University of Southern Denmark, Odense
Faculty of Business and Social Sciences

02/06/2020
Keystrokes: 104.505 (without spaces)

Master's thesis

An efficiency–equality tradeoff? Exploring the relationship between productivity and inequality

This thesis is prepared by:

Nicholas Martin Ford (Nick Ford)
DOB: 25/11/1984, cand.oecon, 4th semester

Supervisor:

Lars Lønstrup
Department of Business and Economics, SDU

Abstract

How does productivity growth affect income inequality? Can income inequality weaken the long-term potential for productivity growth? These twin questions are the focus of this theoretical exploration of the relationship between productivity and inequality.

This thesis presents an endogenous growth model in an overlapping generations framework where individuals choose their level of education; a choice that partly depends on the productivity growth rate. As the average level of education across the labour force increases, productivity growth increases, and income inequality declines. The results of this model are discussed with reference to observed trends in the developed world and relevant research on productivity and inequality.

I hereby solemnly declare that I have personally and independently prepared this paper. All quotations in the text have been marked as such, and the paper or considerable parts of it have not previously been subject to any examination or assessment.

Nick Ford

2 June 2020

Contents

Introduction ... 5

1 Setting the scene ... 7

1.1 What is productivity? ... 7

1.1.1 Why is productivity relevant? ... 8

1.1.2 Trends in productivity ... 10

1.2 What is inequality? ... 14

1.2.1 Why is income inequality relevant? ... 16

1.2.2 Trends in inequality ... 17

1.3 Why might productivity and inequality be linked? ... 19

2 The theoretical context ... 22

2.1 Origins of almost everything ... 22

2.2 An efficiency–equality tradeoff? ... 24

2.3 Skill-biased technical change ... 28

2.4 The human capital – productivity nexus ... 32

3 The model ... 38

3.1 Functional outline ... 38

3.2 Application ... 42

3.2.1 The individual's human capital function ... 42

3.2.2 Aggregate human capital: two ability types ... 44

3.2.3 Efficiency growth ... 45

3.2.4 Returns to labour ... 47

3.3 Extending the model 47

3.3.1 Multiple ability types 47

3.3.2 Establishing the direction of bias in technical change 48

3.3.3 The production function 49

4 Discussion 51

4.1 Inclusive growth 51

5 Conclusion 54

References 56

An efficiency–equality tradeoff?
Exploring the relationship between productivity and inequality

‘Productivity isn’t everything. But in the long run, it is almost everything.’ That fundamental observation by Krugman (1994) distils in a few words a key conclusion of workhorse models of economic growth theory. Growth in output per capita over the long term depends not on increasing inputs, but on efficiency improvements in relation to how those inputs are converted into final goods and services. In that context, improving productivity growth is key to sustaining higher average living standards.

The average in living standards — dividing output by population — masks considerable variation. Productivity improvements, while raising output per capita, might not translate to higher living standards for every member of society. If the gains from productivity growth are distributed unevenly — and persistently so — then inequality will also rise.

Is this fundamentally a problem? Yes, if inequality has the effect of dampening long-term economic growth. International economic policy institutions, including the Organisation for Economic Co-operation and Development (OECD), International Monetary Fund (IMF) and the World Bank, have advanced the concept of ‘inclusive growth’. The fundamental idea is that for economic growth to be sustained over the long term, the gains must be broadly distributed across the population. The alternative, whereby economic growth translates to rising living standards for only a segment of the population, carries longer-term risks — whether due to rising political instability, or the economic and social consequences of inhibiting individuals from fully realising their potential.

The question is whether these two strands of thought — ‘productivity is almost everything’ and ‘inclusive growth’ — can be reconciled. This raises two related questions: first, can productivity growth drive long-term economic growth without contributing to long-term inequality? And second, can constraining or reducing inequality support long-term economic growth without hampering productivity growth?

