Growth and Redistribution: The Hedging Perspective*

Larry Samuelson
Yale University

Jakub Steiner
University of Zurich, CERGE-EI and CTS

August 12, 2024

Abstract

We investigate the impact of wealth redistribution on economic growth, building on Kelly's (1956) optimal investment portfolio theory. A growth-optimal policy redistributes wealth from 'lucky' overperforming individuals to underperforming individuals, minimizing the systematic component of this redistribution in a myopic fashion. That is, the policy minimizes the discrepancy between endowments and outcomes, counterfactually taking outcomes as independent of endowments. Alternatively, we reinterpret this result in terms of maximum likelihood estimation of a distribution over both latent and observable variables. Beliefs derived from the estimated joint distribution fail Bayes' plausibility due to misspecification; however, the estimate myopically minimizes this failure.

*This is a short version of our working paper, Samuelson and Steiner (2023). We have benefited from the comments of Mira Frick, Tomáš Gavenčiak, Ben Hébert, Ryota Iijima, Yuhta Ishii, Roman Kotecký, Jan Kulveit, Rava da Silveira, Ludvig Sinander, Colin Stewart, and various seminar and workshop audiences. This work was supported by ERC grant 770652 and GAČR grant 24-10145S.

Growth and Redistribution: The Hedging Perspective

1 Introduction

Does economic growth generate inequality, or perhaps does inequality inhibit growth? Can wealth redistribution foster economic growth? The literature identifies various channels through which inequality and redistribution affect economic growth, primarily explored in deterministic models that emphasize effects on productivity via incentives, the allocation of resources, and technological advancement.

We investigate a relatively underexplored aspect of the relationship between inequality, redistribution, and economic growth. Economic growth is inherently stochastic, with some fortunate individuals experiencing rapid wealth accumulation while others lag behind, even when controlling for productivity. Maximizing the economic growth rate requires redistribution to hedge against these disparities in luck.

To emphasize the hedging aspect of wealth redistribution, we set aside all the conventional modeling elements of economic growth. Incentive effects, production, technological progress, and other factors are all condensed into a black box, that determines each individual's end-of-period wealth based on their initial wealth and the realization of a random state. Different individuals have distinct such *return functions*, reflecting persistent differences in their skills, education, opportunities, and other factors. How should a planner with the ability to periodically redistribute wealth respond optimally to individuals' ex-ante differences and varying degrees of luck?

Formally, we build upon Kelly's (1956) classical study of optimal private investment portfolios. Kelly's investor allocates wealth to assets with uncertain returns to maximize the long-term growth rate of the portfolio's value. Kelly provides two main insights. First, a growth-optimal portfolio involves hedging—it comprises a variety of assets, including those with inferior expected returns, and is maintained by periodically redistributing wealth from unexpectedly over-performing to underperforming assets. Second, the optimal portfolio eliminates the systematic part of this redistribution—the expected wealth share of each asset must remain the same before and after rebalancing.

In our economic application, we view individuals as assets with uncertain returns. The optimal policy periodically redistributes wealth to maintain a growth-maximizing allocation, akin to rebalancing a financial portfolio. To fit our economic context, we extend the standard portfolio choice setting in three ways. First, we allow the planner to control the state-contingent returns of individuals, subject to feasibility constraints, thereby capturing the effects of economic policies on individuals' varying productivities. Second, to address inequality concerns, we impose constraints on the planner's choice of wealth allocation. Finally, we allow returns and wealth allocations to interact, recognizing that redistribution may affect incentives and, consequently, individual returns.

To a first approximation, our results echo Kelly's insights on the growth-optimal portfolio. The growth-optimal economic policy involves hedging, allocating wealth among a variety of individuals, including those with inferior expected returns. Additionally, the growth-optimal policy routinely redistributes wealth from accidental economic winners to losers, thereby maintaining an optimal hedge. However, constraints may render the full elimination of systematic redistribution infeasible in our setting, and, instead, the growth-optimal policy *myopically* minimizes this redistribution.

To formulate this result, we compare two wealth distributions. The *endowment* distribution refers to the wealth distribution periodically controlled by the planner via redistribution. The *outcome* distribution represents the resulting wealth distribution, after endowments have been augmented by returns, averaged across random states. Our main result states that the growth-optimal endowment distribution minimizes the Kullback-Leibler divergence from the induced outcome distribution, treating the latter as given.

Somewhat counterintuitively, this minimization is myopic; the optimality condition requires the planner to treat the outcome distribution as fixed and approximate it with an endowment distribution as closely as possible, disregarding that the outcomes themselves depend on the endowments. Acknowledging that the outcomes depend on the endowments could further reduce systematic redistribution, but at the expense of the economy's growth.

Our analytical approach is novel. We express the economy's growth rate as a function of economic policy and the resulting path of wealth circulation, which we define to capture how wealth moves among individuals through redistribution. Although wealth circulation depends on policy, our proof construction *decouples* them, allowing us to treat policy and the path of wealth circulation as independent inputs determining the growth rate. The source of the myopia in our main result is then clear: maximizing the growth rate requires each input to be individually optimal when the other is fixed at its respective optimum. We complete the result by linking the path of wealth circulation to the outcome distribution.

This decoupling has an economic interpretation: we divide aggregate wealth into infinitesimal fractions, each of which stochastically circulates throughout the economy, randomly moving among individuals through redistribution. These wealth fractions grow at their own rates, depending on their distinct realized paths of circulation. The fastest-growing fraction ultimately prevails and determines the economy's overall growth rate. Maximizing the economy's growth rate thus becomes a joint optimization over two decoupled controls: the endowment distribution and the path of wealth circulation.

Section 5 reinterprets the setting in terms of maximum likelihood estimation. The analogy between economic growth maximization and maximum likelihood estimation arises because both problems can be viewed as competitions among growth processes. In the latter case, the likelihood for each considered hypothesis, defined as the product of the likelihoods of all sample points, grows in a stochastic growth process with an expanding sample. Thus, the analyst asymptotically selects the hypothesis with the highest long-run growth rate, analogous to the planner selecting the policy with the highest economic growth rate. This analogy clarifies why distributed growth processes, such as the economic growth in our main application, may resemble inference.

Our main result translates into a generalization of the familiar Bayes' plausibility condition. First, an analyst estimates a joint distribution of observable and latent variables and then forms a posterior belief about the latent variable for each given value of the observable variable. For illustration, consider an analyst observing the choices (such as labor force participation) of many individuals and modeling these choices as stochastic consequences of the individuals' latent characteristics (such as reservation wages). The analyst estimates the joint distribution of choices and characteristics, then forms Bayesian posteriors about each individual's characteristics given the observed choices. If the analyst is well-specified, Bayes' plausibility implies that the average of the analyst's posteriors equals her marginal belief about the latent characteristics. However, if the analyst is misspecified and forms an incorrect estimate, Bayes' plausibility fails: the empirical average of her posteriors differs from her estimated marginal belief about the latent characteristics.

