Does repeated application of the Kaldor-Hicks criterion generate Pareto improvements?

Brad Hackinen

University of Victoria

Author note

Brad Hackinen, Honours Program, Department of Economics, University of Victoria

Supervisors: Elisabeth Gugl, Merwan Engineer, University of Victoria

contact: bradhackinen@gmail.com

Abstract

The Kaldor-Hicks criterion is frequently invoked in government decision making. However, surprisingly little research has been performed on what happens when the Kaldor-Hicks criterion is applied to many decisions over a long period of time. One hypothesis is that, in the absence of transfers, benefits and costs will average out in a way that makes everyone better off in the long run (Hotelling, 1938; Hicks, 1941). This paper presents a model in which the Kaldor-Hicks criteria is applied to many proposals with randomly distributed costs and benefits. The model demonstrates that repeated application of the Kaldor-Hicks criterion concentrates wealth. In the long run, this can lead to a winner-takes-all process where one lucky individual accumulates almost all of the wealth. Under these conditions, the Kaldor-Hicks criterion is unlikely to be Pareto improving. Progressive transfers may be necessary to prevent this perverse outcome.

Contents

Does repeated application of the Kaldor-Hicks criterion generate Pareto improvements?

In politics, many policies and proposals are promoted by arguing that they increase the total amount of wealth available to society as a whole. As John F. Kennedy famously proclaimed, "a rising tide lifts all boats." (Kennedy, 1963) The implication of this statement is that, if the total amount of wealth is increased, each person will become wealthier individually. Economists occasionally make similar arguments, but use more careful language. For example, Frank (2000) states, “when the pie is larger, everyone can have a larger slice.” The key difference between Frank’s statement and the preceding John F. Kennedy quote is that Frank only asserts the *possibility* that everyone becomes wealthier individually. This is a readily defensible position. But, it leaves one wondering: If we systematically choose to maximize the total amount of wealth, will everyone increase in wealth individually? Or, will the benefits accrue to a small minority?

The latter outcome may be more consistent with empirical data. Consider the distribution of income in the United States. Over the last 30 years, the top 1% of earners have seen rapid increases in income (fig. 1). In contrast, the bottom 99% of earners have experienced very small gains in income on average (Piketty & Saez, 2010). The Occupy Wall Street protests are the most visible response to this disparity. The protests demonstrate that, for many people, the difference between increasing the net amount of wealth and making everyone richer individually is critically important.

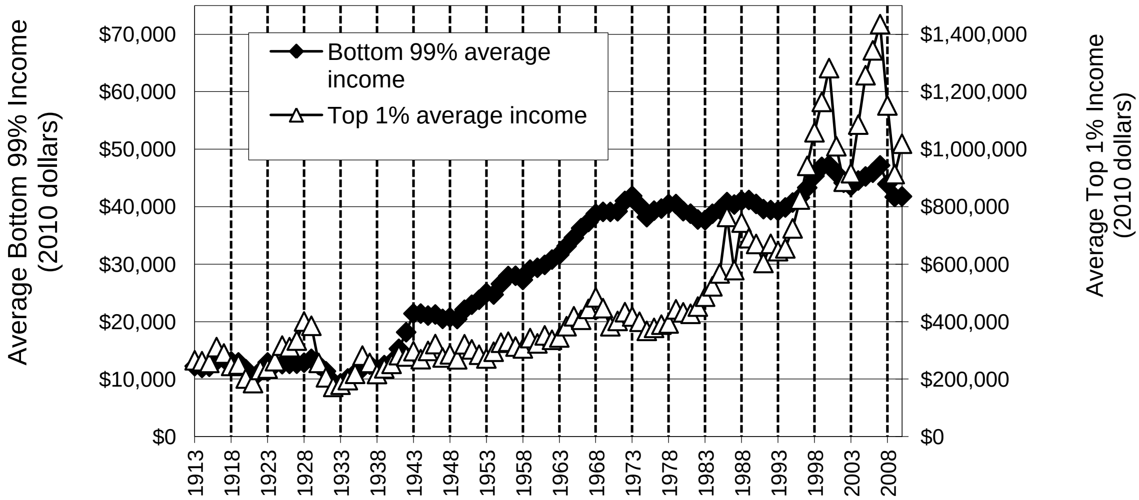

Figure 1. Changes in income for the top 1% and bottom 99% of earners in the United States (Piketty & Saez, 2010)

The idea that a that a policy or project should increase the total amount of wealth is formalized in the Kaldor-Hicks (KH) criterion. This rule states that a proposal should be implemented if and only if the benefits from the policy could be redistributed so that at least one person is made better off and no one is made worse off. For this reason, it is also known as the Potential Pareto Improvement criterion. Unlike the true Pareto Improvement criterion however, the KH criterion does not require that any redistribution actually take place.

Canadian federal regulatory policy is a useful example of how the KH criterion is applied in practice. The federal government makes approximately 350 regulatory changes per year, and requires that every change with significant economic impacts undergo a cost-benefit analysis (Treasury Board of Canada, 2007). Each regulatory change is evaluated as a binary choice between a baseline scenario and the proposed change. The impacts of the change are measured by estimating willingness-to-pay, or changes in social surplus where market data is available. Final approval for each regulatory change is decided at the discretion of a regulatory authority, but federal policy places a strong emphasis on net benefits. It is difficult to say how much regulatory authorities actually consider net benefits in practice. But the policy advocated by the federal government is close approximation of the KH criterion.

What is the effect of applying the KH criterion to hundreds of government decisions per year? One hypothesis is that, in the absence of transfers, benefits and costs will average out in a way that will make everyone wealthier in the long run (Hotelling, 1938; Hicks, 1941). On the other hand, Polinsky (1972) argues that applying the KH criterion to a single project with a long duration and stochastic effects will not necessarily generate Pareto improvements. But to this author's knowledge no one has carefully examined the effect of repeated, compounding applications of the KH criterion. Unfortunately, Hotelling and Hicks' hypothesis has been accepted as fact by practitioners of Cost-Benefit Analysis, with little evidence to support the claim (Boardman et al, 2011).

This paper presents a model of a population of entrepreneurs and one government regulator. The regulator evaluates many potential regulatory changes ('proposals') using the KH criterion. Each proposal, if implemented, has a random effect on each entrepreneur's output. The regulator discards any proposals with negative net benefits, and implements those that have positive net benefits. This process truncates the distribution of proposals and alters the expected value. Over time, the distribution of wealth follows a random walk, with each step determined by the truncated distribution. A general description for the path of this random walk is difficult to derive. However, a number of clear results emerge from examining the expected change each period.

First, applying the KH criterion favours individuals that experience a larger range of potential costs and benefits. I call this variation "exposure". When exposure is proportional to wealth, wealthy individuals benefit from the KH criterion more than poor individuals. The outcomes for each individual are not independent. The mere existence of a wealthy individual in the economy can make poor individuals worse off. Furthermore, in the absence of progressive transfers, this pattern is sufficient to guarantee that relative inequality increases every period. In the long run, one lucky individual ends up with the vast majority of wealth and Pareto improvements become unlikely.

A second important result is that the timing of proposal evaluations matters. Proposals can be evaluated simultaneously, with no interaction between them, or they can be evaluated in series so that one proposal is implemented before the next is evaluated. In the simultaneous case, the law of large numbers causes the distribution of outcomes each period to converge on the expected value. In this case, the gains and losses for each individual do average out to some degree. However, inequality still increases each period, and the probability of Pareto improvements each period declines over time.

Third, I consider the impact of relaxing some of the strict assumptions in the simple model. I show that inequality can still increase if exposure scales less than linearly with wealth, and that correlated proposal effects alter inequality and the chance of Pareto improvements. Finally, I argue that progressive transfers can reduce inequality and increase the chance of Pareto improvements, though the current tax system in Canada does not address the problem in an entirely satisfactory way.

Literature Review

This paper draws on two broad strands of literature. One strand is a large literature which applies the KH criterion to a single proposal. This first strand includes both theoretical research, such as axiomatic analysis of the fair division problem, and more practical guides, critiques, and defenses of the KH criterion as it is applied in Cost-Benefit Analysis. The second strand is a large literature on modeling the long-term distribution of wealth in an economy. This second strand of literature often uses intergenerational models of many agents with stochastic effects on wealth. This paper bridges these two strands by considering the effect of applying the KH criterion in a model of the long-term distribution of wealth.

Repeated Application of the KH criterion

There is little direct research on the effects of repeatedly applying the KH criterion. However, when the criterion was originally being formulated in the early 20th century, the question of long-term effects naturally arose. John Hicks and Harold Hotelling each suggested that the gains and losses to each individual might naturally average out over the long term as many proposals are evaluated, even without transfers. Hotelling (1938) claims, in regard to the criterion,

> a rough randomness in distribution should be ample to ensure such a distribution of benefits that most persons in every part of the country would be better off by reason of the program as a whole. (pp. 259)

Hicks (1941) echoes this assertion, arguing,

> although we could not say, that all the inhabitants of that community would be necessarily better off than they would have been if the community had been organised on some different principle, nevertheless there would be a strong probability that almost all of them would be better off after the lapse of a sufficient length of time. (pp. 111)

Thus, Hotelling and Hicks hypothesize that repeated application of the KH criterion will eventually result in a Pareto improvement, or something very close to it. Neither develops a rigorous argument to back up their assertion. Nonetheless, this hypothesis has found its way into textbooks on Cost-Benefit analysis as a justification for using the KH criterion. In *Cost benefit analysis concepts and practice* (4th edition), the authors state,

> it is likely that different policies will have different sets of winners and losers. Thus, if the rule is consistently applied to government activity, then costs and benefits will tend to average out across people so that each person is likely to realize positive net benefits from the full collection of policies. (Boardmen et al, 2011, pp. 32)

The first (and apparently only) attempt to develop a more definitive model of repeated application of the KH criterion is Polinsky (1972). Polinsky attempts to evaluate the probabilistic claim made by Hicks and Hotelling. His model considers the case of a policy that has a long duration, so that people might experience different benefits and costs over time. Polinsky uses Markov analysis to show that there is no reason to expect the payoff from the KH criterion to be positive for each individual. But he does not consider an environment where many different proposals are evaluated. This paper expands on Polinsky's analysis by considering an environment with compounding proposal effects. In doing so, it fills a gap in the literature that has remained vacant for decades.