This thesis considers the interaction between productivity and inequality. The approach is a theoretical one, building on the well-established literature associated within both fields. It considers various models that attempt to explain the observed trends among developed countries with respect to both productivity growth and income inequality.

The structure of this thesis is as follows:

- Section 1 introduces the key themes of this thesis, defining both productivity and inequality and discussing their relevance. Key trends are also presented, providing useful context as to the research question being explored. The section concludes with a discussion of 'human capital' — the accumulation of abilities and skills that qualitatively differentiate individuals — as the essential link between productivity and inequality.
- Section 2 delves into the theoretical underpinnings of productivity and inequality and how they are related. It starts with a discussion of the central model of long-term economic growth, the Solow model; discussing the economic philosophy of Okun's 'big tradeoff' between efficiency and equality and how it can be applied in a theoretical setting; and presents contemporary research on how productivity improvements may tend to be biased towards high-skilled labour.
- Section 3 builds on the theoretical background, with a model of how education choice contributes to productivity and influences inequality over time. Specifically, I take Galor and Moav (2000) as a starting point, augmenting their model with an individual choice about how much education one pursues — a choice that serves to maximise labour income given an individual's characteristics and the prevailing productivity growth rate.
- Sections 4 and 5 provide a concluding discussion on the key lessons that can be taken from this theoretical presentation, and comments on the implications in terms of inclusive growth.

I am grateful for the motivation and intellectual support provided by my supervisor, Lars Lønstrup. Our conversations helped to clarify my thinking on any number of issues during the inception and delivery of this thesis. I would also like to acknowledge the students I have had the pleasure of working with as an instructor in the subjects Methods for Dynamic Economics and Macroeconomic Analysis. Their perceptive questions routinely challenged my own thinking, and encouraged me to think about models and assumptions from alternative perspectives. *Tusind tak*.

Any mistakes are entirely my own.

1 Setting the scene

Both productivity and inequality are broad concepts that can be viewed through multiple lenses. How one understands the concepts inevitably influences what conclusions one draws about them. This section describes how productivity and inequality are defined for the purposes of this thesis, and outlines relevant trends that lay a foundation for the theoretical discussion and analysis that follows in subsequent sections.

In the discussion of trends and data in this section, I consider only developed economies. This is partly due to data availability, but in larger part to focus on outcomes across broadly comparable countries.

1.1 What is productivity?

Productivity is, at the most fundamental level, a question of how inputs are transformed into outputs. If a factory with a fixed amount of raw materials, machinery and workers produces 100 widgets in one week, and then 120 widgets in another week, we observe some change in output that is not explained by any change in the level of inputs. It could be that workers become better at completing tasks the more they perform those tasks. It could be that the existing machinery can successfully be used in a different way. It could be that the factory's manager applies some new insight about how to assign and coordinate tasks.

However, the observation that the same quantity of inputs can be used to produce a higher quantity of output does not necessarily imply an increase in productivity. What if the 120 widgets have been rushed through the production process and are now more prone to failure? What if the workers' improved techniques or the manager's new ideas about managing the factory are the consequence of a training programme? It is not sufficient to view productivity purely as proportional changes in quantity; one must also control for changes in quality.

Even in this single factory example, the challenge should be apparent. Productivity is not something we observe freely roaming in the wild. We measure it only as a residual of the things we can observe and account for (Solow 1957). Though one can appreciate at a conceptual level why productivity matters, our ability to account for it is remarkably limited. As Abramovitz puts it:

> Since we know little about the causes of productivity increase, the indicated importance of this element may be taken to be some sort of measure of our ignorance about the causes of economic growth … and some sort of indication of where we need to concentrate our attention. (Abramovitz 1956, 11)

This 'measure of our ignorance' is complicated by a need to disaggregate what is a change in productivity from what is 'merely' a change in the composition or quality of an input or output. If data do not accurately pick up improvements in the quality of inputs (or outputs), then productivity growth will be overstated (understated).