In the inference context, our main result states that the maximum likelihood estimate myopically minimizes the extent of this Bayes' plausibility failure. Specifically, the analyst's marginal belief about the latent characteristics minimizes the Kullback-Leibler divergence from the average of her posteriors, treating the latter as given. This tendency to align the prior with the average posterior arises from fitting the model to the data rather than from Bayes' rule. Thus, Bayes' plausibility may serve as a good approximation for agents who form their beliefs by fitting models to data, even when they are misspecified.

2 Literature

Inequality and Growth. The theoretical literature identifies several channels through which inequality might boost growth. In an early survey, Aghion et al. (1999, p. 1615) note that one's 'textbook' reaction is that inequality will foster growth by providing incentives. Inequality may also foster growth if the wealthy have higher saving rates (Kaldor, 1957), if inequality allows some individuals to become educated and entrepreneurs (Barro, 2000), if it increases the median voter's support for public education (Saint-Paul and Verdier, 1993), or if complementarities (Benabou, 1996a) or externalities (Galor and Tsiddon, 1997a) in human capital formation are strong enough. Galor and Tsiddon (1997b) explain that allowing high-ability workers to congregate increases both inequality and growth. Conversely, one can readily identify channels through which inequality retards growth (e.g., Todaro and Smith (2020, Chapter 5)). In Galor and Zeira (1993), inequality discourages investment in human capital. In Galor and Moav (2004), inequality enhances growth when the return to physical capital is relatively high but is detrimental to growth when the return to human capital is high. In De La Croix and Doepke (2003), inequality increases the fertility of the poor, thereby reducing human capital accumulation and growth. Mdingi and Ho (2021) provide a review of the theoretical links, noting arguments for both positive and negative effects of inequality on growth.

The survey of Aghion et al. (1999) concludes that inequality retards growth in some circumstances, and hence that attention should turn to the impact of redistribution. Once again, the theoretical implications are mixed. Redistribution is typically viewed as dampening incentives and thus reducing growth. In contrast, Stiglitz (1969) suggests that redistribution from the rich to the poor fosters growth by alleviating diminishing returns and credit constraints. In Chou and Talmain (1996), redistribution can either boost or impede growth, depending on the curvature of the labor supply curve.

We abstract from all these considerations to focus on optimal hedging. This is not meant to diminish the importance of existing models. Instead, we view our analysis as complementary, highlighting that the inherently stochastic nature of growth introduces considerations for inequality, redistribution, and growth that are often absent from the literature.

The empirical literature on inequality and growth, originating from work that added inequality to cross-country growth regressions pioneered by Barro and Sala-i-Martin (1995), is similarly mixed. An early view held that inequality has a weak negative effect on growth. Benabou (1996b, p. 13) concludes that "initial inequality is detrimental to long-run growth." Barro (2000, p. 5) summarizes his results as showing "little overall relation between income inequality and rates of growth and investment." Forbes (2000, p. 869) summarizes the literature as finding "a negative and just-significant coefficient on inequality, leading most economists to conclude that inequality has a negative impact on growth."

Recent research shows less consensus, both regarding the direction and strength of the results. According to Forbes (2000, p. 869), improved data show that "an increase in a country's level of income inequality has a significant positive relationship with subsequent economic growth." Berg et al. (2018, p. 260) describe their findings as showing that "lower net inequality is robustly correlated with faster and more durable growth... redistribution appears benign in terms of its impact on growth, except when it is extensive...." In contrast, El-Shagi and Shao (2019) find that inequality generally strengthens growth, particularly when education levels are high, but that redistribution can also promote growth. Breunig and Majeed (2020) find that inequality has a negative effect on growth, especially in countries with a high incidence of poverty. Cingano (2014) reports similar findings. Brueckner and Lederman (2018) find that inequality boosts growth in low-income countries, with the relationship reversing at higher incomes. Mdingi and Ho (2021) again offer a useful survey.

Variational Inference. Our analytical method links the classical information theory treatment of investment portfolios by Kelly (1956) to the modern variational inference literature. This literature, originating with Jordan et al. (1999) and Kingma and Welling (2013), provides approximations to Bayes' rule and maximum likelihood estimation. One can see this connection in Proposition 3. Within our focus on economic growth, Proposition 3 expresses growth maximization as an optimization over both a policy and a path of wealth circulation within the economy. The same optimization problem appears in the variational inference literature, albeit with a different interpretation. There, one of the controls is the analyst's statistical model of the data-generating process, while the other control is the analyst's system of posterior beliefs about latent variables. In this standard interpretation, instead of growth maximization, the focus is on maximizing statistical fit. Thanks to the homeomorphism between the growth optimization and the inference problem, our main result can be interpreted both as an optimality condition for wealth redistribution and as a relaxed Bayes' plausibility condition in the inference context.

There is a growing interest in variational inference methods in behavioral economics. Aridor et al. (2020, 2024) and Strzalecki (2024) modify the variational inference problem to accommodate behavioral biases. In Samuelson and Steiner (2024), we employ the variational inference approach to study interactions between misspecification and frictions in Bayesian updating.

This paper is one of our three projects that highlight the central role of large deviations—atypical sequences of draws—in stochastic growth processes. In Robson et al. (2023), we observed that the wealth concentration generated by stochastic growth without redistribution can be studied by exploiting an equivalence to a rational-inattention problem. This paper focuses on the impact of redistribution on growth and highlights the equivalence between growth with redistribution and the variational inference problem. In Samuelson and Steiner (2024), we focus on inference, using the connection to growth in the background to provide a micro-founded interpretation of the variational inference problem.

Misspecified Learning. The standard result on misspecified learning, originating from White (1982) and Berk (1966), identifies the asymptotic estimate as the best feasible approximation of the true generating distribution. Esponda and Pouzo (2016) incorporate misspecified learning into an equilibrium concept.[1] These learning models are concerned with observable variables. In contrast, we consider an analyst who, in addition to the observable variable, also reasons about its latent counterpart. Incorporating the latent variable introduces new considerations into the asymptotic characterization of the misspecified learning outcome, effectively turning it into a fixed point.

[1] For useful points of entry into the misspecified learning literature, see Frick et al. (2023) and Bohren and Hauser (2023) and their references.

3 Economic Growth

Although we describe the model in terms of economic growth, it allows for alternative interpretations. Refer to Section 5 for a reinterpretation within the context of statistical inference.