The Fair Division problem

The task facing the federal government is closely related to the fair division problem. A new policy offers potential benefits and costs which can be seen as 'common property' until the government assigns them to individuals by implementing the project and any associated transfers. There are many constraints on this procedure. Moulin (1988) shows that, when the amount of common property can grow, there is no allocation which satisfies Pareto-optimality, resource monotonicity, and individual rationality from equal division. Resource monotonicity requires that increasing the initial endowment does not make any one person worse off, and individual rationality from equal division requires that each person is at least as well off with their allocation as if they simply consumed an even share of the endowment.

Serizawa (2002) shows that, for an exchange economy with a finite number of agents and a finite number of goods, there is no strategy-proof, Pareto-efficient and individually rational allocation rule on the class of classical, homothetic and smooth preferences. Similarly, there is no strategy-proof, Pareto-efficient and symmetric allocation rule on the same class of preferences (Serizawa, 2002). If an allocation rule is not strategy proof, citizens may have an incentive to hide their preferences, making effective implementation unrealistic. This implies that in practice, governments are unlikely to be able to distribute wealth however they like.

The KH criterion can be seen as a response to these constraints: it maximizes net wealth by ignoring the restrictive requirements of distributional fairness. Disregard for distributional effects is not necessarily undesirable. Kaldor (1939) argues that questions of efficiency and distribution can and should be separated. Kaldor believed that the role of economic policy is to maximize the total amount of wealth, while abstaining from commenting on the division of that wealth. He saw distribution as a purely political question. From this perspective, the KH criterion is a distribution-neutral tool for economic analysis. But separating efficiency and distribution depends critically on the ability of the government to reallocate wealth in accordance with political will. Progressive taxes and welfare payments represent a compromise in this regard. They each reduce inequality, but do not guarantee that every loss is compensated or that the amount of compensation reflects the magnitude of the loss. Instead, these general redistribution schemes rely on statistical effects to average out costs, benefits and compensation for each individual.

Models of Wealth Distribution

Research into the distribution of wealth has a rich history of its own. Vilfred Pareto famously argued that the distribution of wealth follows a universal law, now known in economics as a 'Pareto distribution'. In the Pareto distribution, the distribution of wealth in society is highly skewed, so that the number of people at each level of wealth decreases as wealth increases: poor people being very common, and the rich very rare.

Chapernowne (1953) provided a seminal analysis of the Pareto distribution, exploring how it could be generated by stochastic changes in wealth over time. This has led to more advanced models, such as the one presented by Banerjee & Newman (1991). When people can purchase insurance or leverage investments with loans, the amount that each person is affected by stochastic processes may be at least partially under the control of that individual. Banerjee & Newman incorporate risk aversion and lending into a multi-generational model with stochastic affects and show that it also generates a skewed distribution of wealth (although not nearly as skewed as the Pareto distribution).

More recently, Fargione, Lehman & Polasky (2011) present a much simpler model as a reminder that stochastic effects alone can explain a large amount of inequality in the distribution of wealth. They consider a population of entrepreneurs who earn a random return each generation, and then leave a fraction of that wealth as a bequest to their children. The model generates a log-normal distribution of wealth, which is similar to the Pareto distribution at high levels of wealth (although Fargione, Lehman & Polasky argue that the log-normal distribution actually provides a much better fit for the data). Fargione, Lehman & Polasky show that as time progresses, wealth becomes more and more concentrated among a few entrepreneurs. However, there is still a lot of mobility - the wealthiest entrepreneur may not stay wealthy for long.

The simplicity of the Fargione, Legman & Polasky model makes it a useful starting point for building more complex models. I have used it for that purpose in this paper. One could view the model presented in this paper as an extension of the Fargione, Lehman & Polasky model, in which the government applies the KH criterion and eliminates periods in which entrepreneurs earn net-negative returns.

The model

Overview

The model consists of a population of n entrepreneurs and one government regulator. Time progresses in discrete periods, with each period representing one generation of entrepreneurs. At the start of each period, all entrepreneurs are identical except that each receives a different initial endowment of capital. Each entrepreneur invests his or her entire endowment in a new firm. The government takes investments as given and attempts to maximize net output by adjusting regulations according to the Kaldor-Hicks criterion. Firms produce output after the government has finished adjusting regulations; output is altered by the government's actions. Finally, each entrepreneur consumes a fraction of their firm's output, and leaves the rest as a bequest for their child. The bequest becomes the child's capital endowment in the next period. This process is repeated for many periods so that the distribution of wealth evolves over time.

Entrepreneurs

I choose to focus on entrepreneurs for a number of reasons. The first is that the existence of capital is very important in the model. In real life, government decisions have a broad range of effects on individuals, and many of these effects will be independent of wealth. But those government decisions that depend on wealth may also affect wealth. This creates the potential for feedback loops to arise. While it is conceivable that feedback loops might be possible in other areas (perhaps health or education), the logic is less obvious. The example of capital is therefore somewhat illustrative. Furthermore, the magnitude of effects created by wealth feedback loops are so large that I expect them to dwarf all others. Thus, as a first approximation, I believe it is reasonable to ignore government decisions that do not have effects related to an individual's wealth.

Of course, many people hold capital without being entrepreneurs. In this respect, treating all agents as entrepreneurs is a simplifying assumption. I follow Fargione, Lehman & Polasky (2012) who choose to model entrepreneurs because of their investing habits. Entrepreneurs tend to invest primarily in their own venture, and are remarkably undiversified as investors. This makes them somewhat idiosyncratic compared to the population at large, but also makes modeling their behavior much easier. Fargione, Lehman & Polasky argue that entrepreneurs also make up a large portion of the wealthiest individuals in society, so that any patterns found in the changes of wealth of entrepreneurs is likely to affect the broader distribution of wealth, even if it does not directly apply to normal investors. For a discussion of what occurs when the assumption that all agents are entrepreneurs is relaxed, see the later section on diversification and nonlinear effects on wealth.

Investment

Each entrepreneur is initially endowed with capital k_i. At the start of their life, each entrepreneur starts a new individually owned and managed firm. Each firm is affected uniquely by government regulation. The source of this variation is not explicitly modeled, but one could imagine that (as in reality) firms are differentiated along many different dimensions such as location, production technologies, and material inputs. I assume that entrepreneurs are able to design their firms so that, in expectation, each firm receives the same output per unit of capital. However, entrepreneurs do not know how the regulatory environment will change after they make their investment.

Regulatory Change

The government maintains a set of regulations which define a *regulatory environment* for the entrepreneurs. The regulatory environment determines the productive efficiency of each firm. The government wants to choose the set of regulations which maximizes total output. However, it has a limited ability to gather and process information. It cannot consider all possible regulatory environments and choose the optimum. Instead, the government proceeds by iteration: first, it takes a guess and generates a set of regulatory changes, grouped together as one proposal. I assume that each proposal, if implemented, adjusts every firm's production efficiency by a different amount ε_i, where ε_i is normally distributed with mean μ and variance σ^2. Then the government evaluates the effects of implementing the proposal with the KH criterion. The government knows that each firm will produce,

$$y_i = (1 + \varepsilon_i)k_i$$

Let $\delta_i = \varepsilon_i k_i$ be the total change in production resulting from a given proposal. Then,

$$\delta_i \sim N(\mu k_i, \sigma^2 k_i^2)$$

And the proposal passes the KH criterion if,

$$\sum_{i=1}^{n} \delta_i > 0$$

If the proposal passes, the government implements it immediately. Otherwise the proposal is discarded and forgotten. The rejection of proposals that fail the KH criterion truncates the distribution of δ_i below the hyperplane $\sum_{i=1}^{n} \delta_i = 0$. I denote the random change in wealth for each entrepreneur after the KH criterion has been applied as δ_i^{KH}.

Evaluating a proposal takes time, so the government evaluates a set of S proposals simultaneously each period (and only implements a subset). These evaluations take place in parallel, meaning that the outcome of each proposal evaluation cannot be considered in analysis of other proposals this period and the effects of each implemented proposal are statistically independent. The total effect of all simultaneous implemented proposals on a given entrepreneur is then,

$$\Delta_i = \sum_{s=1}^{S} \delta_{ist}^{KH}$$

Consumption and Bequests

At the end of each period, each entrepreneur produces output,

$$y_{it} = k_{it} + \Delta_{it}$$

Entrepreneurs consume a fixed fraction c of their output. The remaining fraction is left as a bequest to their offspring in the next period. Thus,

$$k_{i(t+1)} = (1 - c)(k_{it} + \Delta_{it})$$

The assumption that entrepreneurs consume a fixed fraction of their output is a another important simplification. In a more complex model, one might want to include the possibility that rich and poor entrepreneurs save differently. However, I am again following Fargione, Lehman & Polasky (2011) in ignoring this potential complication.

Results

Analytic Approach

The model describes a random walk that is very difficult to solve in general. Two analytical approaches that seem promising at first are ruled out for most cases. However, in the specific case where the number of simultaneous projects is very large, the central limit theorem provides a useful approximation of long-term changes in wealth.

There are two related reasons why solving the model is hard. First, the probability density function of δ_i^{KH} depends on the wealth of all entrepreneurs. Second, the change in wealth from each implemented proposal is related to each agent's current wealth. Therefore, changes in wealth not statistically independent from one period to the next.

