making these distributional forecasts has been derived from substantial analysis and testing, ultimately it is extremely easy to implement - it requires estimating the mean and variance of the growth rates of the data, and code one formula: the pdf of the normal distribution with mean equal to the point forecast (Eq. 6), and variance equal to the variance of forecast errors (Eq. 9, or rather the exact version in Appendix A).

One can even *make a distributional forecast after observing only two values for costs*: with two observations, it is possible to compute $\hat{\mu} = \ln(y_m/y_0)/m$, and knowing $\hat{\mu}$ one can guess the standard deviation using the relation $\hat{\sigma}_\eta = 0.02 - 0.76\hat{\mu}$, which was empirically derived by Farmer and Lafond (2016) ($R^2 = 0.87$; they also report a log-log fit).

Before I turn to Wright's law, one final remark. The prediction Eq. 8 is for the *logarithm* of the cost, or, since the distribution is normal, the median of the distributional forecast for cost[6]. What

[6]To be clear, this is a distributional forecast for the concept of cost that underlies the data. Usually, the data is for *average* costs, so the point forecast is the median of the predicted distribution of average costs in the future. Put differently, the average cost is observed as a scalar and treated as a known value, but an interesting avenue for further research would

about the expected cost? Using properties of lognormals,

$$E[\hat{c}_{m+\tau}] = \hat{c}_m \exp\left(\hat{\mu}\tau + \frac{1}{2}\sigma_\eta^2\left(\tau + \frac{\tau^2}{m}\right)\right). \tag{10}$$

Thus, as the forecast horizon goes to infinity, no matter how negative the drift is, the mean of the distribution of predicted values for cost is (positive) infinity, a somewhat disturbing result. If the estimated drift is not negative enough, $(-\hat{\mu}) < \sigma_\eta^2/2$, then the expected value of future cost is always higher than today's cost and keeps rising. If it is negative enough, $(-\hat{\mu}) > \sigma_\eta^2/2$, expected cost goes down initially but plateaus at some forecast horizon $\tau^* = m((-2\hat{\mu} - \sigma_\eta^2)/(2\sigma_\eta^2)$, reaches the same level as today's costs at $2\tau^*$, and exceeds it more and more after that. In practice, it is better to think in terms of point forecasts (the median) and forecast error uncertainty (from the percentiles); the mean itself drifts further and further away into upper percentiles - it is not a good indicator of "central tendency" (see Mandelbrot (1997) for an insightful discussion of the moments of lognormals). It is possible that for long term forecasts, a model with mean reversion would be more appropriate, but the data available so far has not allowed us to explore this.

2.3.2 Wright's law

Typically, learning curves are estimated using the "level" model, $\ln c_t = \alpha + \omega \ln z_t + \epsilon$. As we did in Farmer and Lafond (2016) for the Moore's law model, in Lafond et al. (2018) we proposed to use instead a model where the noise follows a random walk, that is

$$\Delta \ln c_t = \omega \Delta \ln z_t + \eta_t, \tag{11}$$

where again for simplicity $\eta_t \sim$ i.i.d.$\mathcal{N}(0, \sigma_\eta^2)$. The parameters ω and σ_η can be estimated by OLS (being careful to run a regression through the origin, that is, without intercept). The forecasts are then made *conditional* on future values of experience,

$$\ln \hat{c}_{m+\tau} = \ln c_m + \hat{\omega}\Big(\ln z_{m+\tau} - \ln z_m\Big), \tag{12}$$

where $\hat{\omega}$ is the estimated value of ω. This formulation of the stochastic model has the advantage, like for Moore's law as a random walk, that the forecast origin corresponds to the actual observed datapoint (setting $\tau = 0$ gives $\ln \hat{c}_m = \ln c_m$). The forecast errors are again mean zero and

$$\mathrm{Var}\Big(\mathcal{E}_\tau\Big) = \sigma_\eta^2\left(\tau + \frac{\left(\sum_{i=m+1}^{m+\tau} \Delta \ln z_i\right)^2}{\sum_{i=0}^{m} (\Delta \ln z_i)^2}\right). \tag{13}$$

If experience grows deterministically, $\Delta \ln z_t = r, \forall t$, and we see again a manifestation of "Sahal's identity", as Wright's law Eq. 11 becomes Moore's law Eq. 5 with $\mu = \omega r$ and the same noise terms – Moore and Wright's law are observationally equivalent; the forecast errors are also the same, as Eq. 13 becomes Eq. 7. An insightful way to rewrite Eq. 13 is

$$\mathrm{Var}\Big(\mathcal{E}_\tau\Big) = \sigma_\eta^2\left(\tau + \frac{\tau^2}{m} \cdot \frac{\hat{r}_{(f)}^2}{\hat{\sigma}_{z,(p)}^2 + \hat{r}_{(p)}^2}\right), \tag{14}$$

be to model the empirical distribution of costs. So far, too little data has been available. This may change as micro-data underlying aggregate price indices becomes available.

where $\hat{\sigma}^2_{z,(p)}$ and $\hat{r}_{(p)}$ are the sample mean and variance of experience growth rates in the past data $(0 \dots m)$, and $\hat{r}_{(f)}$ is the sample mean of future experience growth rates.

This can be compared directly to the equivalent for Moore's law, Eq. 7, where the only difference is the final ratio within the parenthesis. It is instructive to look at it in detail. First, note that it is associated with the τ^2/m term, that is, it is linked to forecast errors resulting from errors in parameter estimation. Second, the fluctuations of experience (high $\sigma^2_{z,(p)}$) reduce the forecast errors. This is because a high variance of the regressor makes the estimates of the slope more precise (lower standard errors of ω). Third, a high growth rate of experience in the past (high $r_{(p)}$) also reduces the forecast errors, as experience spans a larger range, making estimation of the slope easier. Fourth, when experience in the future grows very fast (high $r_{(f)}$), costs are going down faster so the target to predict is "further away" on the learning curve, and therefore any error in the estimated slope leads to larger error (said differently, if experience does not increase much it is easy to predict that costs don't change much). Finally, since the predictions are conditional on observed experience in the future, fluctuations around the future trend (which I would have denoted $\sigma_{z,(f)}$) do not appear in the formula, as expected.

This discussion is helpful to understand whether and when, assuming that Wright's law is correct, we expect Wright's law to actually perform better than Moore's law out-of-sample. This will be the case when there have been a lot of accumulated experience in the past (fast growth, i.e. high $r_{(p)}$), but not a strong acceleration in the future (low $r_{(f)}$); and, crucially, cumulative production growth rates should have fluctuated a lot (high $\sigma_{z,(p)}$). This latter point is critical because this is typically not the case, simply because experience is a *cumulative* variable. Indeed, in Lafond et al. (2018) we were able to derive the following important result. Let production follow a geometric random walk with drift g and volatility σ_q. Then, cumulative production (experience) has the same average growth rate $(E[\Delta \ln z] \equiv r \approx q)$[7], but its volatility is

$$\sigma_z^2 \approx \sigma_q^2 \tanh(g/2) \approx \sigma_q^2 (g/2), \tag{15}$$

where tanh is the hyperbolic tangent function, and the second approximation is very good for growth rates of the magnitude typically observed (less than 0.3, say). This is a remarkable formula as it makes it clear that the volatility of cumulative production is expected to be very small; when production grows at 5% per year, the volatility of cumulative production is $\sqrt{0.05/2} \approx 16\%$ of the volatility of production, more than six times smaller. This approximation holds very well for $\sigma_q \geq 0.1$, see Appendix B for further discussion of this and related formulas, which have applicability well beyond experience curves. The appendix also shows that, perhaps surprisingly, while cumulative production is much less volatile, it is as "uncertain" as production, in the sense that its cross-sectional ("ensemble") dispersion is almost the same.

For completeness, in Lafond et al. (2018) we found that to give a good fit to the forecast errors, as in the case of Moore's law, we had to extend the model to allow for (moving average) autocorrelation. In this case, it is possible to derive a formula that extends Eq. 14 in the same way that for Moore's law, Eq. 9 extends Eq. 7. More specifically, as derived in the Appendix,

$$\mathrm{Var}\left(\mathcal{E}_\tau\right) = \sigma_\eta^2 \frac{(1+\rho)^2}{1+\rho^2} \left(\tau + \frac{\tau^2}{m} \cdot \frac{\hat{r}^2_{(f)}}{\hat{r}^2_{(p)} + \hat{\sigma}^2_{z,(p)}} \right), \tag{16}$$

where ρ is the moving average autocorrelation parameter. Appendix A shows the exact formula and discusses implementation. In Lafond et al. (2018), based on testing on pooled data, we suggested using $\rho = 0.19$.

[7] This reflects the fact that the exponential function is defined as the only one whose derivative is itself.

2.4 Causality

An important criticism of experience curves is that they are used to predict costs conditional on experience, thereby assuming a clear causal relation, while they may only be a statistical relationship with no causal underpinning. While it is plausible that output (cumulative or not) causes lower cost, it is also very plausible that lower costs drive demand up. To examine this issue, we have looked at military production during World War II, when it was clear that the demand for weapons was driven by battlefield needs, relatively independently of cost considerations. But before diving into this, let us discuss a more direct source of concern: spurious regressions.

Spurious regressions. We have assumed that log costs follow a random walk, that is, are "integrated of order 1", I(1). Testing for unit roots is fraught with difficulties, particularly in small samples, but this appears to be a reasonable assumption. Similarly, it is reasonable to assume that log experience is I(1). It is well known that, when two variables are I(1), regressing one on the other in levels will produce "non-sense", that is, it will pick up a highly significant relationship between the two, with high R^2, even when none exists.

More formally, when the random walks have drifts, Entorf (1997) showed that the coefficient tends to the ratio of the drifts. If costs have a drift μ and experience has a drift r, and if they are completely independent, running the regression $\ln c_t = \alpha + \omega \ln z_t + \epsilon$ will give $\hat{\omega} \approx \mu/r$ even though by assumption $\omega = 0$.

The well-known solution is to estimate the relation in first differences, but *with an intercept*. In other words, running the regression $\Delta \ln c_t = \alpha + \omega \Delta \ln z_t + \epsilon$ will produce $\hat{\omega} \approx 0$, as desired. But there is an important problem with this solution. Adding an intercept corresponds to estimating another model, in fact a mix of Moore and Wright's law, of the form $c_t = c_0 e^{\mu t} z_t^{\omega}$. In principle, adding an irrelevant variable does not bias OLS estimates, but it makes them less precise; this turns out to be a huge problem here, because, as discussed in the previous section, $\Delta \ln z_t$ has very low variance, so this regression will typically have huge standard errors and be untrustworthy. What if we omit the intercept, then?

In this case, one can show (Appendix C) that the expected value of the first-difference estimator *without a constant* is $\hat{\omega} \approx \frac{\mu/r}{1+(\sigma_z/r)^2}$. Thus, the estimate $\hat{\omega}$ goes to the "spurious" result $\hat{\omega} = \mu/r$ as the volatility of experience $\sigma_z \to 0$, and goes to the "correct" result $\hat{\omega} = 0$ if $\sigma_z \to \infty$. If the variance of the growth rate of cumulative production is high, we can safely do a regression in first differences without a constant; but as we have emphasized, σ_z tends to be very low. This makes it almost impossible to test whether the relation is spurious ($\omega = 0$): first differences regressions with a constant do not produce a definite answer because the standard errors are too large, first difference regressions without a constant produce an answer biased toward the spurious regression result.