3.1 Growth Maximization

The population consists of a finite set I of individuals i, and there is a finite set Ω of states ω. The economy begins in period 1 with an initial quantity of perfectly divisible wealth. A time-invariant *endowment distribution* $p \in \Delta(I)$ assigns a share $p(i)$ of the current aggregate wealth to each individual $i \in I$ at the beginning of every period $t = 1, 2, \ldots$. Afterwards, in each period, nature independently draws a state ω_t according to an interior distribution $q^0 \in \Delta(\Omega)$. Subsequently, each individual i earns a nonnegative gross return per unit of her endowment. The return depends on the current state ω_t and the individual i; we denote this time-invariant *return function* as $r(i, \omega_t)$. The resulting aggregate wealth is then redistributed at the beginning of the next period according to p, and so on.

The policy pair p and r jointly determine the long-run growth rate of the aggregate wealth,

$$\mathrm{E}_{q^0(\omega)} \ln \Big(\sum_i p(i) r(i, \omega) \Big). \tag{1}$$

A planner chooses a policy pair from a set $\mathcal{P} \subseteq \Delta(I) \times \mathbb{R}_+^{I \times \Omega}$ of feasible policy pairs to maximize the growth rate (1). The set $\mathcal{P}$ is compact and contains at least one policy pair resulting in a finite growth rate. The existence of an optimizer is then ensured.[2] The *growth-maximizing* policy pair is denoted by $p^*(i)$ and $r^*(i, \omega)$.

Naturally, societies face constraints in both elements of the policy pair. Equity constraints may restrict feasible endowments $p(i)$, while technological constraints on capital formation may restrict the returns $r(i, \omega)$. The feasibility of endowments and returns is typically interconnected. The choice of the redistribution scheme may affect incentives and, consequently, the feasibility of return functions, resulting in a constraint set $\mathcal{P}$ that is not a product set. Unlike much of the economic research that derives the set $\mathcal{P}$ from microfoundations,

[2] If there exists a policy pair yielding a finite growth rate g, then the subset of $\mathcal{P}$ of policy pairs that produce growth rates of at least g is compact. The objective function in equation (1) is continuous on this set, ensuring the existence of a maximizer.

we consider $\mathcal{P}$ as a primitive and develop results applicable to all such sets.[3]

The restriction to stationary policy pairs is without loss of generality because the optimal stationary policy induces at least as high a growth rate as any history-dependent policy. Let a feasible *history-dependent* policy associate a policy pair from $\mathcal{P}$ with every history $(\omega_1, \ldots, \omega_{t-1})$, for each t.

Proposition 1 (Cover and Thomas (2006))**.** *Let the random variable S_t denote the aggregate wealth accumulated under a feasible history-dependent policy over the first t periods. Let S_t^* denote the aggregate wealth accumulated under the growth-maximizing stationary policy pair. Then, almost surely,*

$$\limsup_{t\to\infty} \left(\frac{\ln S_t}{t} - \frac{\ln S_t^*}{t} \right) \leq 0.$$

Cover and Thomas (2006, Theorem 16.3.1) prove this result for an exogenous return function. Theorem 1 from our working paper, Samuelson and Steiner (2023), provides a straightforward extension to the setting at hand.

3.2 Wealth Distributions

Our main result contrasts two wealth distributions. First, the endowment distribution $p(i)$, as introduced above and controlled by the planner, specifies the wealth shares at the beginning of each period. Second, we denote the wealth distribution at the end of each period as

$$o_{p,r}(i \mid \omega) = \frac{p(i)r(i,\omega)}{\sum_{j\in I} p(j)r(j,\omega)}, \tag{2}$$

which specifies the wealth shares of individuals i at the end of each period in which the state $\omega_t = \omega$. Furthermore,

$$o_{p,r}(i) = \mathrm{E}_{q^0(\omega)}\, o_{p,r}(i \mid \omega) \tag{3}$$

denotes the long-run end-of-period wealth distribution; this represents the probability that a dollar, randomly selected from the aggregate wealth at the end of a random period, belongs to individual i. We refer to $o_{p,r}(i \mid \omega)$ and $o_{p,r}(i)$ as the (state-contingent) *outcome distributions*.

[3] A straightforward reinterpretation incorporates consumption. Let individual i consume a share $c(i,\omega) \in [0,1]$ of her endowment in state ω and generate a gross return $r(i,\omega)$ on her residual wealth. Then our analysis applies with return function $r'(i,\omega) = (1 - c(i,\omega))r(i,\omega)$.

Our main result establishes a necessary condition for the growth-maximizing policy pair. The growth-optimal policy pair must minimize a measure of systematic redistribution, while taking the outcome distribution as given. To formulate this, we recall that the Kullback-Leibler divergence between two probability distributions, $o(i)$ and $p(i)$, is defined as follows:[4]

$$\mathrm{KL}\left(o(i) \parallel p(i)\right) := \sum_{i \in I} o(i) \ln \frac{o(i)}{p(i)},$$

and quantifies the discrepancy between the two distributions.

3.3 Main Result

The optimal policy pair must satisfy a fixed-point condition. To state this condition, let $\mathcal{E}^* = \{p(i) : (p(i), r^*(i, \omega)) \in \mathcal{P}\}$ be the set of endowment distributions to which the planner can deviate, starting from the optimal policy pair, without altering the optimized returns $r^*(i, \omega)$.

Proposition 2 (Myopically Minimal Redistribution)**.** *If the endowment distribution $p^*(i)$ and the return function $r^*(i, \omega)$ jointly maximize the aggregate growth rate, then $p^*(i)$ minimizes the* KL*-divergence from the induced outcome distribution:*

$$p^*(i) \in \underset{p(i) \in \mathcal{E}^*}{\arg\min} \, \mathrm{KL}\left(o_{p^*,r^*}(i) \parallel p(i)\right). \tag{4}$$

Thus, the planner myopically minimizes the systematic redistribution of wealth. Among the feasible endowment distributions in combination with the optimized returns, the planner chooses the one most closely aligned with the outcome distribution. This optimization is myopic because the planner proceeds as if the outcome distribution $o_{p,r^*}(i)$ is fixed at $o_{p^*,r^*}(i)$ and independent of the control $p(i)$. As a result, the planner generally does not choose the endowments $p(i)$ that minimize $\mathrm{KL}\left(o_{p,r^*}(i) \parallel p(i)\right)$. Such a minimization would further reduce this measure of systematic redistribution by leveraging the link between the endowment distribution and the induced outcome distribution. However, this would decrease the growth rate.

A simple case arises when the planner is unconstrained in her choice of an endowment distribution. Then, the growth-optimal policy pair fully eliminates systematic redistribution.[5]

[4] We use the standard convention $0 \ln 0 = 0$.

[5] Recalling that $o_{p^*,r^*}(i) = \mathrm{E}_{q^0(\omega)} \frac{p^*(i) r^*(i,\omega)}{\sum_j p^*(j) r^*(j,\omega)}$, this corollary implies that the expected

Corollary 1 (Unconstrained Endowments)**.** *Suppose that any endowment distribution $p(i) \in \Delta(I)$ is feasible along with each of the feasible return functions $r(i, \omega)$. Then, the optimal policy pair eliminates systematic redistribution, so that $p^*(i) = o_{p^*,r^*}(i)$.*

Cover and Thomas (2006, Section 16.2) derive the corollary statement from the first-order conditions in a special case with exogenous return functions. Therefore, Proposition 2 generalizes this classical result to cases where constraints prevent the full elimination of systematic redistribution and where the planner controls the return function.