If the government implemented every proposal regardless of the net benefits, the resulting changes in wealth each period could be made statistically independent by taking the log transform of each entrepreneur's wealth. This simplification would essentially replicate the model analyzed by Fargione, Lehman and Polasky (2011). However, the probability density function of δ_i^{KH} is not simplified by taking a log transform. The truncation boundary created by the KH criterion becomes a more complicated curve, and the shape of this curve changes with the wealth of each agent. This changing shape breaks the statistical independence gained by taking the log transform.

Another approach that is ruled out in most cases is to approximate the model as a random walk with drift. This approach depends on calculating the expected value of δ_i^{KH} and then invoking the central limit theorem to show that that as the number of proposals increases the distribution of outcomes becomes tightly clustered around the net expected value. However, the central limit theorem requires statistical independence between each random variable, so this strategy cannot be applied across multiple periods.

The central limit theorem can, however, be applied within a single period. Simultaneous proposals are statistically independent by assumption, so each period can be treated *individually* as a random walk with drift. If the number of simultaneous proposals is sufficiently large, the outcome of each period is close to certain. Luckily, the assumption that there are a large number of simultaneous proposals appears to be reasonable. The federal government reports making on the order of 350 regulatory changes every year (Treasury Board of Canada, 2007). It seems unlikely that these proposals are highly coordinated, or that the results from each proposal evaluation affect many other proposals that year. Thus, it seems reasonable to suppose that there are hundreds of proposal evaluations that are simultaneous in terms of the model. This quantity is large enough for the central limit theorem to provide an effective approximation of changes in wealth each period.

The approach adopted in this paper is to focus on changes in wealth when the number of simultaneous proposal evaluations is large. Under these conditions, changes in wealth each period are close to deterministic. Then I use simulations to shed light on how changing the number of simultaneous proposals affects the results.

Changes in wealth when *S* is large

The change in wealth for each entrepreneur each period depend entirely on δ_i^{KH}. The distribution of δ_i^{KH} is easiest to visualize for a population of two entrepreneurs:

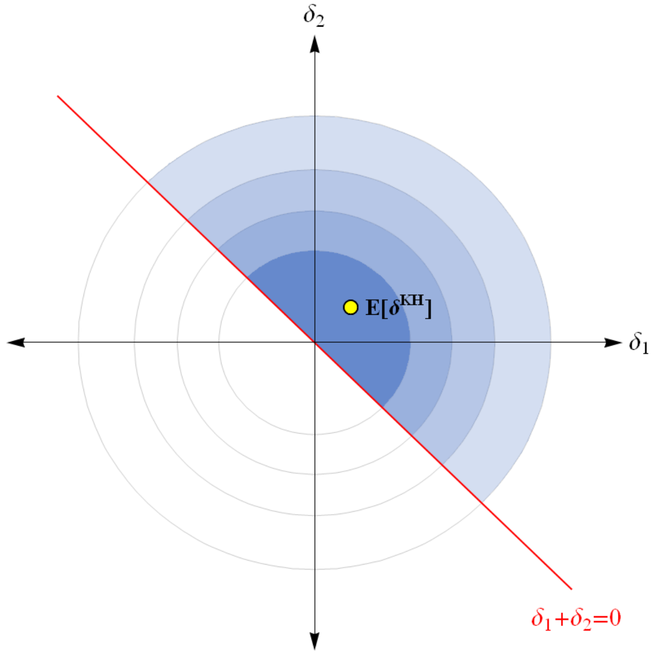

Figure 2. The distribution of proposal effects for a population of two entrepreneurs

Figure 2 shows the changes in wealth created by each random proposal for a population of two entrepreneurs. Each axis represents the change in wealth for one entrepreneur. Proposals are randomly distributed, with the shading showing the probability density. The KH criterion eliminates all proposals with negative net value (everything below the -45 degree line). This truncates the distribution and changes the expected value. It can be shown (see Appendix 1) that the new expected value of the distribution for each entrepreneur is,

$$E[\delta_i^{KH}] = \frac{e^{-\frac{a^2}{2}}\sigma^2 k_i^2}{\sqrt{2\pi n\left(v^2 + \bar{k}^2\right)}} + \frac{1}{2}\operatorname{erfc}\left[a/\sqrt{2}\right]\mu k_i$$

$$a = -\frac{\mu\bar{k}}{\sqrt{\frac{1}{n}\left(v^2 + \bar{k}^2\right)}}$$

Where $\bar{k}$ is the average amount of capital per entrepreneur, and v^2 is the population variance in capital ownership among the entrepreneurs. In this simple version of the model, $E[\delta_i^{KH}]$ is always positive.

The KH criterion also makes the changes in wealth for each entrepreneur interdependent. In particular, $E[\delta_i^{KH}]$is decreasing in $\bar{k}$ and v^2. Thus, increasing the wealth of one entrepreneur raises $\bar{k}$ and v^2, decreasing $E[\delta_i^{KH}]$ for every other entrepreneur. In the limit, as one entrepreneur becomes infinitely wealthy, $E[\delta_i^{KH}]$ goes to zero for all other entrepreneurs.

Is a Pareto improvement likely? The change in wealth for entrepreneur i in period t is,

$$\Delta_{it} = \sum_{s=1}^{S} \delta_{i\,s\,t}^{KH}$$

Thus, a Pareto improvement occurs if $\Delta_{it} \geq 0$ for all i. The probability of a Pareto improvement is then,

$$P\{\Delta_{it} \geq 0 \ \forall \ i\} = \prod_{i=1}^{n} P\{\Delta_{it} \geq 0 \}$$

The central limit theorem implies that Δ_{it} becomes approximately normally distributed for large S:

$$\Delta_{it} \sim N(S \cdot E[\delta_i^{KH}], S \cdot Var[\delta_i^{KH}]) \ as \ \mathrm{S} \to \infty$$

Let Φ denote the cumulative distribution function of a standard normal distribution. Then,

$$\lim_{S\to\infty} P\{\Delta_{it} \geq 0 \} = 1 - \lim_{\mathrm{S}\to\infty} \Phi\left(\frac{0 - S \cdot E[\delta_i^{KH}]}{\sqrt{S \cdot Var[\delta_i^{KH}]}}\right) = 1 - \lim_{S\to\infty} \Phi\left(\frac{-\sqrt{\mathrm{S}} \cdot E[\delta_i^{KH}]}{\sqrt{Var[\delta_i^{KH}]}}\right)$$

$$= \begin{cases} 1 - \Phi(-\infty) = 1 \text{ for } E[\delta_i^{KH}] > 0 \\ 1 - \Phi(\infty) = 0 \text{ for } E[\delta_i^{KH}] < 0 \end{cases}$$

$E[\delta_i^{KH}]$ is strictly positive for every entrepreneur, so if the number of simultaneous proposals is sufficiently large, a Pareto improvement will occur every period. However, the number of simultaneous proposals required to generate a Pareto improvement may be impractically large. In particular, as $\bar{k}$ and v^2 increase, $E[\delta_i^{KH}]$ becomes very small, and the probability of a Pareto improvement declines for any finite S (this is not immediately obvious from the equations above, because $Var[\delta_i^{KH}]$ is also changing, but it is clearly demonstrated by simulation results in the next section). In the limit, as portion of wealth held by one entrepreneur goes to one, the probability of a Pareto Improvement goes to zero. Unfortunately, the KH criterion makes this outcome inevitable in the long run.

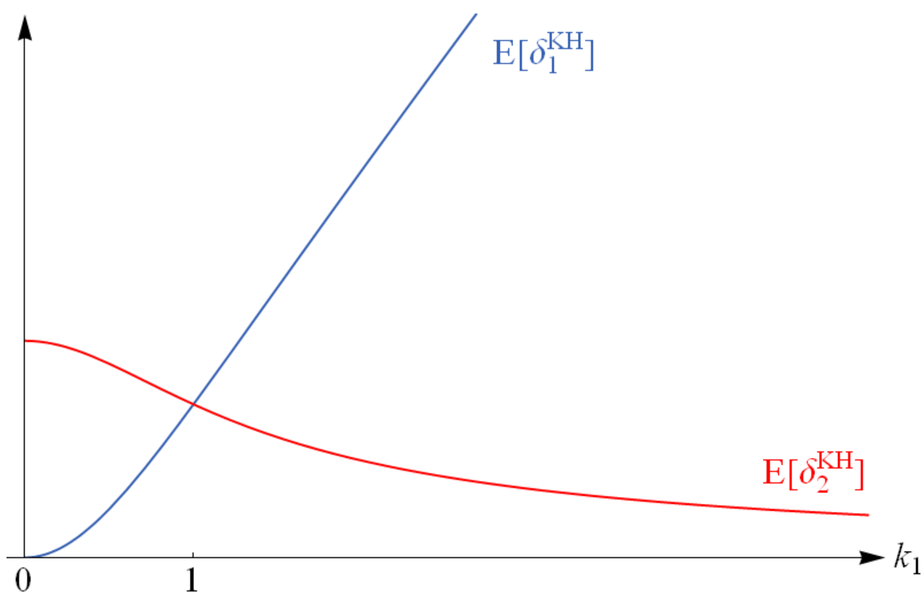

Figure 3. The effect of changing the wealth of one entrepreneur $(k_2 = 1)$

Inequality rises each period because $E[\delta_i^{KH}]$ increases with the *square* of k_i for each entrepreneur. On average, rich entrepreneurs receive proportionally larger benefits than poor entrepreneurs. This pattern is a direct consequence of the KH criterion: before the criterion is applied, the distribution is 'fair' in the sense that the expected value and standard deviation are proportional to each agent's capital. After the government applies the KH criterion, rich entrepreneurs benefit proportionally more than poor entrepreneurs. When the number of simultaneous projects is sufficiently large, inequality increases every period (see proof in Appendix 2). As inequality increases, the payoffs to poor agents become smaller and smaller and Pareto improvements become less and less likely to occur.