Of course, an alternative solution would exist if the two time series were co-integrated, but we have not found good evidence that this is the case (although acknowledging again the difficulty of testing in small samples). In any case, from a theoretical point of view, there are no a priori good reasons to expect a co-integrated relationship, which would assume that there is an intrinsic relationship between the variables in level, so that short-term departures from this long-run relation would be corrected over time[8]. Instead, because experience curves may be based on learning, there are good reasons to think that a first difference relationship is appropriate. What we are postulating here is a relationship between technological *progress* (the *growth rate* of unit costs, not unit costs), and the *accumulation* of experience (not the *level* of experience per se).

[8]Note also that the growth rate of experience should remain non-negative, so the error-correction equations may need to be amended.

While this discussion gives a somewhat disillusioned treatment of "learning curves", it turns out that a deeper concern (reverse causality) has led us on a path toward datasets that do have high fluctuations of experience, and thus make it possible to also estimate a separate exogenous time trend.

Are learning curves just flipping the axes of a demand curve? Nordhaus (2014) criticized learning curves by noting the issue of reverse causality. Let's assume that unit costs depend on both experience and an exogenous time trend, with the form

$$c_t = c_0 z_t^{-b} e^{-at}. \tag{17}$$

The parameters a and b are expected to be positive. Assume that production is equal to demand and that price is equal to unit cost, and let demand be a constant elasticity function of cost, with an additional exponentially growing exogenous demand at rate d,

$$q_t = D_t = D_0 c_t^{-\epsilon} e^{dt}, \tag{18}$$

with demand elasticity $\epsilon > 0$. Taking logs and first differences of Eqs. 17 and 18 , assuming $\Delta \ln q \approx \Delta \ln z$, and solving the system gives

$$\Delta \ln c = \frac{-a - bd}{1 - b\epsilon}, \tag{19}$$

$$\Delta \ln z = \Delta \ln q = \frac{a\epsilon + d}{1 - b\epsilon}. \tag{20}$$

Nordhaus (2014)'s critique is that experience curve studies typically assume that experience is the single explanatory factor for costs, that is they assume that

$$\omega = \frac{\Delta \ln c}{\Delta \ln z} = \frac{-a - bd}{a\epsilon + d}, \tag{21}$$

it is clear that in general $\hat{\omega}$ does not identify the effect of experience, or "learning" parameter $-b$. However, we can recover b from simple regressions in two cases. First, if there is no exogenous technological progress ($a = 0$), then Eq. 21 becomes $\omega = -b$. Second, if demand is exogenous ($\epsilon = 0$), then Eq. 21 becomes $\omega = -(a/d) - b$, that is, since in this case $\Delta \ln z = b$, $\Delta \ln c = -a - b\Delta \ln z$. In other words, if demand is exogenous a simple regression of costs on experience will retrieve both the rate of exogenous technological progress and the experience coefficient.

In Lafond et al. (2022), we studied learning curves during World War II, following a long tradition but with a renewed emphasis on the specific "natural experiment" aspect of this data: we can argue that demand was "exogenous", dictated by battlefield needs. While there may have been some substitutions between different plants or kinds of weapons depending on their changing relative costs, overall, costs were not a big factor overall in deciding output, as the goal was often simply to produce as much as possible. While the war solves the issue of reverse causality, exogenous demand in such a context has another benefit: it is far from simply exponential, exhibiting an increase, plateau, and decline as the war ended. This means that the volatility of production and cumulative production tended to be higher than in modern datasets, allowing a better identification.

Lafond et al. (2022) studied three different datasets: one dataset with plant-level data and long time series, but limited to a few categories of products; one with almost all products, but with only two time-series points per product; and one with long time series covering all kinds of products, but very aggregated. None of these datasets is perfect, but they differ in their drawbacks and advantages. Lafond et al. (2022) estimate $\Delta \ln c_{it} = \alpha + \beta \Delta \ln z_{it} + \epsilon$, that is, assuming constant exogenous and endogenous learning parameters across technologies. The key question is: in a context where demand

is exogenous, and the variance in the data makes it possible to estimate both α ad β, is $\beta \neq 0$? In all three datasets, they found that roughly $\hat{\alpha} \approx \Delta \ln c_{it}/2$, that is, exogenous technological progress accounts for about only half of technological progress - the rest can be attributed to the effects of experience.

2.5 Moore's law vs Wright's law

While this is good evidence of some causality from experience to costs, defenders of Moore's law over Wright's law have found evidence of cases where costs decreased in absence of increased experience (Funk and Magee, 2015).

It is unclear how externally valid the World War II results are, and whether pre-commercial rapid improvement implies that a pure Moore's law would be more predictive than Wright's law once deployment has started. However, given the "impossibility" of separating the effect of a time trend from that of deployment in most existing datasets, I think that the "half-half" result provides a useful prior; in fact, rather than choosing between Moore and Wright, it may be useful to attempt an estimation of the mixed model using the half-half result to center a prior in a Bayesian setting.

Moore's law is much simpler to implement. To recall, using Wright's law requires knowing not only some of the past data on production, but in fact having an estimate of initial cumulative production, and a good prediction for future production.

The case for Wright's law rests on its usefulness for policy evaluation, to capture the endogenous effects of our decisions on technology costs. A case in point is the construction of technology portfolio.

2.6 Technology portfolios

An important application of stochastic experience curves is for constructing technology portfolios. Consider a set of technologies that produce a perfectly substitutable product, and follow Wright's law but perhaps with different parameters. Now imagine that a policy maker has a fixed budget to invest. Investing in a technology brings down its cost, so they have an incentive to invest everything into one technology - this is classic "increasing returns imply lock-in" argument (Arthur, 1989), well-known in the literature on technological change. However, because experience curves are stochastic, a risk-averse decision maker may prefer to hedge their bets and invest in different technologies, as their "noise realizations" will average out - this is a classic "diversification reduces risks" argument (Markowitz, 1952), well-known in the finance literature.

In Way et al. (2019), we studied this trade-off theoretically, in the simple case of two technologies for two periods. The solution exhibits a cusp bifurcation. When risk aversion is high, it is always better to diversify - there is a single local minimum, and it is for a diversified portfolio. But when risk aversion is low, there are two local minima, one for each technology being highly dominant in the portfolio.

In perhaps the most useful and comprehensive application of stochastic experience curves so far, Way et al. (2022) compared the costs of various energy portfolios. Energy systems are complicated, as different sources are not perfectly substitutable (mostly due to intermittency and the need for electricity storage, and the difference between primary, final and useful energy), so optimizing is not practical. However, they showed that the expected present discounted cost of a fast transition scenario is trillions of dollars lower than that of a business as usual scenario - at least 5-15 trillion USD cheaper over a 50-year horizon for any plausible discount rate. This is because fossil fuels have a zero learning rate, while some renewables are very likely to become cheaper and cheaper as we invest in them.

One area that has received little attention is the issue of cross-sectional dependence. In Way et al. (2019), we showed that, as expected, increasing correlation between the noise of competing technologies limits the benefits of diversification. However, real data is so scarce that there has been very limited empirical investigation of this issue.

3 Ongoing work and research agenda

3.1 Cost curves

An important issue with cost curves is that they are typically based on world-level data. For highly standardized and commoditized products that are manufactured and sold globally, such as solar cells, this is justified. However, in practice, what matters for users include other elements of costs which are substantially "local" in nature, such as installation cost, or, to some extent, the cost of capital. Baumgärtner and Farmer (2025) are investigating cost curves for solar PV and onshore wind at the national level, decomposing levelized costs of electricity into its components. Their key finding is that solar and wind technologies behave quite differently, with "balance-of-system" costs consistently declining for solar, but not for wind. They argue that this implies that wind energy will approach a floor cost around $35/MWh, whereas solar will continue declining for some time, reaching a likely cost around $3-15/MWh by mid-century.

There are many other interesting avenues for further research. A first is to consider panel data rather than independent time series, and model cross-sectional correlations. Very little has been done so far to evaluate the covariance matrix of technologies. This is relevant for forecasting, and also for technology portfolio construction, as positive correlations reduce the benefits of diversification (Way et al., 2019). Relatedly, for portfolios an interesting question arises of whether it is easier to predict the aggregate directly, or predict the components and aggregate the forecasts. The theoretical answer to this is that it depends on the statistical structure of the data (relative noise variances, weights, etc.), and therefore this should be examined empirically in specific cases.

A second is to predict observed *distributions*. Consider, for instance, hydropower: we may be able to observe the costs of several projects in a given year, and they may vary a lot. We might thus not be very interested in making a distributional for the average cost, but rather, we would like to predict the distribution of costs of various projects in the future.

A third avenue for research is to consider the associated decision theory seriously. The forecast methods presented so far are typically optimal under mean square loss for the log of the cost, essentially for convenience. However, decision makers may care more about the "risk" that carbon capture and storage does *not* make progress, or that artificial intelligence makes "too much" progress. The forecasting and decision theory literature offers tools to deal with asymmetric loss.

3.2 S-curve diffusion

We have seen that production and cumulative production often grow so closely to exponentially that this creates estimation issues. This is somewhat surprising, as a large literature has studied "S-curves" for technology diffusion, including a long research programme at the International Institute for Applied Systems Analysis to understand technology substitution in energy systems (e.g. Grübler et al. (1999)). Here I describe ongoing research efforts in our group to analyze technology diffusion. The classic approach is to assume a logistic curve, that is,

$$Y_t = \frac{L}{1 + \exp(-k(t - t_0))}, \tag{22}$$

where Y_t is the global stock of a technology in year t (e.g the global installed capacity of solar PV), L is the asymptotic level of diffusion, k is a parameter that tunes the speed of the diffusion, and corresponds to the rate of exponential growth in the initial phase, and t_0 is the year in which diffusion has reached 50% of the total, that is, one can easily check that $Y_{t_0} = L/2$. We have experimented with other S-curve models, starting from the most general (Richard's curve), and found that this classic, simple model works as well as more complex models, although the Gompertz model works well too.

Distributional forecasts. In Wagenvoort et al. (2024), we are developing a method to make distributional forecasts for logistic curves. This involves building a method to pool the data from various technologies so that we can create a good model for the structure of the errors; developing parameter estimation methods, as naive estimates of S-curve based on technologies that have not passed the inflexion point can be highly biased; and testing that out-of-sample distributional forecasts are well calibrated.

Is technology diffusion accelerating? Tankwaa et al. (2025) shows that the same countries always tend to be early adopters of new technologies, and that technology diffusion has accelerated over time, in the sense that more recent technologies tend to have a faster diffusion speed k.

In addition to these ongoing efforts, there are a number of further avenues for research. One is the hope that collecting data where diffusion is *not* exponential will help settle the Moore versus Wright debate. Another is the development of multivariate models, which predict both series simultaneously.

3.3 Patents

Patents are one of the most widely used indicators of technological change in the academic literature, because they constitute a large and very well documented record.

A particularly interesting feature of patent data is that it has a lot of metadata and relational data that can be used to perform network analysis. We briefly discuss technology codes and citations.

Technology codes and their combination. Patents are classified into one or more technology classes (or codes)[9], which can be used for several tasks. For instance, Strumsky and Lobo (2015) and Youn et al. (2015), document that while early patents tend to introduce new classes, more recent patents tend to introduce new recombinations of existing classes. A large literature has developed to quantify the novelty of patents or scientific publications based on how "atypical" code combinations are, which can help predict future impact (Kim et al., 2016) or future combinations (Tacchella et al., 2020; Shi and Evans, 2023). More generally, this points to the underlying idea that technologies (codes or patents) can be represented as points in a high-but-not-so-high-dimensional space where clusters and distances can be computed for downstream tasks, including prediction.