For the purposes of this thesis, the term ‘productivity’ should be considered synonymous with the concept of total factor productivity (TFP).[1] That is, the change in an economy’s output that cannot be attributed to changes in input factors: a ‘something else’ factor (Easterly and Levine 2016). A related concept is ‘technical efficiency’, which I will used to refer to the maximum potential level of productivity at a point in time: the smartest possible way to use a given set of production factors in an economy to maximise output. The fine distinction here is between the realised level and the potential level — a given economy might experience a rise in measured TFP simply through a different allocation of resources, such that actual productivity moves closer to its maximum potential. Much of the theoretical discussion later in this thesis will focus on technical efficiency (and changes in technical efficiency: efficiency growth). But in those theoretical settings the realised and the potential will be one and the same: that is, productivity and technical efficiency will be equal.

A final clarification: much of the theoretical literature on economic growth refers to ‘technology’ as a variable that governs how inputs are used. To reduce the breadth of terminology, I have elected to refer to this as technical efficiency. There is an inherent tradeoff here. While retaining the term technology would maximise consistency with the bulk of the work in this field, it is (in my view at least) intuitively harder to draw the link from ‘technology’ to productivity. Yet in a practical sense, this is precisely where the concept of productivity — and specifically how we measure productivity growth — comes from. While technical efficiency is a less compact term than technology, I regard the former as more precise than the latter.

1.1.1 Why is productivity relevant?

Productivity is ‘almost everything’ — critical to understanding long-term economic growth — but at the same time, not something that can be directly observed. In essence, it is the economic equivalent of ‘dark matter’ in physics (Grubb, Hourcade, and Neuhoff 2014). And as with dark matter, there is considerable debate as to the composition of productivity — what matters most in raising productivity growth.

As defined, productivity is shorthand for how well we use inputs in producing output. Instrumental in this is the stock of knowledge: the sum of all the ideas that humanity has built up over millennia, which inform

[1] TFP is also known as multi-factor productivity (MFP), which is the preferred term of (among others) the OECD. The OECD uses MFP ‘to signal a certain modesty with respect to the capacity of capturing all factors’ contribution to output growth’ (OECD 2001, 125). The Australian Productivity Commission distinguishes between TFP and MFP, using TFP where all production factors (including intermediate inputs) are included, and MFP where only standard inputs (labour and capital) are used (J. Gordon, Zhao, and Gretton 2015). As the bulk of the literature primarily uses TFP as the common terminology, I will proceed to use this term (my own modesty notwithstanding).

the ways we use inputs in production. Our world would be fundamentally different had no one ever considered the potential applications of sharp rock edges as tools and weapons, mastered the use of fire as a source of heat and light, or demonstrated the creativity to use round segments of a tree trunk as wheels. More prosaically, ideas affect productivity either by changing what we produce (for example, enabling new types of output that substitute for existing, lower-quality outputs without requiring more inputs) or by changing how we produce (reducing production costs by allowing output to be produced with fewer inputs).

The problem with this narrative is that it risks being too broad to be useful. Merely observing that new ideas are good says little about the origins of such ideas and how they can be fostered. Moreover, if productivity is simply a measure of ideas — which can easily cross borders — then why should different countries exhibit different levels of TFP (and different rates of TFP growth over time)?

Solow (2001) offers two suggestions for better understanding the mechanisms at play. First, what are the relevant microeconomic foundations in terms of incentives to invent and apply new innovations? Second, what are the broader, non-technological factors that influence TFP (and countries' productive potential)?

The incentives story can be thought of as an effort to formalise a production process for ideas and knowledge. This production process is not the same as a production process for a typical consumer good. In part because ideas and knowledge are not a typical consumer good. Specifically, there is non-rivalry in the use of ideas and knowledge — that is, once an idea is created, anyone can use it without impeding or detracting from anyone else's use. Moreover, the extent to which it is possible to exclude others from using knowledge once created varies: intellectual property rights provide some barriers, but are not universal. These factors constrain at some level the production motive: rivalry in use and excludability are essential properties of most goods, allowing businesses to generate an income. But Romer (1990) notes that market incentives are not the sole driver for invention: much basic science, for example, is developed without a commercial motive. The market incentives matter though for converting knowledge into new tools and technologies that are applied in output production. In this sense, differences in the returns from different types of productivity-enhancing innovations will be instrumental in determining what types of technical changes emerge and are applied — a theme that will be revisited in the theoretical discussion.