Example 1 (Economic Growth)**.** Consider three individuals and three states, where $I = \Omega = \{1, 2, 3\}$. Each individual has a high return in her corresponding state and a low return otherwise: $r(i, \omega) = 2$ if $i = \omega$, and $r(i, \omega) = 1$ if $i \neq \omega$. The a priori probabilities of the individuals' favorable states are $q^0(1) = .05$, $q^0(2) = .35$, and $q^0(3) = .6$, indicating that the individuals' a priori stochastic productivity increases with i.

Let us start with an unconstrained planner who can choose any endowment distribution $p(i) \in \Delta(I)$. She maximizes the economy's growth rate by excluding the low-productivity individual 1 and endowing individuals 2 and 3 with wealth shares $p^*(2) \approx .105$ and $p^*(3) \approx .895$. The resulting growth rate of .418 exceeds the autarky growth rate ($.6 \ln(2) \approx .416$) of the most productive individual 3 as the moderate redistribution involving $i = 2$ serves as an advantageous hedge. As Corollary 1 implies, this unconstrained planner eliminates systematic redistribution. Indeed, numerical computation verifies that the induced outcome distribution $o_{p^*,r}(i)$ coincides with the chosen endowment distribution $p^*(i)$.

Leaving individual 1 impoverished may be deemed unacceptable. Accordingly, we impose a constraint on the endowment distribution by restricting its Theil inequality index to at most $\ln 3 - 1 \approx .099$.[6] Consequently, the constrained-optimal endowment distribution becomes more egalitarian, with $p^*(1) \approx .158$, $p^*(2) \approx .329$, $p^*(3) \approx .513$, and the achieved growth rate drops to .355.

This constrained planner systematically redistributes wealth in favor of individuals 1 and 2, resulting in the outcome distribution of $o_{p^*,r}(1) \approx .118$,

relative return $\mathrm{E}_{q^0(\omega)} \frac{r^*(i,\omega)}{\sum_j p^*(j) r^*(j,\omega)}$ is equalized across all individuals with positive endowments when $p(i)$ is unconstrained.

[6] The Theil index for the endowments $p(i)$ of three individuals is $\frac{1}{3} \sum_i \frac{p(i)}{\overline{p}} \ln \frac{p(i)}{\overline{p}} = -\mathrm{H}\left(p(i)\right) + \ln 3$, where the average endowment $\overline{p} = 1/3$. Thus, the above inequality constraint is equivalent to restricting the entropy $\mathrm{H}\left(p(i)\right)$ to be at least 1.

$o_{p^*,r}(2) \approx .318$, $o_{p^*,r}(3) \approx .564$. Treating this outcome distribution as fixed, the planner cannot match it with the endowment distribution due to the inequality constraint. However, according to Proposition 2, the planner myopically minimizes systematic redistribution. The selected endowment distribution minimizes $\mathrm{KL}\left(o_{p^*,r}(i) \parallel p(i)\right)$ subject to the inequality constraint. Indeed, the numerical optimization yields the constrained growth-optimal $p^*(i)$, as required.

The myopia is optimal. For contrast, consider a planner who minimizes systematic redistribution, measured by the divergence $\mathrm{KL}\left(o_{p,r}(i) \parallel p(i)\right)$ between the outcome and endowment distributions, subject to the same inequality constraint, while taking into account that endowments affect outcomes. This planner selects $p(1) \approx .148$, $p(2) \approx .362$, and $p(3) \approx .490$, achieving a slightly lower growth rate of .354. ▲

4 Analysis

This section proves the main result and offers intuition.

4.1 Generalized Distributions

To facilitate the use of tools from information theory, we now relabel the notation for the policy pairs. We continue to denote the endowment distribution by $p(i) \in \Delta(I)$. However, instead of $r(i, \omega)$, we write $p(\omega \mid i)$, a function $I \times \Omega \to \mathbb{R}_+$, for the return function. We treat $p(\omega \mid i)$, $i \in I$, as generalized conditional distributions (which need not be normalized) and introduce

$$p(i, \omega) := p(i)p(\omega \mid i).$$

Note that $p(i, \omega)$ is a sufficient summary statistic of the policy pair for the aggregate growth rate (1), because this rate depends solely on this product. Therefore, we refer to $p(i, \omega)$ as the *policy*. The set $\tilde{\mathcal{P}}$ of feasible policies consists of those $p(i, \omega) = p(i)p(\omega \mid i)$ that can be expressed as a product of a feasible policy pair $\left(p(i), p(\omega \mid i)\right) \in \mathcal{P}$.

Once again, we treat $p(i, \omega)$ as a generalized (non-normalized) joint distribution and apply standard probability-theory operations to it. For example, we define the conditional policy

$$p(i \mid \omega) = \frac{p(i, \omega)}{\sum_i p(i, \omega)} \tag{5}$$

through a formal application of Bayes' rule. This represents the end-of-period wealth share of individual i in state ω and is equivalent to the state-contingent outcome distribution $o(i \mid \omega)$ induced by the policy $p(i, \omega)$, as previously defined in (2).

We extend the standard definition of KL-divergence to a map that takes any well-normalized joint distribution $q(i, \omega)$ and any policy $p(i, \omega)$ as arguments. It is straightforward to verify that the usual chain rule continues to hold:

$$\mathrm{KL}\left(q(i,\omega) \parallel p(i,\omega)\right) = \mathrm{KL}\left(q(i) \parallel p(i)\right) + \sum_i q(i)\,\mathrm{KL}\left(q(\omega \mid i) \parallel p(\omega \mid i)\right),$$

where the summation is over $i \in \operatorname{supp}(q(i)) \cap \operatorname{supp}(p(i))$. An analogous equality holds when the roles of i and ω are reversed.

4.2 Wealth Circulation

We analyze the economy's growth rate by tracking the circulation of wealth among individuals. To introduce the concept of wealth circulation informally, we imagine each dollar drawn from the initial stock of perfectly divisible aggregate wealth as founding a separate dynasty of subsequent wealth. In each period, each such dynasty is held by an individual i, randomly drawn from the endowment distribution $p(i)$, and in a random aggregate state ω of the economy, independently drawn from $q^0(\omega)$. Consequently, the dynasty's wealth multiplies by the return $p(\omega \mid i)$. Without loss of generality, we assume that the dynasties do not mix. That is, all wealth generated by the compounding returns to each dollar initially founding a dynasty circulates within the economy as separate infinitesimal fractions of the aggregate wealth, each representing a distinct dynasty.