Citations. Another key metadata in patents is the citations between them, which makes it possible to again construct networks between patents or between patent categories. Citations were less common in the early decades of the patent system, and are thus harder to parse, so citation networks

[9]The classification system changes substantially over time, and patents are reclassified so that they are always all classified in the latest vintage of the classification system. Lafond and Kim (2019) document the development of the US patent classification system, propose a model for its development, provides examples of key changes, and explain how using the latest vintage of the system biases historical analysis.

typically cover only a more limited time span. Nevertheless, patent citations have been very popular for technology forecasting because they allow us to see an objective, direct relationship between patents; as in scientific papers, one may assume that highly cited patents are more influential. For instance, in Mariani et al. (2019) we showed that patents that are at the top of their cohort in terms of citations received early are more likely to become part of a list of "significant" patents, as defined by Strumsky and Lobo (2015).

In contrast to networks constructed based on co-classification, networks constructed based on patent citations have a clear direction (see Alstott et al. (2017) for a discussion of the various ways of constructing and normalizing patent networks), from the citing category to the cited category, suggesting that technologies rely on others for their development. This is appealing from a complex systems point of view, and indeed one can study, for instance, the auto-catalytic properties of this network (Napolitano et al., 2018). The idea is that if a technology is developing fast, the technologies that tend to depend on it will also be able to develop faster in the future. Thus, the structure of the network should help predict future patenting rates.

Acemoglu et al. (2016) constructed an influence network between technology classes based on citations and used this to predict future patenting rates. They predicted the number of future patents in a given technology class, showed that including past patenting rates from related classes improved their results, and argued that this demonstrates the existence of network effects.

Because patenting rates are very persistent, however, these results require careful interpretation. The number of patents that are granted in a given class in a given year tend to be very similar to the number of patents that were granted in that class in previous years. As is well known in the spurious regression literature, a persistent time series can be an excellent predictor of other persistent yet independent time series. Thus, to demonstrate the existence of network effects, one must first establish a careful benchmark. As shown in Pichler (2021), however, the network-based predictions of Acemoglu et al. (2016) are substantially worse than network-independent predictions made from simple random walks with drifts.

Motivated by this observation, in Pichler et al. (2020) we developed a network-based model and carefully established its predictive over and above that of a random walk. The model resembles a spatial econometrics model, and is very similar to production network models. The growth rate of a technology depends on the growth rates of its neighbors,

$$g_{it} = \beta \sum_j W_{ji,t} g_{jt} + \epsilon_{it} \Rightarrow \boldsymbol{g}_t = (\boldsymbol{I} - \beta \boldsymbol{W}_t')^{-1} \boldsymbol{\epsilon}_t \tag{23}$$

where $\mathbf{g}_t$ is a vector of growth rates of patents, and $\mathbf{W}_t$ is a matrix where column j gives the "recipe" of technology j, that is, the share of the citations it makes that go to each of the other technologies.

An issue with these models is that, if the shocks ϵ are independent across technologies, this assumes that all co-movement comes from the network. When there is observed co-movement, an alternative hypothesis that is equally compelling a priori but fundamentally different conceptually is that there is a common factor, exogenous to all technologies. The factor model says that all technologies move together because they are driven by a common external force, while the network model says that they move together because they are interdependent. Econometrically, common factors represent "strong" cross-sectional dependence because the dependence between two nodes does not really depend on how many other nodes are present, while network dependence is "weak" because two randomly chosen nodes are less likely to be close as network size goes to infinity. That said, details matter and it is usually difficult to distinguish between the two sources of co-movement.

Foerster et al. (2011) proposed an encompassing model, in the context of production networks, keeping Eq. 23 but assuming that the noise follows a factor model, $\boldsymbol{\epsilon}_t = \boldsymbol{\lambda} f_t + \boldsymbol{\nu}_t$, where $\boldsymbol{\lambda}$ is a vector of loadings, $\boldsymbol{\nu}_t$ is a vector of idiosyncratic shocks and f_t is a scalar, the common factor. An interesting

avenue for further research would be to investigate the source of cross-sectional dependence to establish whether predictability arises from common trends in the patent system, or from the endogenous dynamics inherent to systems with strong network-based interdependencies.

4 Conclusion – towards a Technology Observatory

In many cases, it is very hard to collect historical data on costs, but patent data exists and is straightforwardly available. Can we relate patent data on a technology with progress rates of that technology, for the cases where both are available, so that we can predict improvement rates in cases where only patent data is available? In a potentially important study, Triulzi et al. (2020) argue that centrality in a carefully crafted network (using citations, technology codes, and careful normalization) is predictive of the "Moore's law" technological progress rate. More generally, it is tempting to think that the positions of technologies in an abstract, latent technological space determines their progress rate.

An alternative approach to the same problem is to predict progress rates using properties of a technology; for instance, Malhotra and Schmidt (2020) classify technologies according to their degree of design complexity and the need for customization, arguing that this determines progress rates – actually, in their case, Wright's law exponents. Simple technologies like solar PV have higher learning rates than nuclear power, because their "granularity" implies large economies of scale and learning; when each project is huge, and highly specific, these benefits are harder to materialize.

This suggests an ambitious research agenda to construct a "Technology observatory", combining data on costs, deployment, technical performance, design characteristics, latent demand indicators, patents, scientific publications, products, associated companies, etc. This would make it possible to better understand the fundamental drivers of progress rates, why they are so diverse, and improve predictions when they are most helpful, that is, for technologies that are in very early stages of deployment.

This is likely to be a lively and exciting area of future research, as it is very important but remains difficult, with the application of new data and analytical methods having a genuine potential to improve over existing approaches.

References

Acemoglu, D., Akcigit, U., and Kerr, W. R. (2016). Innovation network. *Proceedings of the National Academy of Sciences*, 113(41):11483–11488.

Albright, R. E. (2002). What can past technology forecasts tell us about the future? *Technological Forecasting and Social Change*, 69(5):443–464.

Alstott, J., Triulzi, G., Yan, B., and Luo, J. (2017). Mapping technology space by normalizing patent networks. *Scientometrics*, 110:443–479.

Argote, L. and Epple, D. (1990). Learning curves in manufacturing. *Science*, 247(4945):920–924.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173.

Arrow, K. J. et al. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Arthur, B. (2025). Combinatorial evolution. In editor, editor, *The Economy as an Evolving Complex Systems IV*, chapter x, page x. SFI Press.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal*, pages 116–131.

Arthur, W. B. (2010). *The nature of technology: What it is and how it evolves*. Penguin UK.

Ayres, R. U. (1969). *Technological forecasting and long-range planning*. McGraw-Hill Book Company. https://archive.org/details/technologicalfor0000ayre_l4a1/.

Baumgärtner, L. and Farmer, J. (2025). Empirical basis for national renewable cost forecasts. *in progress*, x(x):x–x.

Bengio, Y. et al. (2023). Pause giant AI experiments: An open letter. Technical report, Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/.

Benson, C. L., Triulzi, G., and Magee, C. L. (2018). Is there a moore's law for 3d printing? *3D printing and additive manufacturing*, 5(1):53–62.

Ciarli, T., Coad, A., and Rafols, I. (2016). Quantitative analysis of technology futures: A review of techniques, uses and characteristics. *Science and Public Policy*, 43(5):630–645.

Clements, M. and Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge University Press.

Coyle, D. (2025). Why are new ideas getting harder to use? In editor, editor, *The Economy as an Evolving Complex Systems IV*, chapter x, page x. SFI Press.

Duran-Nebreda, S., Vidiella, B., , and Valverde, S. (2025). Coevolution of software and innovation: Constraints, tinkering and symbiosis. In editor, editor, *The Economy as an Evolving Complex Systems IV*, chapter x, page x. SFI Press.

Entorf, H. (1997). Random walks with drifts: Nonsense regression and spurious fixed-effect estimation. *Journal of Econometrics*, 80(2):287–296.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., and Zalányi, L. (2013). Prediction of emerging technologies based on analysis of the us patent citation network. *Scientometrics*, 95(1):225–242.

Farmer, J. D. and Lafond, F. (2016). How predictable is technological progress? *Research Policy*, 45(3):647–665.

Foerster, A. T., Sarte, P.-D. G., and Watson, M. W. (2011). Sectoral versus aggregate shocks: A structural factor analysis of industrial production. *Journal of Political Economy*, 119(1):1–38.

Frenken, K. and Neffke, F. (2025). Economic geography and complexity theory. In editor, editor, *The Economy as an Evolving Complex Systems IV*, chapter x, page x. SFI Press.

Funk, J. L. and Magee, C. L. (2015). Rapid improvements with no commercial production: How do the improvements occur? *Research Policy*, 44(3):777–788.

Gilfillan, S. C. (1937). Prediction of inventions, the. *J. Pat. Off. Soc'y*, 19:623.

Goldin, I., Koutroumpis, P., Lafond, F., and Winkler, J. (2024). Why is productivity slowing down? *Journal of Economic Literature*, 62(1):196–268.

Gordon, R. J. (1990). *The Measurement of Durable Goods Prices*. University of Chicago Press.

Grace, K., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., and Brauner, J. (2024). Thousands of ai authors on the future of AI. *arXiv preprint arXiv:2401.02843*.

Grübler, A., Nakićenović, N., and Victor, D. G. (1999). Dynamics of energy technologies and global change. *Energy policy*, 27(5):247–280.

Halawi, D., Zhang, F., Yueh-Han, C., and Steinhardt, J. (2024). Approaching human-level forecasting with language models. *arXiv preprint arXiv:2402.18563*.

Harper, J. C. (2013). Impact of technology foresight. *Compendium of Evidence on the Effectiveness of Innovation Policy Intervention, Manchester Institute of Innovation Research, Manchester Business School, University of Manchester*.

Jantsch, E. (1967). *Technological forecasting in perspective*. OCDE.

Kahn, H. and Wiener, A. J. (1967). The next thirty-three years: A framework for speculation. *Daedalus*, pages 705–732.

Kim, D., Cerigo, D. B., Jeong, H., and Youn, H. (2016). Technological novelty profile and invention's future impact. *EPJ Data Science*, 5:1–15.

Koh, H. and Magee, C. L. (2006). A functional approach for studying technological progress: Application to information technology. *Technological Forecasting and Social Change*, 73(9):1061–1083.

Koh, H. and Magee, C. L. (2008). A functional approach for studying technological progress: Extension to energy technology. *Technological Forecasting and Social Change*, 75(6):735–758.

Lafond, F., Bailey, A. G., Bakker, J. D., Rebois, D., Zadourian, R., McSharry, P., and Farmer, J. D. (2018). How well do experience curves predict technological progress? a method for making distributional forecasts. *Technological Forecasting and Social Change*, 128:104–117.

Lafond, F., Greenwald, D., and Farmer, J. D. (2022). Can stimulating demand drive costs down? world war ii as a natural experiment. *The Journal of Economic History*, 82(3):727–764.

Lafond, F. and Kim, D. (2019). Long-run dynamics of the us patent classification system. *Journal of Evolutionary Economics*, 29(2):631–664.