The non-technology story is a question of the wider framework that an economy exists in — beyond the direct choices involved in any production process. Different countries have different starting points in terms of growth and productivity, owing to differences in (among other things) natural endowments, climatic/environmental conditions and institutions (Mankiw, Weil, and Romer 1992). Endowments and climate are not factors that policy makers or economic agents more widely can choose to change. A mining company cannot simply invent new resources to extract from their land. A country cannot (by and large) pick itself up and relocate to a more favourable climate. And while cultural factors — the innate qualities and preferences of a given society — can evolve over the long term, seismic shifts in the short term are hard to achieve (at least in any stable way).

By contrast, institutions are (relatively speaking) more amenable to change. A coup can in short order replace a dictator with a democracy (or the reverse). Governments can pass laws that strengthen (or weaken) property rights. And of course, governments make choices about economy policy, and thus quite directly can influence productivity.

It is hard to distil the panoply of non-technological factors into a precise expression of what matters most for productivity. Islam (2008) makes a worthwhile contribution with a large empirical examination, testing the effect on productivity of a wide suite of factors (that is, defining productivity as the outcome of interest, and seeing what contribution other variables make to it). The factors are defined across four categories: economic, political/institutional, social/cultural and geographic/environmental. There are two simple messages from his study: one, there is no single factor that drives productivity growth; two, not all factors that drive productivity growth are consequences of individuals' choices or public policy prescriptions. But among the things that plausibly can be influenced, and which clearly make a difference to productivity, is education investment. This point will be explored below.

1.1.2 Trends in productivity

It is hardly a revelation to note that global living standards have increased over time. Nor, following the preceding discussion, is it surprising that productivity has also increased. Figure 1 illustrates the magnitude of this growth among developed countries using data available from the Long-Term Productivity database, a project started by researchers at France's central bank (Banque de France) (Bergeaud, Cette, and Lecat 2016). The current version of the database provides a compilation of core macroeconomic data from 23 OECD countries,[2] including long-term estimates of TFP and output (Gross Domestic Product, GDP) stretching back to the end of the 19th century.