Pareto improvements may still occur over longer time periods, but the initial level of inequality matters. Consider the case where all entrepreneurs start with the same amount of wealth at time $t = 0$. Initially, inequality is low and all entrepreneurs will tend to gain in wealth. As time progresses, inequality rises and the gains each period will decrease for most entrepreneurs. Eventually most entrepreneurs will receive random benefits with expected value of zero. However, it is very likely that all entrepreneurs will still be better off than at time $t = 0$.

Alternatively, consider a case where inequality is already quite high at $t = 0$. Then poor agents will already have a fairly low probability of increasing in wealth. Many poor entrepreneurs will be unlucky enough to lose wealth on average over the long term.

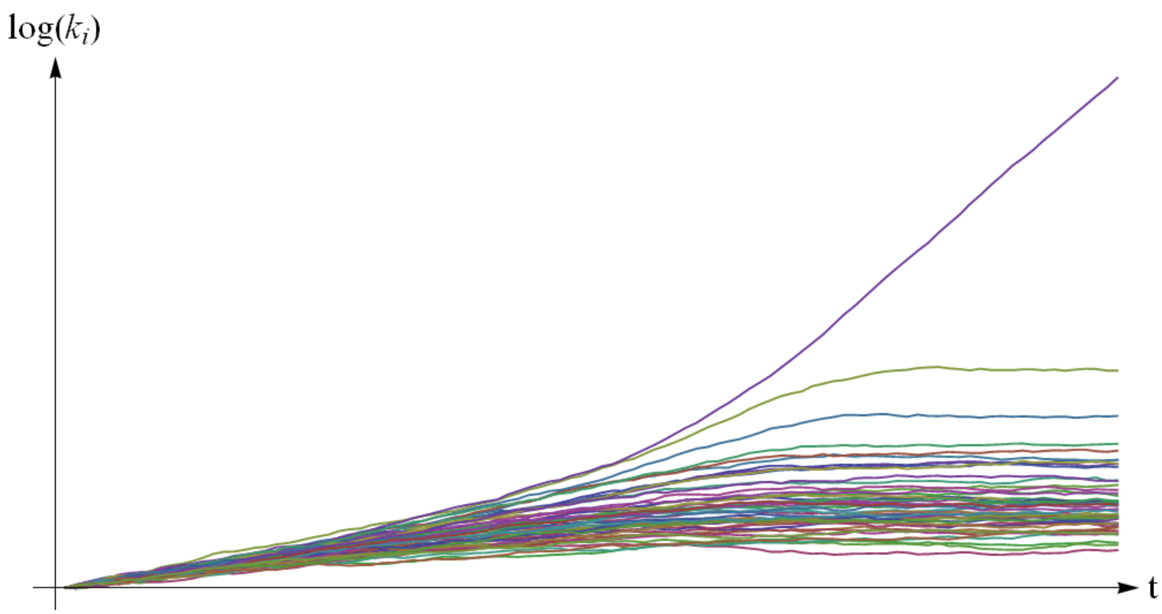

Figure 4. Typical simulation output for $n = 50,\ S = 350,\ T = 100,\ \mu = 0,\ \sigma = 0.002$

Figure 4 shows the simulated effect of applying the Kaldor-Hicks criterion to many random projects when the number of simultaneous proposals is large. Each line represents one person's wealth as it changes over time. All entrepreneurs start with $k_i = 1$, but eventually one lucky individual ends up with most of the gains while the rest lag behind.

Changes in wealth when S is small

As discussed earlier, it is much harder to determine what happens when the number of simultaneous proposals is small. Changes in wealth for each entrepreneur differ from the expected value, and these differences can compound over multiple periods. It is also harder to prove that inequality will increase. For example, in the case where $S = 1$, the fact that expectation is a linear operator only implies,

$$\mathrm{E}\left[\frac{\Delta_{it}}{\Delta_{jt}}\right] = \mathrm{E}\left[\frac{\delta_{it}^{KH}}{\delta_{jt}^{KH}}\right] \neq \frac{E[\delta_{it}^{KH}]}{E\left[\delta_{jt}^{KH}\right]}$$

I use simulations to overcome these difficulties. The following output shows mean wealth, inequality, and the portion of entrepreneurs who increase in wealth for $S = 1{,}10{,}100$ and 1000. To make the output comparable, the parameter σ is defined as $\sigma = 1/S$. This adjustment holds $\mathrm{E}[\Delta_{it}]$ constant for a given distribution of wealth. Then adjusting S is equivalent to comparing the effects of many small proposals each period with fewer, larger proposals each period.

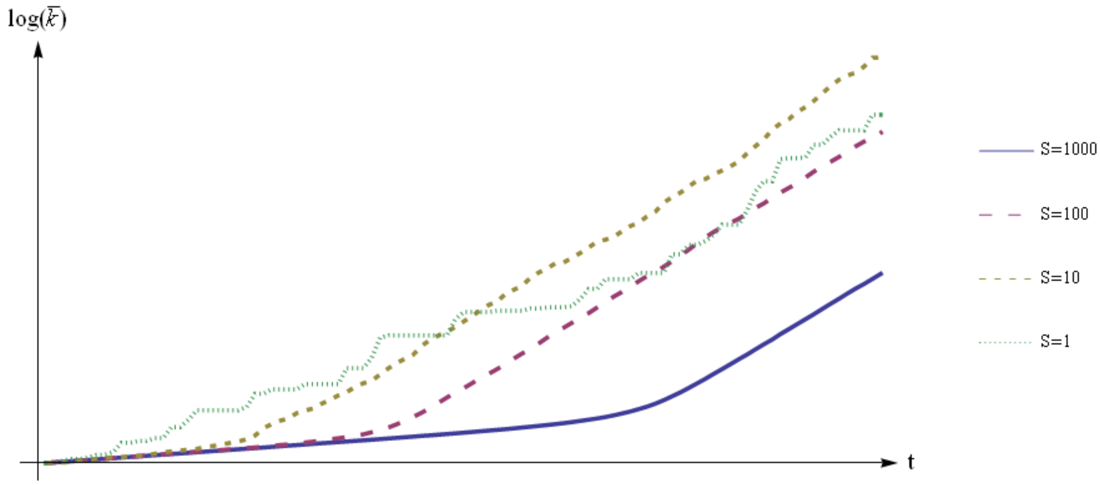

Figure 5. Effect of varying S on changes in mean wealth

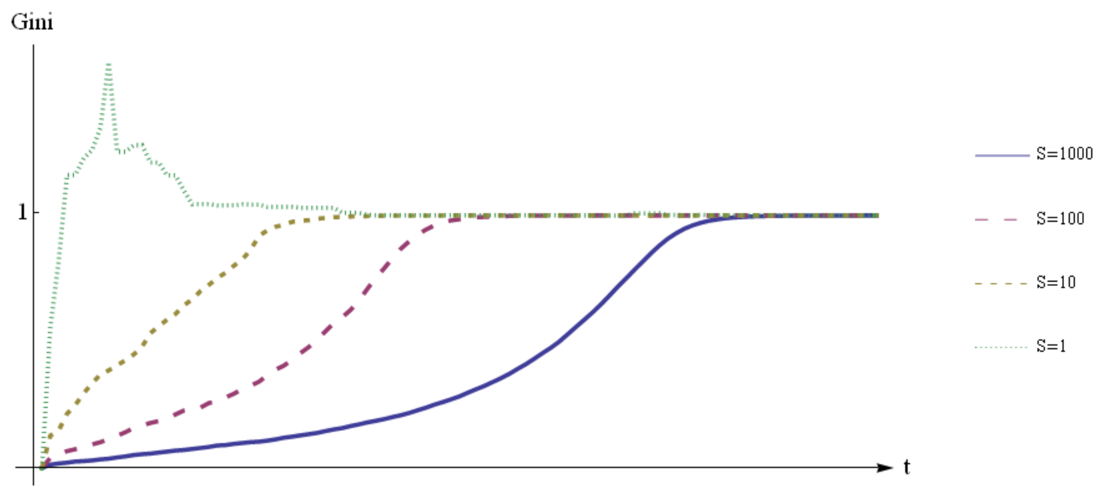

Figure 6. Effect of varying S on changes in inequality (as measured by the Gini coefficient)

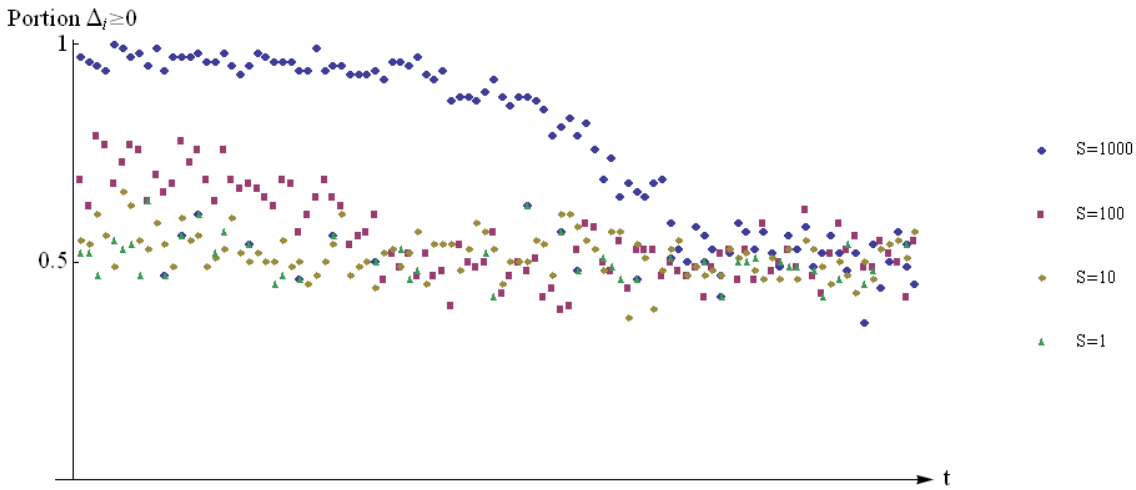

Figure 7. Effect of varying S on the portion of entrepreneurs who increase in wealth each period

The simulations reveal that, when holding $\mathrm{E}[\Delta_{it}]$ constant, the large S scenario places a lower bound on increases in mean wealth and inequality, and a rough upper bound on the portion of entrepreneurs who increase in wealth each period[1]. In the short run, there appears to be a tradeoff between growing wealth and rising inequality. But eventually all scenarios converge to a similar pattern: rapidly rising wealth accompanied by near-perfect inequality, with most entrepreneurs gaining wealth only 50% of the time. Lowering S speeds up the transition to the long-run pattern.