Malhotra, A. and Schmidt, T. S. (2020). Accelerating low-carbon innovation. *Joule*, 4(11):2259–2267.
Mandelbrot, B. (1997). A case against the lognormal distribution. *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk. Selecta Volume E*, pages 252–269.
Mariani, M. S., Medo, M., and Lafond, F. (2019). Early identification of important patents: Design and validation of citation network metrics. *Technological forecasting and social change*, 146:644–654.
Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.
Martino, J. P. (1993). *Technological forecasting for decision making*. McGraw-Hill, Inc.
Meng, J., Way, R., Verdolini, E., and Diaz Anadon, L. (2021). Comparing expert elicitation and model-based probabilistic technology cost forecasts for the energy transition. *Proceedings of the National Academy of Sciences*, 118(27):e1917165118.
Nagy, B., Farmer, J. D., Bui, Q. M., and Trancik, J. E. (2013). Statistical basis for predicting technological progress. *PloS one*, 8(2):1–7.
Nagy, B., Farmer, J. D., Trancik, J. E., and Gonzales, J. P. (2011). Superexponential long-term trends in information technology. *Technological Forecasting and Social Change*, 78(8):1356–1364.
Napolitano, L., Evangelou, E., Pugliese, E., Zeppini, P., and Room, G. (2018). Technology networks: the autocatalytic origins of innovation. *Royal Society Open Science*, 5(6):172445.
Neffke, F., Sbardella, A., Schetter, U., and Tacchella, A. (2025). Economic complexity analysis. In editor, editor, *The Economy as an Evolving Complex Systems IV*, chapter x, page x. SFI Press.
Nordhaus, W. D. (2014). The perils of the learning model for modeling endogenous technological change. *The Energy Journal*, 35(1):1–14.
Pfeiffer, A., Hepburn, C., Vogt-Schilb, A., and Caldecott, B. (2018). Committed emissions from existing and planned power plants and asset stranding required to meet the paris agreement. *Environmental Research Letters*, 13(5):054019.
Philippon, T. (2023). Additive growth. Technical report, https://pages.stern.nyu.edu/~tphilipp/papers/AddGrowth_macro.pdf.
Pichler, A. (2021). *Network-dependent dynamics of innovation and production*. PhD thesis, University of Oxford.
Pichler, A., Lafond, F., and Farmer, J. D. (2020). Technological interdependencies predict innovation dynamics. *arXiv preprint arXiv:2003.00580*.
Porter, A. L. and Cunningham, S. W. (2005). Tech mining. *Competitive Intelligence Magazine*, 8(1):30–36.
Sahal, D. (1979). A theory of progress functions. *AIIE Transactions*, 11(1):23–29.
Sampson, M. (1991). The effect of parameter uncertainty on forecast variances and confidence intervals for unit root and trend stationary time-series models. *Journal of Applied Econometrics*, 6(1):67–76.
Schoenegger, P., Tuminauskaite, I., Park, P. S., and Tetlock, P. E. (2024). Wisdom of the silicon crowd: Llm ensemble prediction capabilities match human crowd accuracy. *arXiv preprint arXiv:2402.19379*.
Shi, F. and Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14(1):1641.
Solé, R. V., Valverde, S., Casals, M. R., Kauffman, S. A., Farmer, D., and Eldredge, N. (2013). The evolutionary ecology of technological innovations. *Complexity*, 18(4):15–27.
Strumsky, D. and Lobo, J. (2015). Identifying the sources of technological novelty in the process of invention. *Research Policy*, 44(8):1445–1461.
Tacchella, A., Napoletano, A., and Pietronero, L. (2020). The language of innovation. *PloS one*, 15(4):e0230107.
Tankwaa, B., Bassat, L. V., Barbrook-Johnson, P., and Farmer, J. D. (2025). Increasing inequality between countries in key renewable energy costs. *INET WP 2025-01*.
Tetlock, P. E. and Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.
Triulzi, G., Alstott, J., and Magee, C. L. (2020). Estimating technology performance improvement rates by mining patent data. *Technological Forecasting and Social Change*, 158:120100.
Verdolini, E., Anadón, L. D., Baker, E., Bosetti, V., and Reis, L. A. (2018). Future prospects for energy technologies: insights from expert elicitations. *Review of Environmental Economics and Policy*.
Wagenvoort, B., , Dyer, J., Lafond, F., and Farmer, J. D. (2024). Universality and predictability of technology diffusion. *in progress*, x(x):x–x.
Way, R., Ives, M. C., Mealy, P., and Farmer, J. D. (2022). Empirically grounded technology forecasts and the energy transition. *Joule*, 6(9):2057–2082.

Way, R., Lafond, F., Lillo, F., Panchenko, V., and Farmer, J. D. (2019). Wright meets markowitz: How standard portfolio theory changes when assets are technologies following experience curves. *Journal of Economic Dynamics and Control*, 101:211–238.

Youn, H., Strumsky, D., Bettencourt, L. M., and Lobo, J. (2015). Invention as a combinatorial process: evidence from us patents. *Journal of the Royal Society interface*, 12(106):20150272.

Zadourian, R. (2018). *Model-based and Empirical Analyses of Stochastic Fluctuations in Economy and Finance*. PhD thesis, Technische Universität Dresden.

Zadourian, R. and Klümper, A. (2018). Exact probability distribution function for the volatility of cumulative production. *Physica A: Statistical Mechanics and its Applications*, 495:59–66.

Appendix (Online only)

The code to reproduce the figures in this appendix (and a little more) can be downloaded from https://francoislafond.info/research/

A Forecast errors

We consider the general model

$$\ln \frac{c_t}{c_{t-1}} = \omega \ln \frac{z_t}{z_{t-1}} + \eta_t, \tag{24}$$

$$\eta_t = \epsilon_t + \rho \epsilon_{t-1}, \tag{25}$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2). \tag{26}$$

For future reference, the variance of η is

$$\sigma_\eta^2 = (1+\rho^2)\sigma_\epsilon^2. \tag{27}$$

Throughout, we will assume for simplicity that we know the autocorrelation parameter ρ and the variances. While we will carefully distinguish between ω and its estimated counterpart $\hat{\omega}$, we will always use σ^2 and ρ without a hat, and use true values in our simulations.

Model	Experience	Autocorrelation	Section
Moore	$\Delta \ln z_t = r$	$\rho = 0$	A.1
Moore + MA(1)	$\Delta \ln z_t = r$	$-1 < \rho < 1$	A.2
Wright	unrestricted	$\rho = 0$	A.3
Wright + MA(1)	unrestricted	$-1 < \rho < 1$	A.4

Table 1: Restrictions of the general model Eqs. 24-25 in each specific model, and associated Section.

Table 1 shows the various restrictions we consider below. It is useful to start from the simple random walk with drift with no autocorrelated noise, and work our way up.

Throughout, we assume that the process starts at $t = 0$ (although noise starts at $t = -1$), and we observe data for $m+1$ periods, that is, until $t = m$. We normalise to $c_0 = z_0 = 1$. We seek to predict τ steps ahead, to $t = m + \tau$.

This is a detailed treatment of highly specific cases; in particular, even when we assume that the noise is autocorrelated, we do not use this to make better forecasts. See Sampson (1991) and Clements and Hendry (1998) for more general treatments.

A.1 The random walk with drift a.k.a. "Moore's law"

We call "Moore's law" the random walk with drift. This corresponds to Eq. 24 where experience grows at exactly the same rate every period, $\Delta \ln z_t = r$. In other words, if experience grows at exactly the same rate every period, its effects are indistinguishable from simply time ticking.

Starting from the case with no autocorrelation, Eqs. 25 with $\rho = 0$, our model here is

$$\ln c_t - \ln c_{t-1} \equiv Y_t = \mu + \eta_t, \tag{28}$$

where $\mu \equiv \omega r$. We estimate it from a sample of the first $m+1$ periods (i.e. m growth rates) as the simple mean

$$\hat{\mu} = \frac{1}{m}\sum_{i=1}^{m} Y_i = \mu + \frac{1}{m}\sum_{i=1}^{m} \eta_i \sim \mathcal{N}\left(\mu, \frac{\sigma_\eta^2}{m}\right), \tag{29}$$

where the last step follows from a direct application of rules for summing up independent random normal variables.

Considering the data generating process (28), since future noise is not predictable and since we do not know μ but we have the estimate $\hat{\mu}$ from (29), the forecast for τ-periods ahead is

$$\ln \hat{c}_{m+\tau} = \ln c_m + \hat{\mu}\tau. \tag{30}$$

The forecast error then is

$$\begin{aligned} \mathcal{E} \equiv \ln c_{m+\tau} - \ln \hat{c}_{m+\tau} &= (\mu - \hat{\mu})\tau + \sum_{i=m+1}^{m+\tau} \eta_i, \\ &\sim \mathcal{N}\left(0, \tau^2 \frac{\sigma_\eta^2}{m}\right) + \mathcal{N}\left(0, \tau\sigma_\eta^2\right), \\ &\sim \mathcal{N}\left(0, \sigma_\eta^2\left(\tau + \frac{\tau^2}{m}\right)\right). \end{aligned} \tag{31}$$

The third step follows from assuming independence, which is correct in this case: noise from the past, which affects parameter estimation, is independent from noise in the future. This calculation makes clear that the prediction variance is the sum of two terms - a term τ is associated with future noise, and a term τ^2/m is associated with parameter estimation; both are premultiplied by the variance of the noise.

To get a distributional forecast at horizon τ using data until time m, denoted $\mathcal{P}_{m,\tau}$, one usually uses the point forecast (30) plus the forecast error (31), that is

$$\mathcal{P}_{m+\tau} = \ln \hat{c}_{m+\tau} + \mathcal{E} \sim \mathcal{N}\left(\ln c_m + \hat{\mu}\tau, \sigma_\eta^2\left(\tau + \frac{\tau^2}{m}\right)\right), \tag{32}$$

see e.g. Sampson (1991) for the same specific case of the geometric random walk with drift[10].

[10] Note that this derivation conditions on past data, but treats the estimated parameter as stochastic, which may appear contradictory. Here, one considers the ensemble for the entire time series $(0 \ldots m \ldots (m+\tau))$, rather than just for the future conditional on one specific realisation of the past (a single realisation $(0 \ldots m)$ and an uncertain future $((m+1) \ldots (m+\tau))$). In terms of simulations, it means drawing one full time series, splitting into train and test, evaluating one forecast, and repeating by drawing an entirely new full time series at the next iteration. This is different from drawing one time series until m, making a forecast, and comparing it to the distribution of possible futures conditional on that one specific realisation.

An alternative derivation of formula (32) arises if we proceed as in Way et al. (2022) have done for Wright's law. We will call this "the simulation method". Way et al. (2022) are interested in making predictions, but do not always have access to the raw data, so they sometimes have to pick parameter values from the literature. Let us start from the last observed value, and project forward by first drawing a drift parameter from its distribution, taken to be

$$\tilde{\mu} \sim \mathcal{N}(\hat{\mu}, \sigma^2/m), \tag{33}$$

where σ^2/m can directly be a standard error reported in the literature, and then draw a set of future noises from

$$\eta_{m+\tau} \sim \mathcal{N}(0, \sigma^2).$$

Now let us denote a future value generated in this way by $\ln \tilde{c}_{m+\tau}$. Instead of simulating, we can actually compute its probability exactly as

$$\Pr(\ln \tilde{c}_{m+\tau} = x) = \int \Pr(\ln \tilde{c}_{m+\tau}|\tilde{\mu}) \Pr(\tilde{\mu}) d\tilde{\mu}, \tag{34}$$

where the first factor in the integral is $\mathcal{N}(\ln c_m + \tilde{\mu}\tau, \sigma_\eta^2 \tau)$ by well-known properties of random walks, and the second factor is from Eq. 33. Inserting the formulas for the Gaussian probability densities with the appropriate means and variances, and performing the integral gives (32) as expected. Note that this is essentially a Bayesian calculation; in fact, it can be shown that the same formula arises in a Bayesian framework with an improper prior on μ.

While this is intuitive, we shall now see that a complication arises when one introduces autocorrelated noise. The reason to show this is that in practice, to make a forecast, we may have to pick the parameters (drift and its standard error, and variance of the noise) from the literature. In these cases we do not have the luxury to estimate all the parameters ourselves, including in a Bayesian way. Reassuringly, we will find that the predictions from the "simulation method" reproduce very well the exact distribution of forecast errors.

A.2 Moore's law with moving average noise

When noise is autocorrelated, the realisations of the noise in the latest periods of the estimation sample affect the parameter estimation (of course), *but also future noise*, since future noise depends on past noise.