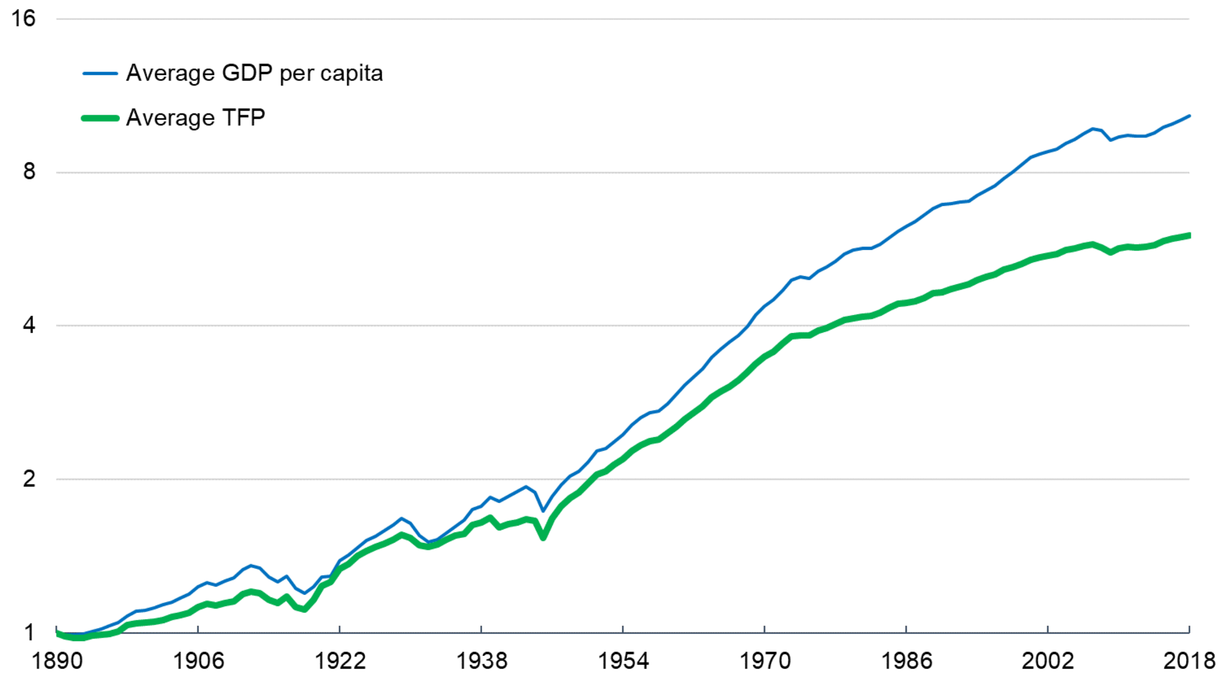
A cursory glance of the chart highlights the effect of the geopolitical and economic volatility of the first half of the 20th century — most notably, two world wars and the Great Depression. The subsequent period is considerably more stable, with striking growth (both in TFP and GDP per capita) through the 1950s and 1960s followed by a relatively more subdued growth period through to the present day (briefly interrupted

[2] The data used here excludes two countries: Chile and Mexico. While OECD members, the countries are not in other settings classed as 'developed' or 'advanced' economies — see, for example, IMF (2019).

by the 2007–08 Global Financial Crisis and its aftermath). Average annual TFP growth across the selected countries during the two full decades after World War II (1950–1969) was 3.2 per cent, while from 1970 to 2018, TFP grew on average by 1.2 per cent per year. The average annual TFP growth since 2000 has been lower still: an anaemic 0.7 per cent. The equivalent figures in terms of growth in GDP per capita are 4 per cent (1950–69), 1.9 per cent (1970–2018) and 1.1 per cent (2000–2018).

The gap (in both levels and growth rates) between TFP and GDP per capita is worth commenting on. The same pattern — though with different magnitudes — is evident when looking at subsets of the data. For example, the countries in the sample that today are members of the Eurozone recorded (on average) a nearly sixfold increase in GDP per capita (5.98) between 1950 and 2018, while TFP rose by a factor of 3.94. In the United States, GDP per capita was 3.76 times greater in 2018 than in 1950, while TFP was 2.43 times greater.

Figure 1 Average GDP per capita and TFP: selected OECD countries, 1890–2018
Logarithmic scale (y-axis), values normalised to 1 in 1890



Notes: Countries included in the average measures are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States. The original dataset also includes Chile and Mexico, which have not been included. All measurements standardised to US dollars (2010) on a purchasing power parity basis.

Data source: Long-Term Productivity database (Bergeaud, Cette, and Lecat 2016)

The gap between TFP and GDP does not disprove the assertion that productivity is 'almost everything' when it comes to long-term economic growth: short-term effects exists as well. In the short term, for example, output can be increased by increasing inputs. But this effect can only be sustained by continually sourcing new inputs. And these are scarce. This is why improving how inputs are used (lifting productivity) matters in the long term.

Figure 2 shows the average annual change in TFP in each decade from 1950 for selected countries from the Long-Term Productivity database. The aggregation of data by decade provides a better long-term overview of productivity growth, stripping out the natural short-term volatility in productivity data associated with peaks and troughs in economy activity.[3] Notwithstanding this attempt at smoothing, there remains considerable noise evident in the chart.