Up to a given time horizon, let the dynasty-specific distribution $q(i, \omega)$ describe the frequency of individuals and states occupied by this dynasty; we refer to $q(i, \omega)$ as the dynasty's path. Since the shocks ω_t are aggregate, all paths satisfy the consistency condition $q(\omega) = q^0(\omega)$ in the long run. Accordingly, we formally define the *path* as any joint distribution $q(i, \omega) \in \Delta(I \times \Omega)$ that satisfies $q(\omega) = q^0(\omega)$.

For any finite horizon, the dynasties differ in the frequencies they have spent in the hands of each individual i, and in how their allocation to various individuals correlates with the individuals' random productivities. By chance, a small proportion of lucky dynasties spend a disproportionate fraction of periods in the hands of individuals when these individuals are particularly productive. As it turns out, these fortunate dynasties generate almost all of the economy's growth.[7]

For each path $q(i,\omega)$ and any given time horizon, we define the wealth of the path as the aggregate wealth of all dynasties whose realized path matches $q(i,\omega)$. We show that the wealth of the path $q(i,\omega)$ under policy $p(i,\omega)$ grows exponentially at a long-run rate

$$\begin{aligned} \gamma(p,q) &= \mathrm{E}_{q(i,\omega)} \ln p(\omega \mid i) - \mathrm{E}_{q^0(\omega)} \mathrm{KL}\left(q(i \mid \omega) \parallel p(i)\right) \qquad (6) \\ &= -\mathrm{KL}\left(q(i,\omega) \parallel p(i,\omega)\right) + \text{const.} \end{aligned}$$

To understand the first equality, we decompose the growth rate of a path into two components. First, the growth rate of each separate dynasty following the path $q(i,\omega)$ equals its long-run log-return $\mathrm{E}_{q(i,\omega)} \ln p(\omega \mid i)$. Second, paths that deviate from the endowment distribution $p(i)$ by visiting individuals i with distinct state-contingent frequencies $q(i \mid \omega) \neq p(i)$ involve atypical draws, causing the measure of dynasties following such paths to diminish over time. The measure of such dynasties, and hence their wealth, diminishes at the rate of $\mathrm{E}_{q^0(\omega)} \mathrm{KL}\left(q(i \mid \omega) \parallel p(i)\right)$; the more significant the deviation from the endowment distribution, the faster the decline.[8] The wealth of all dynasties following a given path then grows at a rate given by the difference between the two rates.

The second equality in (6) follows from straightforward algebra. This rearrangement indicates that the closer the path $q(i,\omega)$ aligns with the policy $p(i,\omega)$, the higher the resulting growth rate of wealth.

As the wealth of each path grows exponentially, the wealth of the fastest-growing path ultimately dwarfs that of all other paths. Consequently, the overall growth rate of the economy under a given policy is determined by the path that minimizes divergence from the policy. The maximal feasible growth rate of the economy is then achieved by the combination of the path and policy that minimize their divergence. The following result formalizes this intuition:

Proposition 3. *A policy $p^*(i,\omega)$ maximizes the growth rate of the economy if*

[7] A probability theorist would refer to these atypical paths as 'large deviations' from the generating distribution $p(i)q^0(\omega)$.

[8] This is a consequence of Sanov's theorem, which implies that the chance of drawing a sample with an empirical frequency $q(i)$ from a distribution $p(i)$ decreases at the rate of $\mathrm{KL}\left(q(i) \parallel p(i)\right)$ as the sample expands.

and only if it solves

$$\begin{aligned}\min_{q(i,\omega),p(i,\omega)} \quad & \mathrm{KL}\left(q(i,\omega) \parallel p(i,\omega)\right) \\ \text{s.t.} \quad & p(i,\omega) \in \tilde{\mathcal{P}} \\ & q(i,\omega) \in \Delta(I \times \Omega) \\ & q(\omega) = q^0(\omega),\end{aligned}$$

together with some minimizer $q^*(i,\omega)$*.*

See the Appendix for the proof. This optimization appears in a different context in Kingma and Welling (2013) as a variational approach to maximum likelihood estimation. We explore this inference context in Section 5. In our parallel project, Samuelson and Steiner (2024), we focus on behavioral constraints imposed on inference.

4.3 Growth-Maximizing Path and Policy

To prove the main result in Proposition 2, we establish an identity between two conceptually distinct objects: the outcome distribution and the growth-maximizing path. For this, we extend the definition of the outcome distribution by introducing a *joint* outcome distribution induced by the policy $p(i,\omega)$, defined as

$$o_p(i,\omega) := q^0(\omega)p(i \mid \omega), \tag{7}$$

where $p(i \mid \omega)$ is given by (5). That is, $o_p(i,\omega)$ is the probability that a dollar, sampled from the aggregate wealth at the end of a random period, is held by individual i in state ω. The outcome distribution, as defined in (3), is then given by the marginalization $o_p(i) = \sum_\omega o_p(i,\omega)$.

Given a policy $p(i,\omega)$, let

$$\begin{aligned} q_p^*(i,\omega) \in \operatorname*{arg\,min}_{q(i,\omega)\in\Delta(I\times\Omega)} \quad & \mathrm{KL}\left(q(i,\omega) \parallel p(i,\omega)\right) \\ \text{s.t.} \quad & q(\omega) = q^0(\omega)\end{aligned} \tag{8}$$

denote the fastest-growing path.

Lemma 1. *For any policy* $p(i,\omega)$*, the fastest-growing path* $q_p^*(i,\omega)$ *equals the joint outcome distribution* $o_p(i,\omega)$*.*

Proof. Using the chain rule and the constraint $q(\omega) = q^0(\omega)$, we express the objective from (8) as

$$\mathrm{KL}\left(q(i,\omega) \parallel p(i,\omega)\right) = \mathrm{KL}\left(q^0(\omega) \parallel p(\omega)\right) + \sum_{\omega} q^0(\omega)\,\mathrm{KL}\left(q(i \mid \omega) \parallel p(i \mid \omega)\right).$$

Minimizing with respect to $q(i \mid \omega)$ implies that $q_p^*(i \mid \omega) = p(i \mid \omega)$ for each ω. The lemma then follows from the definition of $o_p(i,\omega)$ in (7): $o_p(i,\omega) = q^0(\omega)p(i \mid \omega) = q_p^*(i,\omega)$. $\square$

Proposition 3 and Lemma 1 jointly establish that the optimal policy and the resulting outcome distribution are two optimized, decoupled controls. The proof of the main result follows.