In Farmer and Lafond (2016), we chose an MA(1) process for the noise (Eq. 25). The reason was that it worked well for our purpose - it increased the theoretical variance of the forecast errors in a way that made it more similar to the empirical forecast errors; we used it a bit as a fudge factor, providing us with a powerful parametric model for the forecast errors. It is quite possible that more complex ARMAs would be more appropriate[11], but the MA(1) makes analytical calculations very simple because noise at time t depends only on noise at time $t-1$.

Starting again from (31) and using (29), the forecast error is

$$\mathcal{E} = (\mu - \hat{\mu})\tau + \sum_{i=m+1}^{m+\tau} \eta_i = \left(\frac{-1}{m}\sum_{i=1}^{m} \eta_i\right)\tau + \sum_{i=m+1}^{m+\tau} \eta_i. \tag{35}$$

We will now see three methods to compute or approximate the variance of (35).

[11] DGPs with fat tail or long memory noise did not work as well as DGPs with ARMA noise to imply forecast errors similar to those we had empirically, but these two possibilities could be re-investigated with more and better data.

Method 1: Exact calculation from independent noise terms. The first method is to painstakingly separate all the independent noise terms ϵ_i in Eq. 35. Substituting in (25),

$$\begin{aligned}\mathcal{E} &= \underbrace{\left\{\frac{-\tau}{m}\sum_{i=1}^{m}[\epsilon_i+\rho\epsilon_{i-1}]\right\}}_{\text{due to parameter estimation}} + \underbrace{\left\{\sum_{i=m+1}^{m+\tau}[\epsilon_i+\rho\epsilon_{i-1}]\right\}}_{\text{due to future noise}} \\ &= \left\{\frac{-\tau}{m}\left(\rho\epsilon_0+\frac{\tau}{m}\sum_{i=1}^{m-1}(1+\rho)\epsilon_i+\epsilon_m\right)\right\}+\left\{\rho\epsilon_m+\sum_{i=m+1}^{m+\tau-1}(1+\rho)\epsilon_i+\epsilon_{m+\tau}\right\} \\ &= \frac{-\tau}{m}\left(\rho\epsilon_0+\frac{\tau}{m}\sum_{i=1}^{m-1}(1+\rho)\epsilon_i\right)+\underbrace{\left(\rho-\frac{\tau}{m}\right)\epsilon_m}_{\text{“mixed” term}}+\sum_{i=m+1}^{m+\tau-1}(1+\rho)\epsilon_i+\epsilon_{m+\tau}.\end{aligned}$$

The term ϵ_m has gathered a prefactor from parameter estimation (τ/m), and another from future noise (ρ); we will see below that this creates non-null covariance between $\hat{\mu}$ and future noise. For now, now that we have our fully independent noise terms we can get the variance as

$$\mathrm{Var}(\mathcal{E})_{\text{exact}} = \sigma_\epsilon^2\left[\left(\frac{\tau}{m}\rho\right)^2+(m-1)\left(\frac{\tau}{m}(1+\rho)\right)^2+\left(\rho-\frac{\tau}{m}\right)^2+(\tau-1)(1+\rho)^2+1\right],$$

and while the algebra is a little tedious, this simplifies to

$$\mathrm{Var}(\mathcal{E})_{\text{exact}} = \sigma_\epsilon^2\left[-2\rho+\left(1+2\frac{m-1}{m}\rho+\rho^2\right)\left(\tau+\frac{\tau^2}{m}\right)\right]. \tag{36}$$

In this formula, we cannot neatly separate a “τ term associated with future noise” from a “τ^2/m term associated with parameter estimation” since the τ term is multiplied by a factor that includes the sample size for parameter estimation m. In more general settings, this is one of the reasons why forecast error taxonomies are only approximate (Clements and Hendry, 1998, Section 7.2 p.163).

However, assuming that m is large enough (that is, $(m-1)/m \approx 1$) and τ is large enough (-2ρ can be neglected), we get the approximation quoted in the main text,

$$\mathrm{Var}(\mathcal{E})_{\text{exact}} \approx \sigma_\epsilon^2(1+\rho)^2\left(\tau+\frac{\tau^2}{m}\right) = \sigma_\eta^2\frac{(1+\rho)^2}{1+\rho^2}\left(\tau+\frac{\tau^2}{m}\right) \equiv \mathrm{Var}(\mathcal{E})_{\text{approx}}. \tag{37}$$

Since the only issue preventing us from separating the two effects is a single noise term that affects parameter estimation and future noise (ϵ_m), its overall influence decreases if we increase the sample size and the number of future periods. It is good to keep in mind that the problem will be worse the stronger and the longer the autocorrelation is (e.g. more general ARIMAs with “high” parameters).

Method 2: Evaluating the variance due to parameter estimation and future noise separately. The following preliminary is useful. Recall from Eq. 27 that the MA(1) process from Eq. 25 has variance $(1+\rho^2)\sigma_\epsilon^2$. However, if we sum the noise, say m times, the variance is not $m\sigma_\eta^2$, but instead

$$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{m}\eta_i\right) &= \mathrm{Var}\left(\sum_{i=1}^{m}(\epsilon_i+\rho\epsilon_{i-1})\right) = \mathrm{Var}\left(\rho\epsilon_0+\sum_{i=1}^{m-1}\epsilon_i+\epsilon_m\right) \\ &= \sigma_\epsilon^2\left(\rho^2+(m-1)(1+\rho)^2+1\right) = \sigma_\epsilon^2\left(m(1+\rho)^2-2\rho\right).\end{aligned} \tag{38}$$

Note the useful approximation

$$\mathrm{Var}\Big(\sum_{i=1}^{m}\eta_i\Big)\approx\sigma_\epsilon^2 m(1+\rho)^2. \tag{39}$$

Now, we can compute the variance of the forecast error starting from (35), which is exact, as

$$\begin{aligned}\mathrm{Var}(\mathcal{E}) &= \mathrm{Var}\Big(\tau(\mu-\hat{\mu})\Big)+\mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau}\eta_i\Big)+2\,\mathrm{Cov}\Big(\tau(\mu-\hat{\mu}),\sum_{i=m+1}^{m+\tau}\eta_i\Big)\\ &= \tau^2\mathrm{Var}(\hat{\mu})+\mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau}\eta_i\Big)-2\tau\,\mathrm{Cov}\Big(\hat{\mu},\sum_{i=m+1}^{m+\tau}\eta_i\Big)\\ &= \tau^2\mathrm{Var}\Big(\frac{1}{m}\sum_{i=1}^{m}\eta_i\Big)+\mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau}\eta_i\Big)-2\tau\,\mathrm{Cov}\Big(\frac{1}{m}\sum_{i=1}^{m}\eta_i,\sum_{i=m+1}^{m+\tau}\eta_i\Big).\end{aligned} \tag{40}$$

Using Eq. 38, we compute easily the two variances as

$$\mathrm{Var}(\hat{\mu})=\mathrm{Var}\Big(\frac{1}{m}\sum_{i=1}^{m}\eta_i\Big)=\sigma_\epsilon^2\Big(\frac{1}{m}\Big)^2\Big(\rho^2+(m-1)(1+\rho)^2+1\Big) \tag{41}$$

and

$$\mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau}\eta_i\Big)=\sigma_\epsilon^2\Big(\rho^2+(\tau-1)(1+\rho)^2+1\Big). \tag{42}$$

The covariance term is, indeed, not zero,

$$\mathrm{Cov}\Big(\frac{1}{m}\sum_{i=1}^{m}\eta_i,\sum_{i=m+1}^{m+\tau}\eta_i\Big)=E\Big[\frac{\epsilon_m}{m}\rho\epsilon_m\Big]=\frac{\rho}{m}\sigma_\epsilon^2, \tag{43}$$

that is, if $\rho>0$, there is positive correlation between $\hat{\mu}$ and future noise. Substituting Eqs. 41, 42 and 43 back into 40 gives the exact formula derived using method 1, Eq 36.

If instead of the exact formulas for the variance (using (38)), we use the approximation (39), and assume that the covariance is equal to zero, we retrieve the approximation for the forecast errors derived earlier, Eq. 37.

Method 3: Simulation. The third method consists in drawing a value from the distribution of $\hat{\mu}$, and then drawing values from an MA(1) process for the future noise, and repeating this experiment many times to obtain a distributional forecast. The simulation method effectively generates

$$\ln c_{m+\tau}=\ln c_m+\tilde{\mu}\tau+\sum_{i=m+1}^{m+\tau}\eta_i,$$

where

$$\tilde{\mu}\sim\mathcal{N}\Big(\hat{\mu},\mathrm{Var}(\hat{\mu})\Big).$$

with $\mathrm{Var}(\hat{\mu})$ from Eq. 41. The variance of the predicted log cost in the simulation method is thus

$$\mathrm{Var}(\mathcal{E})_{\mathrm{sim}}\equiv\mathrm{Var}(\ln c_{m+\tau})=\tau^2\mathrm{Var}(\hat{\mu})+\mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau}\eta_i\Big),$$

using Eqs. 41 and 42 for an explicit formula.

Put differently, the simulation method is better than the approximation because it produces exact values for the variances (in the limit of infinitely many simulations), but it is not as good as the exact method because it omits the covariance between the distribution of the estimated parameter and the distribution of future noise.

To sum up, we have

$$\mathrm{Var}(\mathcal{E})_{\mathrm{sim}} = \mathrm{Var}(\mathcal{E})_{\mathrm{exact}} - 2\tau \mathrm{Cov}\left(\hat{\mu}, \sum \text{future noise}\right) = \mathrm{Var}(\mathcal{E})_{\mathrm{exact}} - 2\sigma_\epsilon^2 \frac{\tau}{m}\rho. \tag{44}$$

A note on bias. An unbiased forecast of a MA(1) process would use the estimated past noise to make a forecast of future noise

$$E[\eta_{m+1}|\hat{\epsilon}_m] = \rho\hat{\epsilon}_m.$$

For any particular realisation of a time series, our forecast for time $m+1$ is biased because it does not account for this. However, on average over the ensemble, where we cannot condition on a specific realisation of ϵ_m, the forecasts are unbiased.

A.3 Wright's law without autocorrelation

Starting again from the general model (24) but with $\rho = 0$, and using shorthands Y and X for growth rates of costs and experience, respectively, we have

$$Y_t = \omega X_t + \eta_t. \tag{45}$$

We estimate the learning exponent as

$$\hat{\omega} = \frac{\sum_{i=1}^m Y_i X_i}{\sum_{i=1}^m X_i^2} = \omega + \frac{\sum_{i=1}^m X_i \eta_i}{\sum_{i=1}^m X_i^2}, \tag{46}$$

where the last step comes from substituting in (45). This is an OLS regression through the origin (omitting the intercept is important because the intercept would estimate an exponential time trend; see the main text for an informal discussion and Section C for additional discussion in the context of spurious regressions). Note that here and throughout, I treat experience as given (i.e., as "non-stochastic").