One trend that is apparent is that almost all countries in the sample recorded their highest productivity growth in the first two decades of the period included — consistent with the overarching trend discussed above. With the exception of Ireland, no country in the sample has recorded productivity growth of over 25 per cent across a decade (or 2.5 per cent on average per year) since 1980.

The apparent productivity slowdown has been a source of considerable debate, particularly given the technological revolution that the world has witnessed since the rise of personal computing, the internet and mobile technology. As Solow (1987) once remarked, 'you can see the computer age everywhere but in the productivity statistics'. This quip encapsulates what is regarded as a productivity paradox.

The question central to this debate is: have the major ICT advances of recent decades enabled fundamental leaps in output relative to inputs? Broadly speaking, the debate is characterised by two camps: those arguing 'no', that the ICT revolution does not (yet) offer a strong, persistent pro-productivity effect; and those arguing 'yes', that there are productivity gains, which the productivity data are not capturing.

There are various strands of thought on the 'no' side. One argument is that the effect of ICT advances on productivity is limited, because computers and related technologies are simply an evolution of existing technologies, which delivered earlier gains: the telegraph, the telephone and the typewriter were, relatively speaking, more radical innovations (Blinder and Quandt 1997). Moreover, without disputing the productive applications of ICT advances, it is also plausible that a large share of the technological revolution is a story of households and leisure (Gordon 2012). And arguably, just as the internet enables easier access

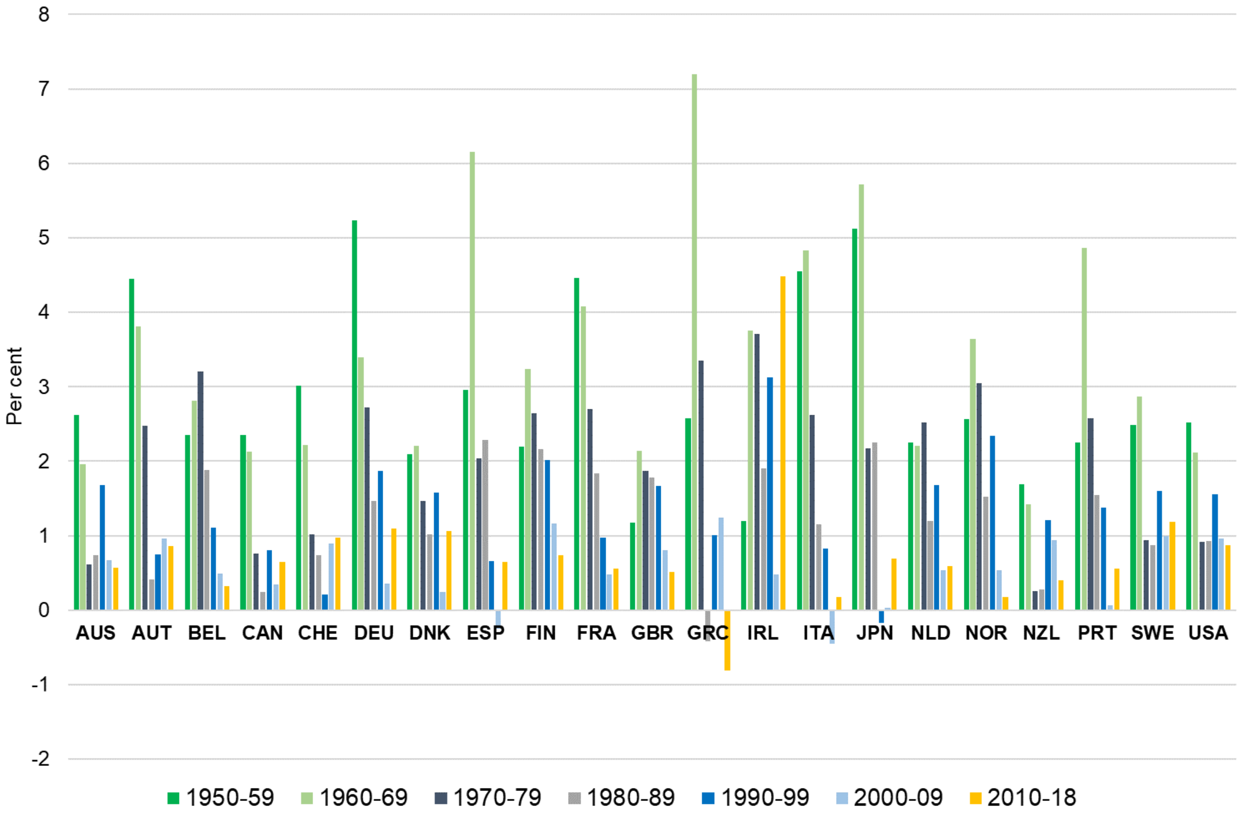
[3] For this reason, the Australian Bureau of Statistics and the Australian Productivity Commission focus on 'productivity cycles' — reporting changes in TFP across the business cycle from peak-to-peak (J. Gordon, Zhao, and Gretton 2015; Parham 2012). However, as business cycles are not consistent across countries, such a cycle-based approach would complicate comparability.