In general, these simulated scenarios are consistent with the mathematical analysis. In the long run, one entrepreneur pulls out ahead of the rest of the population and from then on receives essentially all of the benefits, raising mean wealth, but also creating perfect inequality. The rest

[1] A note about the Gini output: The $S = 1$ scenario often shows Gini values greater than one. This occurs when some entrepreneurs end up with negative wealth. It is not entirely clear if negative wealth has a reasonable interpretation in the model, but correcting the problem requires adding another layer of complexity, so I have chosen not to address it at this time.

of the entrepreneurs experience a mix of positive and negative returns. The mathematic results show that expected value of this mix of returns goes to zero in the long run.

Additional considerations

Diversification and Nonlinear effects on wealth

The basic model discussed so far assumes that each entrepreneur's exposure is proportional to the amount of capital that they own. This is a strong assumption, and one that is probably only rarely true in the strict sense. In particular, most people are not entrepreneurs, and even real entrepreneurs do not invest exclusively in their own firm (Goetzmann & Kumar, 2008). When people have the opportunity to diversify their investments, analysis becomes much more complicated. The effect of each proposal on a given individual's wealth may depend highly on that individual's investment strategy. What happens when the assumption of linearity is relaxed, and what does this imply for more realistic models of investment?

Consider the ratio $E[\delta_i^{KH}]/k_i$. As discussed in the Appendix 2, this ratio gives the proportional change in wealth each period when S is large. If the proportion of wealth is higher for wealthy entrepreneurs than poor ones, inequality increases each period. Taking the derivative of the ratio with regards to k_i gives the rate of increase in the ratio at each level of wealth.

$$\frac{\partial E[\delta_i^{KH}]/k_{it}}{\partial k_{it}} = \frac{e^{-\frac{a^2}{2}}\sigma_{it}^2}{\sqrt{2\pi n\left(v^2+\bar{k}^2\right)}}\left(2\frac{\partial\sigma_{it}}{\partial k_{it}}\frac{k_{it}}{\sigma_{it}}-1\right)+\frac{1}{2}\operatorname{erfc}\left[a/\sqrt{2}\right]\left(\frac{\partial\mu_{it}}{\partial k_{it}}k_{it}-\mu_{it}\right)$$

The derivative is a bit messy, but it shows that the rate of increase in inequality will be an increasing function of the elasticities of σ_i and μ_i with regard to wealth. More specifically, a sufficient condition for inequality to always increase is,

$$\frac{\partial \sigma_{it}}{\partial k_{it}} \frac{k_{it}}{\sigma_{it}} > \frac{1}{2} \ and \ \frac{\partial \mu_{it}}{\partial k_{it}} \frac{k_{it}}{\mu_{it}} > 1$$

If $\mu_{it} = 0$, only the elasticity of σ_i matters and the first restriction is sufficient for inequality to always increase.

Thus, the main results hold in more general situations than strict linearity with regards to capital. One way to interpret the sufficient condition outlined above is to consider the following scenario: imagine that all entrepreneurs are replaced by normal investors. There are a large number of firms in the economy, and the investors purchase shares in the firms. Each investor is risk averse, and diversification is costless. Investors therefore diversify as much as possible. The limiting factor is that there is a minimum share size, so that each investor can only own a number of shares that is proportional to their total capital. When the government adjusts regulation, it still affects each firm's output in proportion to its capital, but the investors are now less affected because of diversification. In fact, the elasticities of σ_i and μ_i with regard to wealth will be exactly $1/2$ and 1 respectively. Thus, this scenario represents the edge case where increasing inequality can no longer be guaranteed. If investors diversify any less than this amount, inequality will increase every period.

In reality, investors diversify *much* less than required by the edge-case scenario. In the U.S., investors only hold 4 or 5 stocks on average, and while wealthy investors tend to hold more stocks than poor investors, the differences are not very large (Goetzmann & Kumar, 2008). Thus it is reasonable to infer that condition above will be met, and that inequality will increase every period.

Correlated regulatory effects

It might seem unreasonable to assume that the effect of regulatory change is statistically independent for all entrepreneurs. In particular, one might imagine that regulations affect the provision of public goods, or alter broad-reaching externalities. In this case, when the government makes a change to regulatory policy, changes in output would be positively correlated across entrepreneurs. Alternatively, one might imagine that firms operate in a competitive environment, where one firm's loss is another firm's gain. In this case, changes in output would be negatively correlated across entrepreneurs. Of course it is likely both of these opposing forces operate in the economy. Without knowing the magnitude of each effect, there is no reason to believe that changes in output should be strictly uncorrelated. What happens when the assumption of statistically independent changes in output is relaxed?

In appendix 1, I derive the expected value of $E[\delta_i^{KH}]$ given the assumption that all changes in output are correlated by a common correlation coefficient ρ. This is a strong assumption, but it is enough to give a sense of what effect correlations can have. Once again, I return to the ratio $E[\delta_i^{KH}]/k_i$. Recall that this ratio gives the proportional change in wealth each period when S is large. If the proportion of wealth is higher for wealthy entrepreneurs than poor ones, inequality increases each period. The effect of changing ρ on this ratio is:

$$\frac{\partial E[\delta_i^{KH}]/k_{it}}{\partial \rho} = \frac{(n\bar{\sigma}_t\sigma_{it} - \sigma_{it}^2)e^{-\frac{a^2}{2}}/k_{it}}{\sqrt{2\pi(n(1-\rho)(v^2+\bar{\sigma}_t^2)+\rho n^2\bar{\sigma}_t^2)}} > 0$$

This derivative is strictly positive because $n\bar{\sigma}_t > \sigma_{it}$ will always be true if there are multiple entrepreneurs with positive amounts of wealth. So increasing ρ always makes every entrepreneur better off. The amount that an entrepreneur is proportionally better off depends on how wealthy they are. Taking the derivative with regards to k_{it} reveals that the proportional benefits of increasing ρ fall as entrepreneurs become more wealthy.

$$\frac{\partial E[\delta_i^{KH}]}{\partial\rho\,\partial k_{it}} = \frac{-(n\bar{\sigma}_t - \sigma_{it})e^{-\frac{a^2}{2}}\left(\frac{\partial\sigma_{it}}{\partial k_{it}}/k_{it}^2\right)}{\sqrt{2\pi(n(1-\rho)(v^2+\bar{\sigma}_t^2)+\rho n^2\bar{\sigma}_t^2)}} < 0$$

Here, the only additional assumption is that σ_{it} is increasing in k_{it}, which is much weaker restriction than required for other results in the paper.

Thus, positively correlated changes to output make all entrepreneurs better off, but the effect tends to reduce inequality because poor entrepreneurs are more affected than rich ones. Negative correlations have the opposite effect: all entrepreneurs are made worse off, but wealthy entrepreneurs are less affected, so inequality increases.

To get an intuition for the source of this pattern, consider an economy with one very rich entrepreneur and one very poor one. When the KH criterion is applied, the very rich entrepreneur is weighted much more heavily than the poor one in the evaluation process. To a first approximation, any proposal that makes the rich entrepreneur better off will be implemented, and any proposal that makes the rich entrepreneur worse off will be discarded. If the effects of regulatory change are uncorrelated, the poor agent will have an even chance of increasing or decreasing in wealth. But if effects are positively correlated, the poor entrepreneur receives, on average, similar returns to the rich entrepreneur. When the proposals are implemented because they make the rich entrepreneur better off, they end up helping the poor entrepreneur as well. If the effects are negatively correlated, the opposite occurs. The poor entrepreneur only benefits from proposals that make the rich entrepreneur worse off, and all of these proposals are rejected. And when the poor entrepreneur is worse off, the rich entrepreneur is benefits, so the proposal is implemented. In this scenario, the rich entrepreneur is basically unaffected by the correlation, but the poor entrepreneur is much better off if the effects of regulatory change are positively correlated.

Transfers and Redistribution

Many of the problems associated with repeated application of the KH criterion can be addressed by transfers of wealth. In fact, if the government has free reign to redistribute wealth as it sees fit, Pareto improvements could be easily achieved each period: as demonstrated in the Results section, when S is large, mean amount of wealth increases every period. If the government simply collected all output and then redistributed it evenly among the entrepreneurs, inequality would disappear, and every entrepreneur would increase in wealth every period. However, as I discussed in the Literature Review, there are important restrictions on the government's ability to redistribute wealth. In practice the government is likely to apply a weaker redistributive policy. Surveying the effect all possible redistributive mechanisms is beyond the scope of this paper. However, one cautionary note is worth raising.