Note for later that, since the noise terms η_t are independent,

$$\hat{\omega} \sim \mathcal{N}\left(\omega, \frac{\sigma_\eta^2}{\sum_{i=1}^m X_i^2}\right). \tag{47}$$

The predictions are made as

$$\ln \hat{c}_{m+\tau} = \ln c_m + \hat{\omega} \ln \frac{z_{m+\tau}}{z_m} = \ln c_m + \hat{\omega} \sum_{i=m}^{m+\tau} X_i. \tag{48}$$

The forecast error then is

$$\mathcal{E} \equiv \ln c_{m+\tau} - \ln \hat{c}_{m+\tau} = (\omega - \hat{\omega}) \sum_{i=m}^{m+\tau} X_i + \sum_{i=m+1}^{m+\tau} \eta_i.$$

Using (47), we can see that the forecast is unbiased ($E[\mathcal{E}] = 0$), and the variance can be computed as

$$\begin{aligned}
\mathrm{Var}(\mathcal{E}) &= \Big(\sum_{i=m+1}^{m+\tau} X_i \Big)^2 \frac{\sigma_\eta^2}{\sum_{i=1}^{m} X_i^2} + \tau \sigma_\eta^2, \\
&= \sigma_\eta^2 \left(\tau + \frac{\left(\sum_{i=m+1}^{m+\tau} X_i \right)^2}{\sum_{i=1}^{m} X_i^2} \right), \\
&= \sigma_\eta^2 \left(\tau + \frac{\tau^2}{m} \mathcal{W} \right),
\end{aligned}$$

where

$$\mathcal{W} \equiv \frac{r_{(f)}^2}{r_{(p)}^2 + \sigma_{z,(p)}^2} \tag{49}$$

is defined from the "sample moments" of experience: $r_{(f)}$ is the mean growth in the future $(1 \dots \tau)$, and $r_{(p)}$ and $\sigma_{z,(p)}$ are the mean and standard deviation in the past $(1 \dots m)$, noting that $\hat{\sigma}_{z,(p)}$ must be calculated without the "Bessel" degrees of freedom correction. I put "sample moments" in quotation mark because experience is taken as non-stochastic - all further derivation are conditioned on one specific realisation of (past and future) experience.

Eq. 49 now looks very similar to the equivalent equation for Moore's law, Eq. 31, and the main text provides intuition into the differences.

A.4 Wright's law with autocorrelation

Finally, we arrive at the more complex model,

$$Y_t = \omega X_t + \eta_t, \qquad \eta_t = \epsilon_t + \rho \epsilon_{t-1}, \tag{50}$$

The parameter $\hat{\omega}$ is estimated as before, Eq. 46. The forecast errors are

$$\mathcal{E} \equiv \ln c_{m+\tau} - \ln \hat{c}_{m+\tau} = (\omega - \hat{\omega}) \sum_{i=m+1}^{m+\tau} X_i + \sum_{i=m+1}^{m+\tau} \eta_i. \tag{51}$$

Substituting $\hat{\omega}$ from Eq. 46,

$$\begin{aligned}
\mathcal{E} &= \left(- \frac{\sum_{i=1}^{m} X_i \eta_i}{\sum_{i=1}^{m} X_i^2} \right) \sum_{i=m+1}^{m+\tau} X_i + \sum_{i=m+1}^{m+\tau} \eta_i, \\
\mathcal{E} &= \sum_{i=1}^{m} \left[X_i \left(\frac{-\sum_{i=m+1}^{m+\tau} X_i}{\sum_{i=1}^{m} X_i^2} \right) \eta_i \right] + \sum_{i=m+1}^{m+\tau} \eta_i, \\
\mathcal{E} &= \sum_{i=1}^{m} H_i \eta_i + \sum_{i=m+1}^{m+\tau} \eta_i,
\end{aligned} \tag{52}$$

where I have defined

$$H_i \equiv X_i \left(\frac{-\sum_{i=m+1}^{m+\tau} X_i}{\sum_{i=1}^{m} X_i^2} \right) = \left(\frac{-\tau}{m} \cdot \frac{r_{(f)}}{r_{(p)}^2 + \sigma_{z,(p)}^2} \right) X_i \equiv \xi X_i. \tag{53}$$

Exact variance of the forecast errors. From (52) we can now carefully separate independent noise terms ("method 1" in Section A.2),

$$\begin{aligned}\mathcal{E} &= \sum_{i=1}^{m} H_i(\epsilon_i + \rho\epsilon_{i-1}) + \sum_{i=m+1}^{m+\tau} (\epsilon_i + \rho\epsilon_{i-1}),\\ &= \left\{\rho H_1\epsilon_0 + \sum_{i=1}^{m-1}(H_i + \rho H_{i+1})\epsilon_i + H_m\epsilon_m\right\} + \left\{\rho\epsilon_m + \sum_{i=m+1}^{m+\tau-1}(1+\rho)\epsilon_i + \epsilon_\tau\right\},\end{aligned}$$

so that the variance of the forecast errors is

$$\mathrm{Var}(\mathcal{E})_{\text{exact}} = \sigma_\epsilon^2\left(\rho^2 H_1^2 + \sum_{i=1}^{m-1}(H_j + \rho H_{j+1})^2 + (\rho + H_m)^2 + (\tau-1)(1+\rho)^2 + 1\right). \tag{54}$$

Justifying an approximation. Let us rewrite Eq. 54 as

$$\frac{\mathrm{Var}(\mathcal{E})_{\text{exact}}}{\sigma_\epsilon^2} = \underbrace{\rho^2 H_1^2 + \sum_{i=1}^{m-1}(H_i + \rho H_{i+1})^2 + H_m^2 + 2\rho H_m}_{(1)} + \underbrace{\rho^2 + (\tau-1)(1+\rho)^2 + 1}_{\approx(1+\rho)^2\tau,\ \text{see (39)}}, \tag{55}$$

If experience grew at fairly similar rates year after year in the past, $X_j \approx X_{j'}$ for any j and j', so $H_j \approx H_{j'}$. In particular, if we just assume that $X_m \approx X_1$, we have

$$(1) \approx \sum_{i=1}^{m}(H_i + \rho H_{i+1})^2 = \xi^2 \sum_{i=1}^{m}(X_i + \rho X_{i+1})^2.$$

using (53) in the last step. This approximation becomes very good as m grows large, as only one out of m terms is approximated. Expanding, we have

$$(1) \approx \xi^2\left(\sum_{i=1}^{m} X_i^2 + \rho^2\sum_{i=1}^{m} X_{i+1}^2 + 2\rho\sum_{i=1}^{m} X_i X_{i+1}\right). \tag{56}$$

To get to a simple formula, the useful final approximation is

$$\hat{r}_{(p)}^2 + \hat{\sigma}_{(p)}^2 = \frac{1}{m}\sum_{i=1}^{m} X_i^2 \approx \frac{1}{m}\sum_{i=1}^{m} X_{i+1}^2 \approx \frac{1}{m}\sum_{i=1}^{m} X_i X_{i+1}. \tag{57}$$

Pre-multiplying (56) by m/m and using (57), we have

$$(1) \approx m(1+\rho)^2\xi^2(\hat{r}_{(p)}^2 + \hat{\sigma}_{(p)}^2) = (1+\rho)^2\frac{\tau^2}{m}\mathcal{W}, \tag{58}$$

where I have used (53) and (49). Substituting (58) into (55) and using $\sigma_\epsilon^2 = \sigma_\eta^2/(1+\rho^2)$ (27),

$$\mathrm{Var}(\mathcal{E})_{\text{approx}} = \sigma_\eta^2\frac{(1+\rho)^2}{1+\rho^2}\left(\tau + \frac{\tau^2}{m}\mathcal{W}\right), \tag{59}$$

which is the formula quoted in the main text.

Error of the simulation method. Again, the simulation method would draw ω from its distribution, and draw future noise terms independently. We would thus have

$$\mathrm{Var}(\mathcal{E})_{\mathrm{sim}} \equiv \mathrm{Var}(\ln c_{t+\tau}) = \Big(\sum_{i=m+1}^{m+\tau} X_i\Big)^2 \mathrm{Var}(\hat{\omega}) + \mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau} \eta_i\Big), \tag{60}$$

where the second term, the variance of future noise, is as in Moore's law, Eq. 42. For the first term, from looking at (46),

$$\hat{\omega} = \omega + \frac{1}{\sum_{i=1}^{m} X_i^2}\Big(\rho X_1 \epsilon_0 + \sum_{i=1}^{m-1}(X_i + \rho X_{i+1})\epsilon_i + X_m \epsilon_m\Big),$$

so its variance is

$$\mathrm{Var}(\hat{\omega}) = \frac{\sigma_\epsilon^2}{[\sum_{i=1}^{m} X_i^2]^2}\Big((\rho X_1)^2 + \sum_{i=1}^{m-1}(X_i + \rho X_{i+1})^2 + X_m^2\Big).$$

As for Moore's law, the simulation method (59) is only approximate because, starting from (51), we should express the variance of forecast errors as the sum of a term linked to parameter estimation, a term linked to future noise, *and* a covariance term ("method 2" in Section A.2),

$$\mathrm{Var}(\mathcal{E})_{\mathrm{exact}} \equiv \Big(\sum_{i=m+1}^{m+\tau} X_i\Big)^2 \mathrm{Var}(\hat{\omega}) + \mathrm{Var}\Big(\sum_{i=m+1}^{m+\tau} \eta_i\Big) - 2\Big(\sum_{i=m+1}^{m+\tau} X_i\Big)\,\mathrm{Cov}\Big(\hat{\omega}, \sum_{i=m+1}^{m+\tau} \eta_i\Big), \tag{61}$$

which can be compared to Eq. 40. The covariance term can be calculated as

$$\mathrm{Cov}\Big(\hat{\omega}, \sum_{i=m+1}^{m+\tau} \eta_i\Big) = E\Big[\Big(\epsilon_m \frac{X_m}{\sum_{i=1}^{m} X_i^2}\Big)\Big(\rho \epsilon_m\Big)\Big] = \sigma_\epsilon^2 \rho \frac{X_m}{\sum_{i=1}^{m} X_i^2}. \tag{62}$$

Comparing the definition for the exact and simulated variance, Eqs. 60 and 61, their difference is

$$\begin{aligned}
\mathrm{Var}(\mathcal{E})_{\mathrm{sim}} &= \mathrm{Var}(\mathcal{E})_{\mathrm{exact}} - 2\Big(\sum_{i=m+1}^{m+\tau} X_i\Big)\mathrm{Cov}\Big(\hat{\omega}, \sum_{i=m+1}^{m+\tau} \eta_i\Big), \\
&= \mathrm{Var}(\mathcal{E})_{\mathrm{exact}} - 2\sigma_\epsilon^2 \rho \frac{\sum_{i=m+1}^{m+\tau} X_i}{\sum_{i=1}^{m} X_i^2} X_m, \\
&= \mathrm{Var}(\mathcal{E})_{\mathrm{exact}} - 2\sigma_\epsilon^2 \frac{\tau}{m} \rho \mathcal{W} \frac{X_m}{r_f},
\end{aligned}$$

where the second step uses Eq. 62, and the last step makes it easier to compare to the equivalent for Moore's law, Eq. 44, remembering that $\mathcal{W}$ is defined in Eq. 49 and equals 1 in this case.

Fig. 1 compares the various formulas. To highlight that Wright's law is different from Moore's law, I assume that experience is a (deterministic) logistic curve, so the growth rates are not constant (Eq. 22, $L = 100$, $k = 0.5$, $t_0 = m + 5$). I take $\sigma_\epsilon^2 = 0.1$, and, to try to highlight the differences between the formulas, a very high autocorrelation ($\rho = 0.9$) and a very small sample size ($m = 5$). The "true" forecast errors are from generating 10^6 true series and forecasts, and the "simulation" is using 10^6 simulated paths.

The figure shows that the simulation method produces almost exact results. In general, it is difficult to find reasonable examples where the simulation method will deviate significantly. By contrast, the intervals based on $\mathrm{Var}(\mathcal{E})_{\mathrm{approx}}$ are perceptibly different, although still very close. A case where they differ dramatically from the exact result is when $\rho < 0$, as one can check that $\mathrm{Var}(\mathcal{E})_{\mathrm{approx}} \to 0$ as $\rho \to -1$.