Proof of Proposition 2. Let $p^*(i,\omega)$ be the growth-maximizing policy and $q_{p^*}^*(i,\omega)$ the associated fastest-growing path. According to Proposition 3, the policy p^* maximizes the growth rate of this path:

$$\begin{aligned} p^*(i,\omega) &\in \arg\min_{p(i,\omega)\in\tilde{\mathcal{P}}} \mathrm{KL}\left(q_{p^*}^*(i,\omega) \parallel p(i,\omega)\right) \\ &= \arg\min_{p(i,\omega)\in\tilde{\mathcal{P}}} \mathrm{KL}\left(o_{p^*}(i,\omega) \parallel p(i,\omega)\right), \end{aligned}$$

where the equality follows from Lemma 1.

By applying the chain rule, we rewrite this last objective as follows:

$$\mathrm{KL}\left(o_{p^*}(i) \parallel p(i)\right) + \sum_{i\in\mathrm{supp}(o_{p^*}(i))\cap\mathrm{supp}(p(i))} o_{p^*}(i)\,\mathrm{KL}\left(o_{p^*}(\omega \mid i) \parallel p(\omega \mid i)\right).$$

By setting the returns $p(\omega \mid i)$ to the optimized returns $p^*(\omega \mid i)$ and restricting optimization of the last inline expression to the control of the endowment distribution $p(i)$ within $\mathcal{E}^*$, we obtain the statement in (4). $\square$

5 Inference

We conclude by observing that our model of economic growth can be reinterpreted as a process of maximum likelihood estimation. This relabeling translates our main result—the minimization of systematic redistribution—into a relaxed version of Bayes' plausibility condition.

To this end, consider an analyst concerned with two random variables: the observable ω and the latent i, with finite support sets Ω and I. The analyst observes a sample $\omega^n = (\omega_t)_{t=1}^n$ drawn independently from an unknown distribution $q^0(\omega)$, and selects a statistical model $p(i,\omega) \in \tilde{\mathcal{P}}$. Each statistical model p is a joint distribution of the observable and latent variable. Here, the set $\tilde{\mathcal{P}} \subset \Delta(I \times \Omega)$ is the collection of models the analyst considers; it specifies the analyst's preconceived knowledge of the distribution $p(i)$ of the latent variable and of the likelihoods $p(\omega \mid i)$. Given the observed sample ω^n, the analyst selects the maximum likelihood estimate

$$p^n(i,\omega) \in \operatorname*{arg\,max}_{p(i,\omega)\in\tilde{\mathcal{P}}} \prod_{t=1}^{n} p(\omega_t),$$

where $p(\omega) = \sum_i p(i,\omega)$ is the marginalized model. As shown by White (1982), when the sample size n diverges, the estimate converges to the set of models,

$$\operatorname*{arg\,min}_{p(i,\omega)\in\tilde{\mathcal{P}}} \mathrm{KL}\left(q^0(\omega) \parallel p(\omega)\right), \tag{9}$$

that minimize the divergence from the true data-generating distribution $q^0(\omega)$. We refer to any of these minimizers as the *asymptotic estimate.*

Our main result, when translated to the inference context, provides a necessary condition on the asymptotic estimate. To this end, given a model $p(i,\omega)$, we define the average posterior belief

$$o_p(i) = \mathrm{E}_{q^0(\omega)}\, p(i \mid \omega).$$

Specifically, we let the analyst sample the observable ω from the true process $q^0(\omega)$, form Bayesian posteriors $p(i \mid \omega)$ about the latent values i associated with the observed values ω using the estimated model $p(i,\omega)$, and compute the long-run average of these posteriors under the true process generating the observables. As the notation indicates, $o_p(i)$ is formally equivalent to the outcome distribution induced by the economic policy p in our main application.

When the analyst correctly learns the true process that generates observables, so that $p(\omega) = q^0(\omega)$, Bayes' plausibility dictates that $p(i) = o_p(i)$—the analyst's prior $p(i)$ over the latent variable coincides with the average of her posteriors. In general, however, the analyst is misspecified and hence does not learn the true observable process, $p(\omega) \neq q^0(\omega)$. Consequently, the analyst generally fails the analogue of the Bayes' plausibility condition: her prior $p(i)$ differs from the empirical average $o_p(i)$ of her posteriors. The next result, however, states that the analyst myopically minimizes the extent of this discrepancy.

For an asymptotic estimate $p^*(i,\omega)$, let again $\mathcal{E}^* = \{p(i) : p(i)p^*(\omega \mid i) \in \tilde{\mathcal{P}}\}$ be the set of the marginal distributions $p(i)$ to which the analyst can deviate without changing the estimated likelihoods $p^*(\omega \mid i)$.

Corollary 2. *The asymptotic estimate myopically minimizes the failure of plausibility as follows:*

$$p^*(i) \in \operatorname*{arg\,min}_{p(i)\in\mathcal{E}^*} \mathrm{KL}\left(o_{p^*}(i) \parallel p(i)\right). \tag{10}$$

This result differs from the standard asymptotic characterization of misspecified learning by White (1982) and Berk (1966). The typical setting in the misspecified learning literature involves only observable variables. In that setting, the asymptotic estimate is the best feasible approximation of the *true* generating distribution, as in (9). In contrast, condition (10) relates the estimated distribution $p^*(i)$ of the latent variable to the *deduced* distribution $o_{p^*}(i)$. Indeed, $o_{p^*}(i)$ represents the analyst's deduction about the frequencies of the latent counterparts $(i_1,\ldots,i_n)$ of the observed values $(\omega_1,\ldots,\omega_n)$. This deduction is endogenous because it depends on the estimate $p^*(i,\omega)$ through the posteriors $p^*(i \mid \omega)$. Thus, the corollary establishes a fixed-point condition.

As in our main result regarding the optimal economic policies, the minimization in (10) is myopic because it treats $o_{p^*}(i)$ as fixed. A global minimization of $\mathrm{KL}\left(o_p(i) \parallel p(i)\right)$, which acknowledges that the deduced sample distribution $o_p(i)$ of the latent variable depends on the model $p(i,\omega)$, would further reduce the plausibility failure but also worsen the fit. Therefore, an analyst employing maximum likelihood estimation minimizes the extent of the plausibility failure myopically but not globally.

Corollary 2 follows directly from our main result in Proposition 2 because the growth-maximizing objective from (1) and the asymptotic-estimate objective from (9) differ only in sign and a constant:

$$\mathrm{KL}\left(q^0(\omega) \parallel p(\omega)\right) = -\,\mathrm{E}_{q^0(\omega)} \ln \sum_i p(i,\omega) + \mathrm{C},$$

with $C = -\,\mathrm{H}(q^0(\omega))$. Intuitively, the equivalence arises because maximum likelihood estimation can be viewed as a competition of growth processes. The likelihood of each considered model $p(i,\omega)$ grows in a stochastic growth process with a (negative) growth rate $\mathrm{E}_{q^0(\omega)} \ln \sum_i p(i,\omega)$ as the sample size n increases. Thus, asymptotically, the analyst selects the estimate with the highest growth rate, similar to the planner who selects the growth-maximizing policy.