Canada uses discrete income tax brackets to determine marginal tax rates, and there is a highest tax bracket with a maximum marginal rate (Canada Revenue Agency, 2012). The tax system overall is progressive because average tax rates increase with income. But in the highest tax bracket, progressivity declines as incomes rise. Repeated application of the KH criterion generates inequality at all levels of wealth, so inequality is likely to increase among top earners. Again, as one entrepreneur becomes very wealthy, the expected change in wealth falls for all other entrepreneurs. When the revenue generated by the tax is distributed evenly among the entrepreneurs, the expected change in wealth for poor entrepreneurs has a positive minimum value. But this minimum value is equal to the fraction of wealth each entrepreneurs gains through redistribution. For the poor entrepreneurs, gains in wealth depend entirely on transfers from the wealthy entrepreneur. In simulations, the consistent outcome is for the population to split into two classes: a single rich entrepreneur who earns high returns, and everyone else. The poor entrepreneurs converge at what could be thought of as a poverty line – the amount of wealth obtained by relying on transfers alone (fig 8).

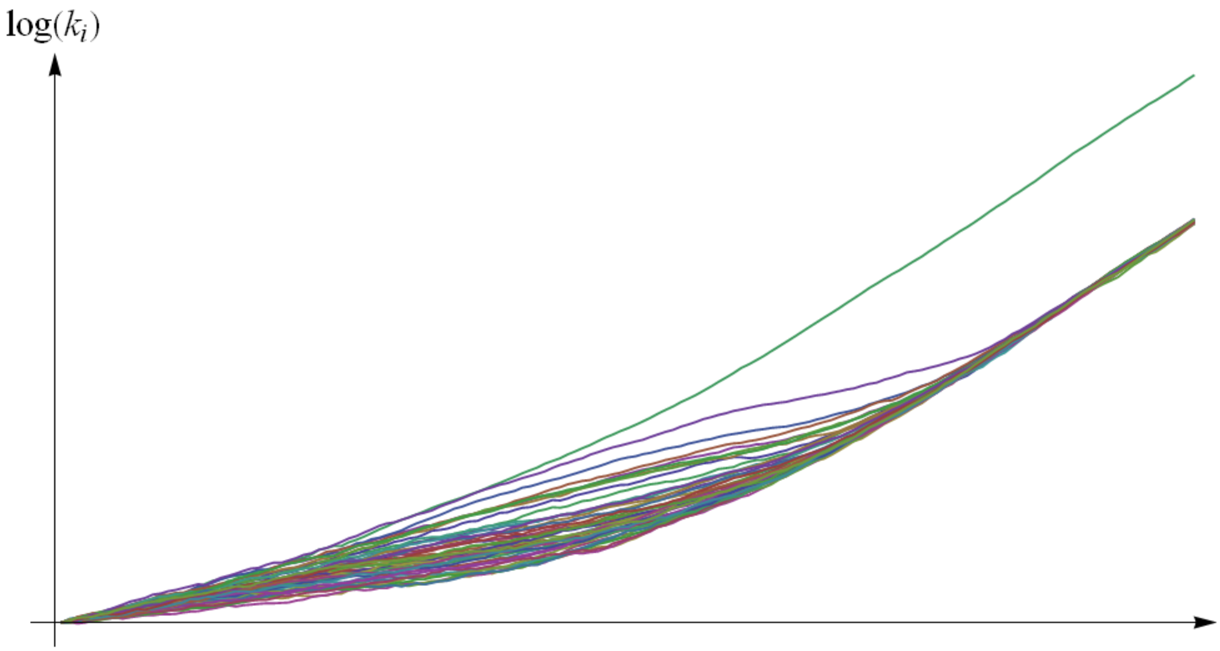

Figure 8. Simulation output with a single tax-bracket redistribution scheme

Thus, the income tax bracket system does succeed in generating Pareto improvements every period and reducing inequality. But the distribution of wealth still converges on the maximum amount of inequality that is possible given the highest marginal tax rate.

Conclusion

This paper fills a gap in the literature by providing a simple model of what happens when the KH criterion is applied to many decisions over time. The model shows that KH criterion can generate Pareto improvements if applied to many simultaneous decisions. But, the likelihood of a Pareto Improvement occurring depends on the distribution of wealth. Applying the KH criterion to many decisions is shown to award proportionally higher expected returns to wealthy individuals relative to poor ones. This causes inequality to rise when proposals are evaluated in sequence. Given enough time, one lucky individual ends up with the vast majority of the wealth. When inequality is extremely high, poor individuals see no gains on average and Pareto improvements become unlikely. These results stand in stark contrast to the Hotelling and Hicks' hypothesis that repeated application of the KH criterion will make everyone better off in the long run. The results are also significantly different than the Fargione, Lehman and Polasky model. In their model, inequality rises, but rich entrepreneurs are still subject to random returns and may not stay rich for long. In my model, whoever becomes rich is certain to stay rich, and whoever becomes poor is certain to stay poor.

My results are driven by three key components: 1) a stochastic source of changes to wealth, 2) an increasing relationship between the absolute magnitude of these changes (what I have termed ‘exposure’) and a given individual’s current wealth and, 3) the KH criterion as a filter on these changes, so that changes with net-negative benefits are prevented from occurring. These three components will not exist in all real-life situations. But where they do, the amount of inequality generated may be very large, undermining any hope of long-term Pareto Improvements. The question then, is whether these three components exist in the real economy in sufficient quantity to affect the distribution of wealth. This is an empirical question, and a very interesting one, but it will have to wait for further research. I have attempted to show that it is at least very plausible that repeated application of the KH criterion has a large influence on the distribution of wealth. At the very least, I hope this paper will shake the confidence of those who follow Hotelling and Hicks’ musings so unquestioningly.

On a practical note, if repeated application of the Kaldor-Hicks criterion is unlikely to generate Pareto Improvements, this paper presents a strong argument for policies which progressively redistribute wealth. In some sense, by applying the KH criterion, the government is responsible for the increase in inequality created. It seems reasonable that the government should also be responsible for dealing with it. Unfortunately, a satisfying solution to this inequality may require changes to the current tax system. Simulations suggest that the existing tax system in Canada may generate in a small number of very wealthy individuals, and a majority dependant on transfers.

It may also be worth extending this analysis to other decision criteria. This paper only considers the KH criterion, so it is difficult to be sure if the results are unique to the KH criterion, or are a result of the modeling approach. The basic framework of this analysis, where the KH criterion is modeled as the truncation of a statistical distribution, could easily be applied to other decision processes. Preliminary research suggests that this approach makes it easy to demonstrate why the Pareto Criterion fails for large populations (the expected change in wealth for each individual falls at a rate of $1/2^n$), and that a majority vote rule results in lower inequality but also lower mean wealth. I suspect that if the government's ability to redistribute wealth is sufficiently unrestricted, the KH criterion will dominates all other rules for any given level of inequality. But, this is an unproven hypothesis and there may be important exceptions that will be revealed by more thorough research.

Acknowledgements

I am indebted to my supervisors Elisabeth Gugl and Merwan Engineer who provided huge amounts of assistance and feedback throughout the process of developing and writing this thesis. I also benefited from regular guidance from directors of the Honours Program, Daniel Rondeau and Herbert Schuetze, and the fellowship of all the 20011-2012 Economics Honours Students. The mathematical tools and techniques applied would have been beyond my grasp without help from Louis Philippe Saumier and Florin Diacu. Finally, I owe a debt of gratitude to Amelja Zoehner for her tireless editing and emotional support. All remaining errors in grammar, math and economic principle are my own.

Appendix 1: Derivation of $E[\boldsymbol{\delta}_i^{KH}]$

Overview

$E[\boldsymbol{\delta}^{KH}]$ is the expected value of the vector of proposal effects for each entrepreneur after the KH criterion has been applied. The KH criterion implements proposals for which $\sum_{i=1}^{n} \delta_i > 0$, and discards all others. When a proposal is discarded, $\boldsymbol{\delta}^{KH} = \mathbf{0}$. When a proposal is implemented, $\boldsymbol{\delta}^{KH}$ is pulled from a truncated probability distribution with $E[\boldsymbol{\delta}| \sum_{i=1}^{n} \delta_i > 0]$. Thus,

$$E[\boldsymbol{\delta}^{KH}] = P\left\{\sum_{i=1}^{n} \delta_i > 0\right\} \cdot E\left[\boldsymbol{\delta}| \sum_{i=1}^{n} \delta_i > 0\right]$$

I assume $\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{M})$, where $\boldsymbol{M}$ is a positive-definite covariance matrix.

The strategy for deriving $E[\boldsymbol{\delta}^{KH}]$ is to transform the distribution into a standard normal distribution, applying the same transformation to the hyperplane described by $\sum_{i=1}^{n} \delta_i = 0$. Once the distribution is standard normal, there exists a vector of independent random variables (also standard normally distributed) such that one dimension of the vector is orthogonal to the transformed hyperplane. The expected value can be solved as a 1-dimensional problem along this vector. Then it is simply a matter of inverting all the transformations to find $E[\boldsymbol{\delta}^{KH}]$.

Lemmas

Two lemmas are used:

Lemma 1: Suppose $\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{M})$. Let $\boldsymbol{AA}' = \boldsymbol{M}$ by Cholesky decomposition. Then $\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu})$ is standard normal $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

Proof: Let $\boldsymbol{\delta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{M})$. By the affine transformation property, $\boldsymbol{A}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu}) = \boldsymbol{A}^{-1}\boldsymbol{\delta} - \boldsymbol{A}^{-1}\boldsymbol{\mu}$ is normally distributed with expected value $\boldsymbol{A}^{-1}\boldsymbol{\mu} + \boldsymbol{A}^{-1}(-\boldsymbol{\mu}) = \boldsymbol{0}$ and variance $\boldsymbol{A}^{-1}\boldsymbol{M}(\boldsymbol{A}^{-1})' = \boldsymbol{A}^{-1}\boldsymbol{AA}'(\boldsymbol{A}^{-1})' = \boldsymbol{I}$.