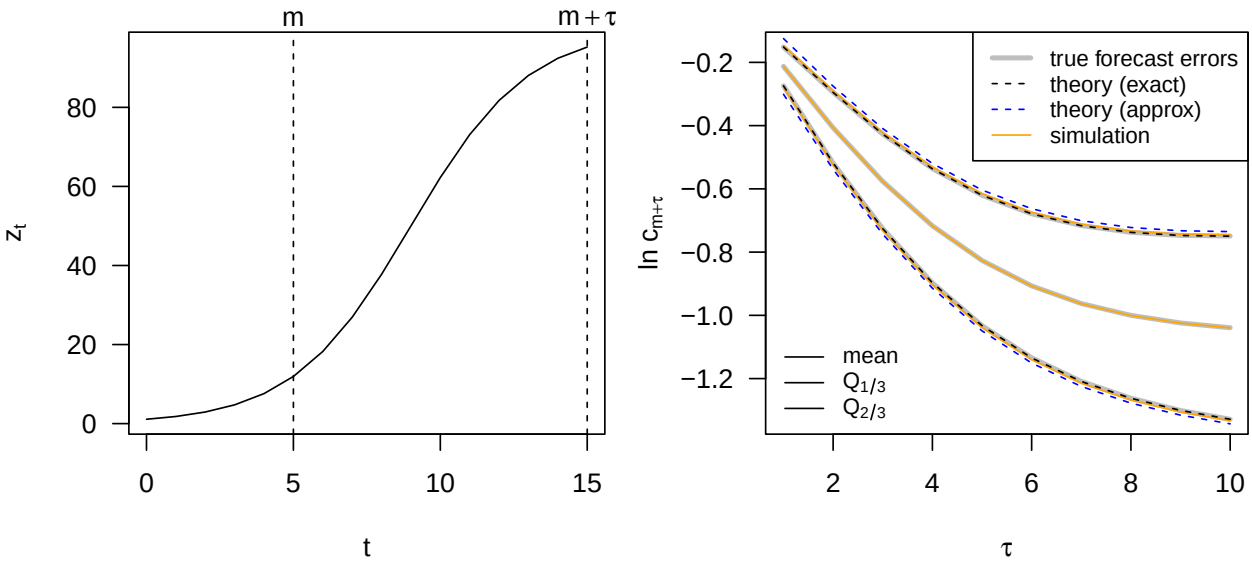

Figure 1: Distributional forecasts for costs conditional on future experience, comparing 3 methods. The left panel shows the evolution of experience, assumed to be a logistic curve, and highlights the assumption for the sample size (m) and forecast horizon (τ). The right panel shows the distributional forecasts, using Eq. 54, Eq. 59, and the "simulation" method (simulating forward paths after drawing from the distribution of estimated parameters). This is compared to the actual distribution of forecast errors obtained by simulating the stochastic process, making forecasts, and recording the errors. The quantiles shown are 1/3 and 2/3; see text for detailed values of the parameters.

B On the volatility of sums of lognormals

If production is a geometric random walk, what are the time series properties of *cumulative* production (experience)?

Let production be a random walk with drift,

$$\Delta \ln q_t = g + \sigma_q u_t,$$

where $u_t \sim \mathcal{N}(0,1)$. Thus, q_t is lognormal, and for future reference note that the variance of its growth rate (its volatility) is σ_q^2 while its ("ensemble" or "cross-sectional") variance grows linearly with time,

$$\mathrm{Var}(\ln q_t) \sim \sigma_q^2 t.$$

To avoid any confusion, let me define the empirical counterparts of the population moments (which will be used in simulations below). The "variance" is the cross-sectional variance of the levels of the log variable

$$\mathrm{Var}(\ln q_t) \equiv \frac{1}{N}\sum_{i=1}^{N}\Big(\ln q_{it} - \langle \ln q_t \rangle\Big)^2,$$

with $\langle \ln q_t \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} \ln q_{it}$. It is computed using N different time series, taken at their time step t.

In contrast, the "volatility" is the variance of the growth rate of one time series. It is computed as

$$\mathrm{Var}(\Delta \ln q_{it}) \equiv \frac{1}{t}\sum_{s=1}^{t}\Big(\Delta \ln q_{is} - \hat{g}_i\Big)^2,$$

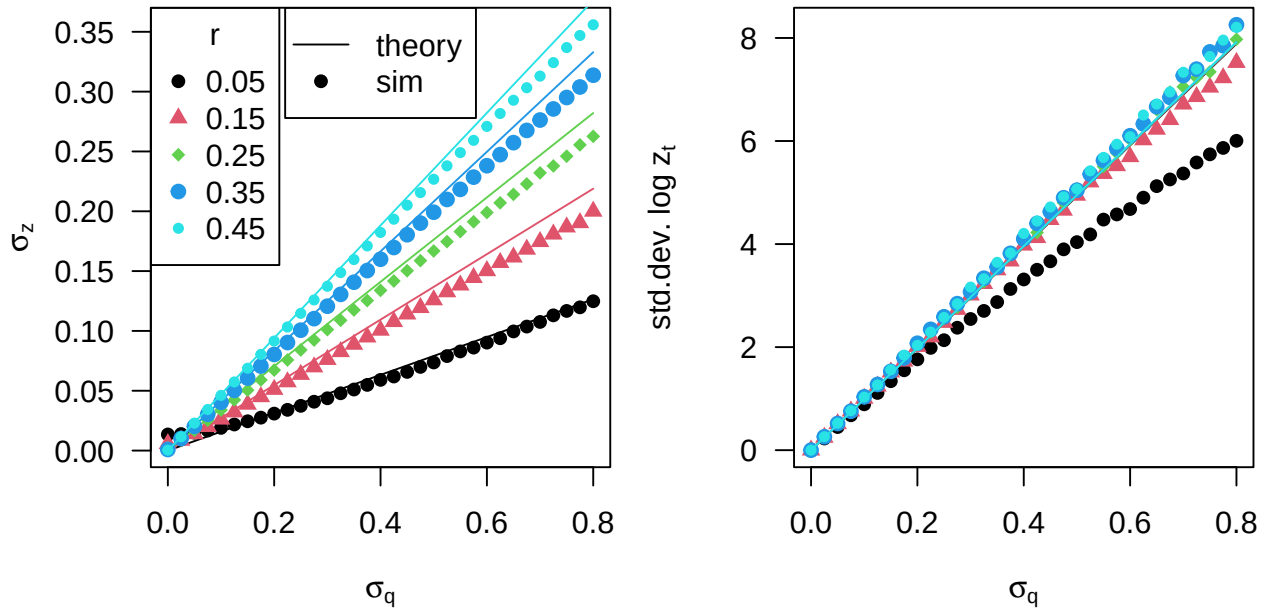

Figure 2: Empirical moments of experience against their predicted value.

where the average growth rate is the estimated drift, $\hat{g}_i = \frac{1}{t}\sum_{s=1}^{t} \Delta \ln q_{is}$. In simulations I will report the average of this quantity over the ensemble.

Now, define experience as cumulative production,

$$z_t = \sum_{j=1}^{t} q_j.$$

Lafond et al. (2018) derived the approximation Eq. 15 in the main text,

$$\sigma_z^2 \approx \sigma_q^2 \tanh(g/2).$$

To test how well it works, I simulate 5,000 time series of production of 110 periods (for each different values of g and σ_q); compute the cumulative sum of each time series to construct "experience"; remove the initial 10 periods, as a "burn-in"; and finally, compute the volatility of each of the 5,000 time series, and average over all simulations.

Fig. 2 (left) shows the results, demonstrating a good agreement between the simulations and the theoretical results for most reasonable parameter values. Zadourian and Klümper (2018) go a lot further, and derive the distribution of experience. Let me now comment on two further results.

Experience is not volatile, but it is uncertain. First, Lafond et al. (2018) also derived an approximation for the variance of the levels as

$$\mathrm{Var}(\ln z_t) \approx \sigma_q^2 \left(\frac{2e^g + 1}{1 - e^{2g}} + t \right).$$

This is interesting because since the first term disappears as $t \to \infty$, this implies

$$\mathrm{Var}(\ln z_t) \sim \mathrm{Var}(\ln q_t),$$

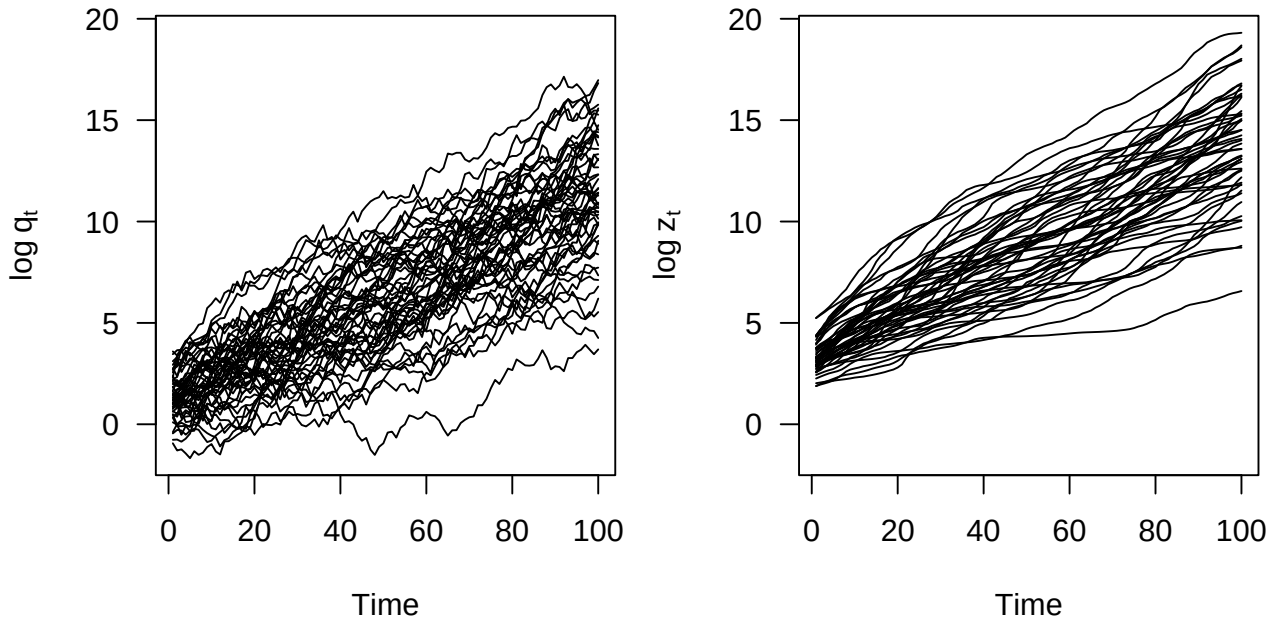

Figure 3: The cumulative (z_t, right) of a geometric random walks with drift (q_t, left) has much lower volatility, but similar cross-sectional dispersion.

where $\sim$ here means that the variables scale together as time goes to infinity (their ratio tends to 1). In other words, the *ensembles* of production and experience spread out in the same way (their cross-sectional variance is almost the same), even though, as we have seen, their volatility is much different. Fig. 2 (right) shows that this formula provides a good approximation.

To drive the point home, Fig. 3 shows 50 simulated time series of production, and the associated time series of experience ($r = 0.1$ and $\sigma_q = 0.3$). The Fig. makes it clear that while the volatility of experience is much lower, its cross-sectional dispersion remains similar. In this sense, we may say that while experience is much less volatile than production, it is just as uncertain.

Depreciation weakens the smoothing effect of cumulating. This phenomenology is interesting well beyond our specific application, given the prevalence of random walks and cumulative sums in various domains of science and engineering. Taking a view from economics, consider that we want to understand the fluctuations of the capital stock, defined as cumulative investment; this is a similar problem, with the important difference that one needs to take into account depreciation. In her PhD thesis, Zadourian (2018) further derived that if depreciation is denoted δ, so that

$$z_t = (1-\delta)z_{t-1} + q_t,$$

then Eq. 15 becomes

$$\sigma_z^2 \approx \sigma_q^2 \tanh\left(\frac{g - \ln(1-\delta)}{2}\right) \approx \sigma_q^2 \tanh\left(\frac{g+\delta}{2}\right),$$

where the last approximation holds for $|\delta| \ll 1$.