Below, we translate Example 1 from the economic context to the inference context.

Example 2 (Inference)**.** The observable and latent random variables attain values in $\Omega = I = \{1,2,3\}$. The analyst is certain of the likelihoods $p(\omega \mid i) = 1/2$ for $\omega = i$ and $p(\omega \mid i) = 1/4$ otherwise. She does not know the marginal distribution $p(i)$ of the latent variable but is confident that i is stochastic; specifically, the analyst is convinced that $\mathrm{H}\left(p(i)\right) \geq 1$. The true distribution $q^0(\omega)$ generating the observables, unknown to the analyst, is $q^0(1) = .05$, $q^0(2) = .35$, and $q^0(3) = .6$. The analyst observes a large sample of the observable variable draws and learns the distribution $p(i)$ using maximum likelihood estimation.

This inference is equivalent to the inequality-constrained planner's problem from Example 1, with the likelihood function $p(\omega \mid i) = r(i,\omega)/4$ equaling the (renormalized) original return function, and with the stochasticity constraint imposed on $p(i)$ corresponding to the original planner's inequality constraint.

Accordingly, the solution from Example 1 applies, and the analyst's estimate of the latent distribution corresponds to the constrained optimal endowment distribution of the planner. Thus, the analyst forms the estimate $p^*(i)$ with $p^*(1) \approx .158$, $p^*(2) \approx .329$, and $p^*(3) \approx .513$. The analyst is misspecified, with the estimated marginal probabilities $p^*(\omega) = \sum_i p^*(i)p(\omega \mid i)$ of states $\omega = 1$, 2, and 3 approximately equal to .29, .33, and .38, respectively, differing from the true generating distribution $q^0(\omega)$.

Due to the misspecification, the analyst's average posterior belief $o_{p^*}(i) = \mathrm{E}_{q^0(\omega)}\, p^*(i \mid \omega)$ differs from her marginal belief $p^*(i)$, with $o_{p^*}(1) \approx .118$, $o_{p^*}(2) \approx .318$, and $o_{p^*}(3) \approx .564$, as in Example 1. However, according to Corollary 2, the estimate $p^*(i)$ myopically minimizes the measure $\mathrm{KL}\left(o_{p^*}(i) \parallel p(i)\right)$ of this plausibility failure, taking the average $o_{p^*}(i)$ of the posteriors as given. Again, this myopicity is optimal for maximizing the model's likelihood. An alternative estimate $p(i)$ with $p(1) \approx .148$, $p(2) \approx .362$, and $p(3) \approx .490$ minimizes the measure $\mathrm{KL}\left(o_p(i) \parallel p(i)\right)$ of the plausibility failure globally, recognizing that the chosen estimate $p(i)$ affects $o_p(i)$. However, it decreases the model's fit. ▲

6 Summary

We have deliberately ignored all of the usual factors that mediate the effect of redistribution on economic growth, in order to highlight a relatively unexplored factor, arising from the stochastic nature of growth. Growth optimization requires regular redistribution. This optimization is forward-looking, with the optimal policy regularly redistributing incidental wealth gains and losses. However, our main result highlights the adverse effect of *systematic* redistribution on growth. The growth-optimal policy myopically minimizes systematic redistribution, taking economic outcomes as independent of the policy. Specifically, the growth-optimal policy does not globally minimize systematic redistribution. This counterintuitive myopia results from the fixed-point property of the optimal policy, which we derive through a decoupling approach with an intuitive economic interpretation. It remains for further work to embed this role of redistribution in a richer growth model.

Our reinterpretation of the problem in terms of inference yields two additional insights. First, because most analysts are misspecified, their empirical average of posteriors does not match their marginal belief about the latent variable. However, if they estimate their models using maximum likelihood estimation, they myopically minimize this failure of Bayes' plausibility. Thus, Bayes' plausibility serves as a reliable first approximation. Second, the homeomorphism between growth optimization and inference may explain why phenomena resembling inference may arise in distributed systems.

A Proof of Proposition 3

A policy $p(i, \omega)$ generates growth rate of aggregate wealth in each state ω equal to:

$$\begin{aligned}
\ln \sum_i p(i, \omega) &= \frac{1}{t} \ln \left(\sum_i p(i, \omega) \right)^t \\
&= \frac{1}{t} \ln \sum_{\mathbb{i} \in I^t} \prod_i p(i, \omega)^{q_{\mathbb{i}}(i) t},
\end{aligned}$$

where $t \in \mathbb{N}$ is arbitrary, $\mathbb{i} \in I^t$ is a sequence $(i_1, \ldots, i_t)$ and $q_{\mathbb{i}}(i) = \frac{1}{t} \sum_{\tau=1}^{t} \mathbb{1}_{i_\tau = i}$ is the empirical distribution of the sequence $\mathbb{i}$. Since all sequences $\mathbb{i}$ with the same empirical distribution generate the same value of the inline summand, we obtain the growth rate in state ω:

$$\ln \sum_i p(i, \omega) = \frac{1}{t} \ln \left(\sum_{q(i|\omega) \in \Delta_t} n_t\big(q(i \mid \omega)\big) \prod_i p(i, \omega)^{q(i|\omega) t} \right), \tag{11}$$

where $\Delta_t \subset \Delta(I)$ is the set of the empirical distributions that can be generated by sequences $\mathbb{i}$ of length t, and $n_t : \Delta_t \to \mathbb{N}$ maps each distribution q to the number $n_t(q)$ of sequences of length t that generate such an empirical distribution.

The number of sequences of length t with an empirical distribution q can be approximated using the entropy $\mathrm{H}(q)$ as follows:

$$\frac{1}{(t+1)^{|I|}} \exp[t \times \mathrm{H}(q)] \leq n_t(q) \leq \exp[t \times \mathrm{H}(q)], \tag{12}$$

for all $q \in \Delta_t$. See Theorem 11.1.3 in Cover and Thomas (2006) for these bounds.