Lemma 2: Suppose $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. Consider $E[\boldsymbol{X_1}|x_1 > a]$, where X_1 is the first element in the vector of random variables that make up $\boldsymbol{X}$. Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the PDF and CDF functions of the standard normal distribution. Then,

$$E[\boldsymbol{X_1}|x_1 > a] = \int_a^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1 f(x_1, x_2, x_3, \dots, x_n) dx_n \dots dx_3 dx_2 dx_1$$

$$= \int_a^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1 \phi(x_1)\phi(x_2)\phi(x_3) \dots \phi(x_n) dx_n \dots dx_3 dx_2 dx_1$$

$$= \int_a^{\infty} x_1\phi(x_1) \int_{-\infty}^{\infty}\phi(x_2)dx_2 \int_{-\infty}^{\infty}\phi(x_3)dx_3 \dots \int_{-\infty}^{\infty}\phi(x_n)dx_n = \int_a^{\infty} x_1\phi(x_1)$$

Therefore, $E[\boldsymbol{X_1}|x_1 > a]$ is the expected value of a univariate truncated normal distribution. The formula for this expected value can be expressed as follows (Forbes et. al, 2011, pp. 147):

$$E[\boldsymbol{X_1}|x_1 > a] = \frac{\phi(a)}{1 - \Phi(a)} = \sqrt{2/\pi}\,\frac{e^{-\frac{a^2}{2}}}{erfc\left(a/\sqrt{2}\right)}$$

And,

$$P\{x_1 > a\} \cdot E[X_1 | x_1 > a] = \Phi(-a)\left(\frac{\phi(a)}{1-\Phi(a)}\right) = \phi(a) = \frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}}$$

Derivation

Let $\boldsymbol{Z} = \boldsymbol{A}^{-1}(\boldsymbol{\delta} - \boldsymbol{\mu})$. Then $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ (by lemma 1) and $\boldsymbol{\delta} = \boldsymbol{AZ} + \boldsymbol{\mu}$.

The equation $\sum_{i=1}^{n} \delta_i = 0$ describes a hyperplane which is the boundary of the truncation created by the KH criterion: all proposal values above this hyperplane are accepted, and everything below is rejected. Note that $\sum_{i=1}^{n} \delta_i = 0$ can be written equivalently as $'\boldsymbol{\delta} = 0$, where $\boldsymbol{\iota}$ is an $(n \times 1)$ vector of ones. Then,

$$0 = \boldsymbol{\iota}'\boldsymbol{\delta} = \boldsymbol{\iota}'(\boldsymbol{AZ} + \boldsymbol{\mu}) = \boldsymbol{\iota}'\boldsymbol{AA}^{-1}(\boldsymbol{AZ} + \boldsymbol{\mu}) = \boldsymbol{\iota}'\boldsymbol{A}(\boldsymbol{Z} + \boldsymbol{A}^{-1}\boldsymbol{\mu})$$

This is an equation for a hyperplane in $\boldsymbol{Z}$ with normal vector $\boldsymbol{A}'\boldsymbol{\iota}$ and fixed point $\boldsymbol{r} = \boldsymbol{A}^{-1}\boldsymbol{\mu}$. By the fact that a standard normal distribution is invariant under rotation, the distribution can be rotated such that such that one dimension of the vector is orthogonal (normal) to the hyperplane without changing the fact that the distribution is standard normal. Let $\boldsymbol{X}$ be the corresponding random vector in this rotated basis. Only the first component of this vector (x_1) must be truncated.

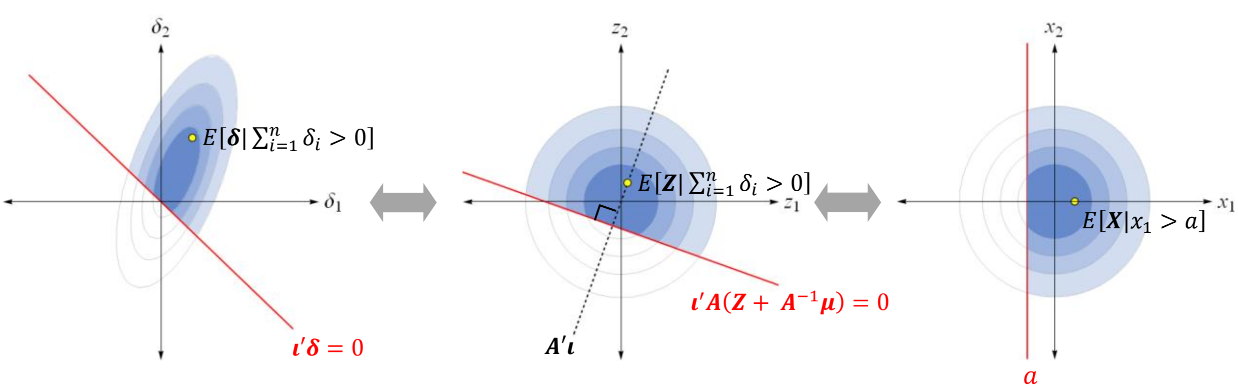

Figure 9. Transforming $E[\boldsymbol{\delta}| \sum_{i=1}^{n} \delta_i > 0]$ into a 1-dimensional problem

By lemma 2,

$$E[\boldsymbol{X_1}|x_1 > a] = \frac{\phi(a)}{1-\Phi(a)} = \sqrt{2/\pi}\,\frac{e^{-\frac{a^2}{2}}}{\text{erfc}\left(a/\sqrt{2}\right)}$$

This equation describes the distance from the origin to the expected value of truncated Z. The direction from the origin will be normal to the truncating hyperplane. Note that the length of $\boldsymbol{A'\iota}$ is,

$$\|\boldsymbol{A'\iota}\| = \sqrt{(\boldsymbol{A'\iota})'(\boldsymbol{A'\iota})} = \sqrt{\boldsymbol{\iota' A A' \iota}} = \sqrt{\boldsymbol{\iota' M \iota}}$$

Thus,

$$E[\boldsymbol{Z}|\boldsymbol{\iota' A}(\boldsymbol{Z} + \boldsymbol{A}^{-1}\boldsymbol{\mu}) > 0] = E[\boldsymbol{X_1}|x_1 > a]\frac{\boldsymbol{A'\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}}$$

The value a can be found by finding the length of $\boldsymbol{0} - \boldsymbol{r} = -\boldsymbol{A}^{-1}\boldsymbol{\mu}$ on projected on $\boldsymbol{A'\iota}$. Thus,

$$a = \frac{(\boldsymbol{A'\iota})'(-\boldsymbol{A}^{-1}\boldsymbol{\mu})}{\|\boldsymbol{A'\iota}\|} = -\frac{\boldsymbol{\iota' A A}^{-1}\boldsymbol{\mu}}{\sqrt{\boldsymbol{\iota' M \iota}}} = -\frac{\boldsymbol{\iota'\mu}}{\sqrt{\boldsymbol{\iota' M \iota}}}$$

Now the expectation in $\boldsymbol{Z}$ can be transformed to an expectation in $\boldsymbol{\delta}$ by applying the transformation $\boldsymbol{\delta} = \boldsymbol{AZ} + \boldsymbol{\mu}$. Thus,

$$E\left[\boldsymbol{\delta}|\sum_{i=1}^{n}\delta_i > 0\right] = \boldsymbol{A}\cdot E[\boldsymbol{Z}|\boldsymbol{\iota' A}(\boldsymbol{Z} + \boldsymbol{A}^{-1}\boldsymbol{\mu}) > 0] + \boldsymbol{\mu} = E[\boldsymbol{X_1}|x_1 > a]\frac{\boldsymbol{AA'\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}} + \boldsymbol{\mu} =$$

$$= E[\boldsymbol{X_1}|x_1 > a]\frac{\boldsymbol{M\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}} + \boldsymbol{\mu}$$

Finally,

$$E[\boldsymbol{\delta}^{KH}] = P\left\{\sum_{i=1}^{n}\delta_i > 0\right\}\cdot E\left[\boldsymbol{\delta}|\sum_{i=1}^{n}\delta_i > 0\right] = P\{x_1 > a\}\cdot\left(E[\boldsymbol{X_1}|x_1 > a]\frac{\boldsymbol{M\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}} + \boldsymbol{\mu}\right)$$

$$= \phi(a)\frac{\boldsymbol{M\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}} + \left(1-\Phi(a)\right)\boldsymbol{\mu} = \frac{e^{-\frac{a^2}{2}}}{\sqrt{2\pi}}\frac{\boldsymbol{M\iota}}{\sqrt{\boldsymbol{\iota' M \iota}}} + \frac{1}{2}\text{erfc}\left[a/\sqrt{2}\right]\mu$$

Suppose that $\boldsymbol{M}$ has entries,

$$\boldsymbol{M}_{ij} = \rho\sigma_i\sigma_j \;\forall\; i \neq j$$

$$\boldsymbol{M}_{ii} = \sigma_i^2 \;\forall\; i$$

Let $\boldsymbol{\sigma^2} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2]'$, $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$, $\bar{\sigma} = \frac{1}{n}\sum_{i=1}^n \sigma_i = \frac{1}{n}\boldsymbol{\iota}'\boldsymbol{\sigma}$,