This shows that more depreciation leads to more volatility of the growth of the capital stock, and thus, thinking in terms of growth accounting equations, more volatility of labor productivity. This has potentially some relevance to the current landscape, where modern economies are shifting from traditional capital (which may have a depreciation $\delta = 1\% - 20\%$), to "intangible" capital like

software, which depreciates very quickly ($\delta \approx 20\% - 50\%$, see Goldin et al. (2024, Table 6).). This equation predicts that, everything else equal, the shift towards intangible capital should lead to higher aggregate fluctuations, or, to phrase it in the context of this paper, higher prediction intervals for predictions of technological progress.

Similarly, due to over-investment in fossil fuels in the last decades, it is likely that some assets will need to be retired earlier than originally planned (Pfeiffer et al., 2018). These stranded assets effectively correspond to higher depreciation, and thus, everything else equals, higher uncertainty of future productivity growth.

C Spurious regressions and experience curves

If both costs and experience were independent random walks with drifts, how would various estimators of their relationship perform? Specifically, which estimator would pick up that this is a spurious regression ($\omega = 0$)? Our goal is to show that the only estimator that works, the first-difference estimator allowing a non-zero intercept, also has very large standard errors, making it very hard to properly test for spurious regression.

Data generating process. Let

$$\Delta \ln c_t \equiv Y_t = \mu + u_t, \tag{63}$$

$$\Delta \ln z_t \equiv X_t = r + v_t. \tag{64}$$

To denote the variances of u and v, which will come up in calculations, I will directly use σ_Y^2 and σ_X^2.

The level estimator. It is defined as the OLS estimator of ω in the equation

$$\ln c_t = \alpha + \omega \ln z_t + \epsilon_t. \tag{65}$$

Entorf (1997) shows that $E[\hat{\omega}] \to \mu/r$. This is the well-known "spurious regression" result: performing a regression with two independent random walks returns the ratio of their drifts, rather than 0, with a high degree of statistical significance.

Entorf (1997) also shows that asymptotically, $\hat{\omega}$ is normal with variance

$$\mathrm{Var}(\hat{\omega}) = \frac{1}{m}\left(\frac{6}{5} \cdot \left(\frac{\mu}{r}\right)^2 \left[\left(\frac{\sigma_Y}{\mu}\right)^2 + \left(\frac{\sigma_X}{r}\right)^2\right]\right). \tag{66}$$

So precision increases when $\sigma_X \to 0$, meaning that in our case of interest, the estimator becomes increasingly precisely wrong.

The first difference (FD) estimator with an intercept. It is defined as the OLS estimator of ω in the equation

$$Y_t = \alpha + \omega X_t + \epsilon_t. \tag{67}$$

In this case, the ϵ_t are stationary, so the OLS regression gives "correct" results, $E[\hat{\omega}] = 0$, but with variance

$$\mathrm{Var}(\hat{\omega}) = \frac{\sigma_Y^2}{m\sigma_X^2}, \tag{68}$$

where I have used the classic result for when regressors are non-stochastic, and substituted $\hat{\sigma}_X$ by σ_X. These standard errors will be huge if σ_X is small, which we should expect (see Appendix B):

experience fluctuates very little so it is hard to distinguish it from "time ticking". Indeed, this model is not our stochastic Wright's law, because the intercept in difference implies a time trend in levels,

$$\ln c_t = \text{cstt} + \alpha t + \omega \ln z_t,$$

that is, a "mix" of Moore and Wright's laws. When we think that the data generating process is our stochastic Wright's law, the OLS estimator of (67) is both an estimator of the incorrect DGP, and an estimator with very poor precision. The right estimator in this case is discussed next.

The first-difference estimator without a constant. It is the OLS estimator of ω in the equation

$$Y_t = \omega X_t + \epsilon_t. \tag{69}$$

This is the model discussed in the main text, Eq. 11. The estimator is defined in Eq. 46, and Appendix A.3 studies its distribution when model (69) is correct.

Here we are interested in the properties of the estimator (46) when the model (69) is *not* correct, but instead, the data is generated as two independent random walks Eqs. 63-64. We can easily find the behaviour of $\hat{\omega}$ heuristically, as follows. By definition of $\hat{\omega}$ (46), inserting the DGP Eqs. 63-64, we have

$$\hat{\omega} = \frac{\sum Y_i X_i}{\sum X_i^2} = \frac{\mu r + r\left(\frac{1}{m}\sum u_i\right) + \mu\left(\frac{1}{m}\sum v_i\right) + \left(\frac{1}{m}\sum v_i u_i\right)}{r^2 + 2r\left(\frac{1}{m}\sum v_i\right) + \left(\frac{1}{m}\sum v_i^2\right)}, \tag{70}$$

where m is the sample size. Let $m \to \infty$ and replace $\frac{1}{m}\sum u_i \to 0$ and $\frac{1}{m}\sum v_i \to 0$. The sums with the v_i^2 is just a consistent estimator of the volatility of X, $\frac{1}{m}\sum v_i^2 \to \sigma_X^2$. By assumption the noises of the two times series are independent, so $\frac{1}{m}\sum v_i u_i \to 0$. Plugging these four results into Eq. 70, we have

$$\hat{\omega}_{|m\to\infty} = \frac{\mu/r}{1 + (\sigma_X/r)^2}. \tag{71}$$

From this it becomes clear that whether the asymptotic $\hat{\omega}$ is "correct" or "spurious" depends on σ_X,

$$\hat{\omega}_{|m\to\infty} \to 0 \text{ as } \sigma_X \to \infty \qquad \text{[correct]},$$

$$\hat{\omega}_{|m\to\infty} \to \mu/r \text{ as } \sigma_X \to 0 \qquad \text{[spurious]}.$$

For the standard errors, a heuristic is to derive it conditional on X, $\text{Var}(\hat{\omega}|X) = \frac{\sigma_Y^2}{\sum_i X_i^2}$, and then, taking expectation over X and assuming that the expectation of the inverse is well approximated by the inverse of the expectation,

$$\text{Var}(\hat{\omega}) = \frac{\sigma_Y^2}{m(r^2 + \sigma_X^2)}. \tag{72}$$

In contrast to the first difference with a constant, Eq. 68, the denominator includes the term r^2 which prevents the standard errors from diverging as $\sigma_X \to 0$. So while the estimator that gives a "correct" result is highly imprecise, this one, somewhat like the spurious regression, is precisely wrong. When $\sigma_X \to \infty$, however, the expectation approaches the correct result, and the estimator becomes more precise.

Numerical results. To show numerically that Eq. 71 works well, I simulate independent random walks and estimate ω according to three methods: level, difference with and without intercept.

The simulation setup is similar to the values reported for Solar PV by Lafond et al. (2018): 41 periods, the mean and std. dev. of Y equal -0.121 and 0.153, and the mean of X is 0.318 (Lafond et al. (2018) report 0.318 for cumulative production, 0.315 for production). I use various values for the variance of X, to see its effects, and run 10,000 replications for each value of σ_X.

To test the formulas, the first set of simulation follows the assumption made for the calculations, that is, that X is a random walk with drift. Next we will see if the formulas still work in the realistic case where X is the growth of the *cumulative* of a geometric random walk with drift.

Fig. 4 (left) shows how each estimator performs at different values of σ_X. Dots show the mean across replications, and error bands show the 0.025-0.975 quantiles. To avoid clutter, the theoretical results for the standard errors are not shown, as they do not work perfectly for small sample size, but I have checked that they work well for larger sample sizes.

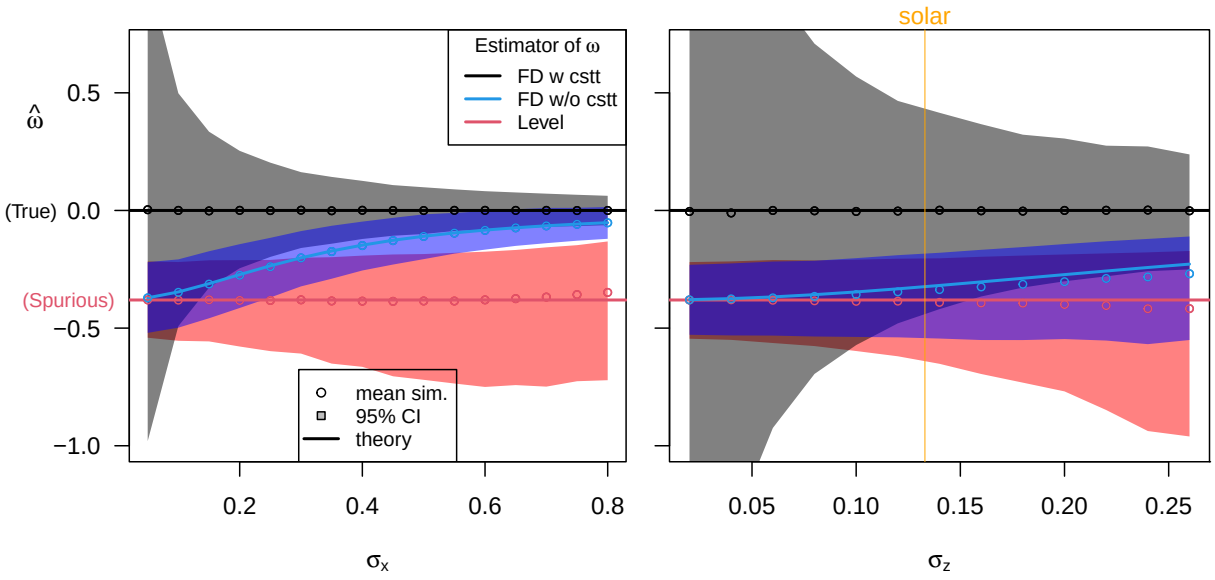

Figure 4: Properties of 3 different estimators of ω as a function of the variance of the regressor, when the DGP is two independent random walks (left). On the right panel, the regressor is "experience", that is, cumulative of geometric random walks.

As seen in the theoretical discussion, the first-difference estimator with an intercept, in black, is correct on average, but features large errors when the variance of the regressor is small, as intuitively known from regressions: a larger range of the regressor is better to pin down the correct slope.

The level model, in red, gives on average the spurious result. A large volatility of the regressor makes it worse, as has been shown by Entorf (1997).

Finally, for the estimator of interest, the first-difference without constant, the blue line shows that the theoretical result (71) agrees very well with the simulations, and the variance of the estimator decreases slightly with more variance of the regressor. For small volatility, there is no chance that it is able to pick up the correct value (zero), but it is substantially less biased than the level estimator.

So the key question is whether in practice, the variance of the regressor (the volatility of experience) is "high enough". To investigate, I first run the same simulations as before, but simulating X as the growth rate of the cumulative of a geometric random walk, rather than as a random walk. This means the simulations are now different from the DGP assumed in (64), but closer to the real-world.

The right panel of Fig. 4 shows again that the formulas still works well. Note that the x-axis now spans a much smaller range; in fact I chose it so that the volatility of experience on the right panel corresponds to the volatility of production on the left panel, using the result 15 (see Appendix B).

Doing this makes it possible to check where "real technologies" lie. I take the value of σ_z measured for solar by Lafond et al. (2018). Unfortunately, for a value of σ_z comparable to empirical data for solar, the difference estimator without intercept is not in the good regime: it still gives a spurious answer. In this regime, the first-difference estimator with a constant is hardly very useful, as its standard errors are so large that the spurious answer is within its confidence interval. Thus, in this regime, none of the two estimators would be very useful to detect whether the regression is spurious.