Substitution of the bounds into (11) yields:

$$\begin{aligned}
& \frac{1}{t} \ln \sum_{q(i|\omega) \in \Delta_t} \exp\left[t \times \left(\mathrm{E}_{q(i|\omega)} \ln p(i, \omega) + \mathrm{H}\left(q(i \mid \omega)\right)\right)\right] - \frac{|I| \ln(t+1)}{t} \\
\leq\ & \ln \sum_i p(i, \omega) \qquad (13) \\
\leq\ & \frac{1}{t} \ln \sum_{q(i|\omega) \in \Delta_t} \exp\left[t \times \left(\mathrm{E}_{q(i|\omega)} \ln p(i, \omega) + \mathrm{H}\left(q(i \mid \omega)\right)\right)\right].
\end{aligned}$$

These bounds further simplify to:

$$
\begin{aligned}
& \max_{q(i|\omega)\in\Delta_t} \left\{ \mathrm{E}_{q(i|\omega)} \ln p(i,\omega) + \mathrm{H}\left(q(i \mid \omega)\right) \right\} - \frac{|I| \ln(t+1)}{t} \\
\leq \quad & \ln \sum_i p(i,\omega) \\
\leq \quad & \max_{q(i|\omega)\in\Delta_t} \left\{ \mathrm{E}_{q(i|\omega)} \ln p(i,\omega) + \mathrm{H}\left(q(i \mid \omega)\right) \right\} + \frac{|I| \ln(t+1)}{t}.
\end{aligned}
$$

For the lower bound, we replaced the sum in lower bound from (13) by its maximal summand. For the upper bound, we replaced each summand in the upper bound from (13) by the maximal summand and noticed that there are at most $(t+1)^{|I|}$ summands. This is because the distributions $q(i \mid \omega) \in \Delta_t$ attain values in $\{\frac{0}{t}, \frac{1}{t}, \ldots, \frac{t}{t}\}$ for each $i \in I$; hence $|\Delta_t| \leq (t+1)^{|I|}$.

Since $\frac{\ln(t+1)}{t}$ vanishes as t diverges, taking the limit $t \to \infty$ yields:

$$
\ln \sum_i p(i,\omega) = \max_{q(i|\omega)\in\Delta(I)} \left\{ \mathrm{E}_{q(i|\omega)} \ln p(i,\omega) + \mathrm{H}\left(q(i \mid \omega)\right) \right\}.
$$

Taking the expectation over $\omega \sim q^0(\omega)$ and straightforward algebraic steps yield the characterization of the aggregate growth rate induced by a policy $p(i,\omega)$:

$$
\begin{aligned}
\mathrm{E}_{q^0(\omega)} \ln \sum_i p(i,\omega) = \quad & \max\nolimits_{q(i,\omega)\in\Delta(I\times\Omega)} \quad \gamma\big(q(i,\omega) \parallel p(i,\omega)\big) \\
& \text{s.t. } q(\omega) = q^0(\omega),
\end{aligned}
$$

where the growth rate $\gamma\big(q(i,\omega) \parallel p(i,\omega)\big)$ is defined as in (6). Finally, optimization over the policy $p(i,\omega)$ yields the proposition.

References

Aghion, P., E. Caroli, and C. Garca-Peñalosa (1999). Inequality and economic growth: The perspective of the new growth theories. *Journal of Economic Literature 37*(4), 1615–1660.

Aridor, G., R. A. da Silveira, and M. Woodford (2024). Information-constrained coordination of economic behavior. Working paper 32113, NBER. Forthcoming, *Journal of Economic Dynamics and Control*.

Aridor, G., F. Grechi, and M. Woodford (2020). Adaptive efficient coding: A variational auto-encoder approach. biorxiv prepring 2020–05, Cold Spring Harbor Laboratory.

Barro, R. J. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth 5*(1), 5–32.

Barro, R. J. and X. Sala-i-Martin (1995). *Economic Growth*. New York: McGraw-Hill.

Benabou, R. (1996a). Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance. *American Economic Review 86*(3), 584–609.

Benabou, R. (1996b). Inequality and growth. *NBER macroeconomics annual 11*, 11–74.

Berg, A., J. D. Ostry, C. G. Tsangarides, and Y. Yakhshilikov (2018). Redistribution, inequality, and growth: New evidence. *Journal of Economic Growth 23*, 259–305.

Berk, R. H. (1966). Limiting behavior of posterior distributions when the model is incorrect. *The Annals of Mathematical Statistics 37*(1), 51–58.

Bohren, J. A. and D. N. Hauser (2023). Behavioral foundations of model misspecification. Technical report, University of Pennsylvania and Aalto University.

Breunig, R. and O. Majeed (2020). Inequality, poverty and economic growth. *International Economics 161*, 83–99.

Brueckner, M. and D. Lederman (2018). Inequality and economic growth: the role of initial income. *Journal of Economic Growth 23*, 341–366.

Chou, C.-f. and G. Talmain (1996). Redistribution and growth: Pareto improvements. *Journal of Economic growth 1*, 505–523.

Cingano, F. (2014). Trends in income inequality and its impact on economic growth. Social, employment and migration working papers no. 163, OECD.

Cover, T. M. and J. A. Thomas (2006). *Elements of Information Theory* (Second ed.). New York: John Wiley and Sons.

De La Croix, D. and M. Doepke (2003). Inequality and growth: why differential fertility matters. *American Economic Review 93*(4), 1091–1113.

El-Shagi, M. and L. Shao (2019). The impact of inequality and redistribution on growth. *Review of Income and Wealth 65*(2), 239–263.

Esponda, I. and D. Pouzo (2016). Berk–Nash equilibrium: A framework for modeling agents with misspecified models. *Econometrica 84*(3), 1093–1130.

Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. *American economic review 90*(4), 869–887.

Frick, M., R. Iijima, and Y. Ishii (2023). Belief convergence under misspecified learning: A martingale approach. *The Review of Economic Studies 90*(2), 781–814.

Galor, O. and O. Moav (2004). From physical to human capital accumulation: Inequality and the process of development. *The review of economic studies 71*(4), 1001–1026.

Galor, O. and D. Tsiddon (1997a). The distribution of human capital and economic growth. *Journal of Economic Growth 2*(1), 93–124.

Galor, O. and D. Tsiddon (1997b). Technological progress, mobility, and economic growth. *The American Economic Review*, 363–382.

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *The review of economic studies 60*(1), 35–52.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1999). An introduction to variational methods for graphical models. *Machine Learning 37*, 183–233.

Kaldor, N. (1957). A model of economic growth. *The economic journal 67*(268), 591–624.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal 35*(4), 917–926.

Kingma, D. P. and M. Welling (2013). Auto-encoding variational Bayes. arxiv preprint arxiv:1312.6114, Cornell University.

Mdingi, K. and S.-Y. Ho (2021). Literature review on income inequality and economic growth. *MethodsX 8*, 101402.

Robson, A., L. Samuelson, and J. Steiner (2023). Decision theory and stochastic growth. *American Economic Review: Insights*. Forthcoming.

Saint-Paul, G. and T. Verdier (1993). Education, democracy and growth. *Journal of development Economics 42*(2), 399–407.

Samuelson, L. and J. Steiner (2023). Growth and likelihood. *CEPR Discussion Papers* (18339).

Samuelson, L. and J. Steiner (2024). Constrained data-fitters. *Yale University and University of Zurich, CERGE-EI, and CTS*.

Stiglitz, J. E. (1969). Distribution of income and wealth among individuals. *Econometrica 37*(3), 382–397.

Strzalecki, T. (2024). Variational Bayes and non-Bayesian updating. *arXiv preprint arXiv:2405.08796*.

Todaro, M. P. and S. C. Smith (2020). *Economic development.* Pearson UK.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–25.