$\bar{\mu} = \frac{1}{n}\sum_{i=1}^n \mu_i = \frac{1}{n}\boldsymbol{\iota}'\boldsymbol{\mu}$ and $s^2 = \frac{1}{n}\sum_{i=1}^n(\sigma_i - \bar{\sigma})^2$. Then,

$$\boldsymbol{M} = (1-\rho)\,\boldsymbol{\Sigma} + \rho\boldsymbol{\sigma}\boldsymbol{\sigma}'$$

$$\boldsymbol{M\iota} = \big((1-\rho)\,\boldsymbol{\Sigma} + \rho\boldsymbol{\sigma}\boldsymbol{\sigma}'\big)\boldsymbol{\iota} = (1-\rho)\,\boldsymbol{\Sigma\iota} + \rho\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\iota} = (1-\rho)\boldsymbol{\sigma^2} + \rho n\bar{\sigma}\boldsymbol{\sigma}$$

$$\boldsymbol{\iota}'\boldsymbol{M\iota} = \boldsymbol{\iota}'\left((1-\rho)\boldsymbol{\sigma^2} + \rho n\bar{\sigma}\boldsymbol{\sigma}\right) = (1-\rho)\sum_{i=1}^n \sigma_i^2 + \rho n^2\bar{\sigma}^2$$

$$\sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n(\sigma_i^2 - 2\sigma_i\bar{\sigma} + \bar{\sigma}^2 + 2\sigma_i\bar{\sigma} - \bar{\sigma}^2) = \sum_{i=1}^n((\sigma_i - \bar{\sigma})^2 + 2\sigma_i\bar{\sigma} - \bar{\sigma}^2)$$

$$= \sum_{i=1}^n(\sigma_i - \bar{\sigma})^2 + \sum_{i=1}^n 2\sigma_i\bar{\sigma} - \sum_{i=1}^n \bar{\sigma}^2 = n\frac{1}{n}\sum_{i=1}^n(\sigma_i - \bar{\sigma})^2 + 2\bar{\sigma}n\frac{1}{n}\sum_{i=1}^n \sigma_i - n\bar{\sigma}^2$$

$$= nv^2 + 2n\sigma^2 - n\bar{\sigma}^2 = nv^2 + n\bar{\sigma}^2 = n(v^2 + \bar{\sigma}^2)$$

$$\Rightarrow \boldsymbol{\iota}'\boldsymbol{M\iota} = n(1-\rho)(v^2 + \bar{\sigma}^2) + \rho n^2\bar{\sigma}^2$$

$$\boldsymbol{\iota}'\boldsymbol{\mu} = n\bar{\mu}$$

Thus,

$$E[\boldsymbol{\delta}^{KH}] = \frac{\left((1-\rho)\boldsymbol{\sigma^2} + \rho n\bar{\sigma}\boldsymbol{\sigma}\right)e^{-\frac{a^2}{2}}}{\sqrt{2\pi(n(1-\rho)(v^2 + \bar{\sigma}^2) + \rho n^2\bar{\sigma}^2)}} + \frac{1}{2}\operatorname{erfc}\big[a/\sqrt{2}\big]\boldsymbol{\mu}$$

$$a = -\frac{n\bar{\mu}}{\sqrt{n(1-\rho)(v^2 + \bar{\sigma}^2) + \rho n^2\bar{\sigma}^2}}$$

Determining expectation of a particular δ_i^{KH} requires extracting row i of $E[\boldsymbol{\delta}^{KH}]$. Denote $(\cdot)_i$ as the i^{th} row of a column vector. Then,

$$E[\delta_i^{KH}] = \frac{((1-\rho)\sigma_i^2 + \rho n\bar{\sigma}\sigma_i)e^{-\frac{a^2}{2}}}{\sqrt{2\pi(n(1-\rho)(v^2+\bar{\sigma}^2)+\rho n^2\bar{\sigma}^2)}} + \frac{1}{2}\text{erfc}\left[a/\sqrt{2}\right]\mu_i$$

For $n = 2$ this simplifies to,

$$E[\delta_i^{KH}] = \frac{\left(\sigma_i^2 + \rho\sigma_i\sigma_j\right)e^{-\frac{a^2}{2}}}{\sqrt{\pi(n(\sigma_1^2+\sigma_2^2)+2\rho\sigma_1\sigma_2)}} + \frac{1}{2}\text{erfc}\left[a/\sqrt{2}\right]\mu_i$$

$$a = -\frac{\mu_1+\mu_2}{\sigma\sqrt{(\sigma_1^2+\sigma_2^2)+2\rho\sigma_1\sigma_2}}$$

In the basic model, $\mu_i = \mu k_i$ and $\sigma_i = \sigma k_i$, while ρ is assumed to be 0. This simplifies the equation to:

$$E[\delta_i^{KH}] = \frac{e^{-\frac{a^2}{2}}\sigma^2 k_i^2}{\sqrt{2\pi n\left(v^2+\bar{k}^2\right)}} + \frac{1}{2}\text{erfc}\left[a/\sqrt{2}\right]\mu k_i$$

$$a = -\frac{\mu\bar{k}}{\sqrt{\frac{1}{n}\left(v^2+\bar{k}^2\right)}}$$

Appendix 2: Proof that inequality increases every period for large S

In the basic model, the fact that inequality increases every period can be easily demonstrated for large S. We can say that inequality has increased if, for all i, j,

$$k_{it} > k_{jt} \ \Rightarrow \frac{k_{i(t+1)}}{k_{j(t+1)}} > \frac{k_{it}}{k_{jt}}$$

Which can be simplified to,

$$k_{it} > k_{jt} \ \Rightarrow \frac{\Delta_{it}}{\Delta_{jt}} > \frac{k_{it}}{k_{jt}}$$

Taking the probability limit gives,

$$\lim_{S\to\infty} \mathrm{E}\left[\frac{\Delta_{it}}{\Delta_{jt}}\right] = \lim_{S\to\infty} \frac{E[\Delta_{it}]}{E[\Delta_{it}]} = \frac{E[\delta_{it}^{KH}]}{E\left[\delta_{jt}^{KH}\right]} > \frac{k_{it}}{k_{jt}}$$

(See Vogelvang (2005, pp. 66) for a brief introduction to probability limits). This last inequality is equivalent to,

$$k_{it} > k_{jt} \ \Rightarrow \frac{e^{-\frac{a^2}{2}}\sigma^2}{e\sqrt{2\pi n(v^2+\bar{k}^2)}} k_{it} + \frac{1}{2}\mathrm{erfc}\left[a/\sqrt{2}\right]\mu > \frac{\sqrt{2/\pi}\ e^{-\frac{a^2}{2}}\sigma^2}{\sqrt{2\pi n(v^2+\bar{k}^2)}} k_{jt} + \frac{1}{2}\mathrm{erfc}\left[a/\sqrt{2}\right]\mu$$

Which is always true.

References

Banerjee, A. V. & Newman, A. F. (1991). Risk-bearing and the theory of income distribution. *The Review of Economic Studies, 58*(2), 211-235. Retrieved from http://www.jstor.org/stable/2297965

Boardmen, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D, L. (2011). *Cost-benefit analysis concepts and practice* (4th ed.). Upper Saddle River, NJ: Pearson Education Limited.

Canada Revenue Agency. (2012). *What are the income tax rates in Canada for 2012?* Retrieved from http://www.cra-arc.gc.ca/tx/ndvdls/fq/txrts-eng.html

Champernowne, D. G. (1953). A model of income distribution. *The Economic Journal*, *63*(250), 318-351. Retrieved from http://www.jstor.org/stable/2227127

Eichhorn, W., Funke, H. & Richter, W. F. (1984). Tax Progression and Inequality of Income Distribution. *Journal of Mathematical Economics*, *13*, 127-3.

Fargione J.E., Lehman, C., & Polasky, S. (2011). Entrepreneurs, chance, and the deterministic concentration of wealth. *PLoS ONE, 6*(7).

Forbes, C., Evans, M., Hasting, N., & Peacock, B. (2011). *Statistical distributions* (4th ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Frank, R. (2000). Why is cost-benefit analysis so controversial? *The Journal of Legal Studies*. *29*(S2), 913-930. Retrieved from http://www.jstor.org/stable/10.1086/468099

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance, 12*, 433-463.

Hicks, J. R. (1941). The rehabilitation of consumers' surplus. *The Review of Economic Studies*, *8*(2), 108-116. Retrieved from http://www.jstor.org/stable/2967467

Hotelling, H. (1938). The general welfare in relation to problems of taxation and of railway and utility rates. *Econometrica*, *6*(3), 242-269. Retrieved from http://www.jstor.org/stable/1907054

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *The Economic Journal*, *49*(195), 549-552. Retrieved from http://www.jstor.org/stable/2224835

Kennedy, J. F. (1963). Remarks in Heber Springs, Arkansas, at the dedication of Greers Ferry Dam. *The American Presidency Project*. Retrieved from http://www.presidency.ucsb.edu/ws/?pid=9455.

Moulin, H., & Thomson, W. (1988). Can everyone benefit from growth?: Two difficulties. *Journal of Mathematical Economics*. *17*(4), 339-345.

Piketty,T. & Saez, E. (2012). Income Inequality in the United States, 1913-1998. *Quarterly Journal of Economics*, 118(1), 1-39. Retrieved from http://elsa.berkeley.edu/~saez/pikettyqje.pdf

Polinsky, M. (1972). Probabilistic compensation criteria. *The Quarterly Journal of Economics*, 86(3), 407-425. Retrieved from http://www.jstor.org/stable/1880801

Serizawa, S. (2002). Inefficiency of strategy-proof rules for pure exchange economies. *Journal of Economic Theory*. *106*(2), 219-241.

Treasury Board of Canada Secretariat. (2012). Canadian cost-benefit analysis guide; regulatory proposals. Retrieved from http://www.tbs-sct.gc.ca/ri-qr/documents/gl-ld/analys/analys00-eng.asp

Vilfredo Pareto. (2012). In *Encyclopædia Britannica*. Retrieved from http://www.britannica.com/EBchecked/topic/443519/Vilfredo-Pareto

Vogelvang, B. (2005). *Econometrics theory and applications with Eviews*. Harlow, Essex: Pearson Education Limited.