to information and reduces the cost of communication, not all of these effects will be productivity enhancing — think of time spent reading and responding to emails rather than measurable production activity, or people checking their social media feeds during work hours (or even while writing a thesis).

An alternative — somewhat more optimistic — argument calls for patience: that growth will come, eventually. In this narrative, ICT advances have the potential to deliver significant long-term gains, but that the long term (as the name suggests) can take a long time to be realised. If so, a period of lower productivity growth may simply be transitory — even if transitory in this context is a period of decades rather than years (David 2000). Gordon (2012), by contrast, argues that the substantive gains have already been realised — in particular, that US productivity growth ticked upwards in the mid-1990s as the internet became widely accessible, before petering out again by 2004.

Figure 2 Average annual TFP change by decades, 1950–2018

2010-18 standardised to ten years



Notes: AUS, Australia; AUT, Austria; BEL, Belgium; CAN, Canada; CHE, Switzerland; DEU, Germany; DNK, Denmark; ESP, Spain; FIN, Finland; FRA, France; GBR, United Kingdom; GRC, Greece; IRL, Ireland; ITA, Italy; JPN, Japan; NLD, Netherlands; NOR, Norway; NZL, New Zealand; PRT, Portugal; SWE, Sweden; USA, United States.
All measurements standardised to US dollars (2010) on a purchasing power parity basis.

Data source: Long-Term Productivity database

The other side of the debate argues that the observed productivity slowdown is overstated due to deficiencies in data measurement. Productivity data is prone to significant measurement error, in large part because it is unobservable. Indeed, one of the key practical applications of Solow's theoretical contributions is the growth accounting framework that is used to estimate productivity (Solow 1957). In brief, TFP is derived backwards from observed levels of output, taking account of observed levels of inputs.[4] Errors or quirks in the measurement of inputs and output will thus carry through to TFP estimates. Such errors hinge to an appreciable extent on accurately adjusting volumes for changes in quality or product attributes (Griliches 1979; Diewert and Fox 1999).

In this light, the 'missing' ICT effect may owe to deficiencies in productivity measurement. As Mokyr argues, aggregate measures such as productivity 'were designed for a steel-and-wheat economy, not one in which information and data are the most dynamic sector' (Mokyr 2014, 88). The ability to measure productivity associated with ICT (or other modern technologies) may simply be worse than for earlier types of technical advances. But this is hard to validate. Even accepting there are problems in aggregate measurements, Gordon (2014) counters that such problems have always existed. Even as a 'measure of our ignorance' (Abramovitz 1956, 11), productivity data may not effectively measure what we do not know.

This apparent paradox, and the debate around it, is not directly relevant to this thesis. But it is worth noting the potential deficiencies in productivity data and debate about their reliability. For the purposes of this thesis, I take the data as it is, and conclude there has been a slowdown in productivity growth. But I acknowledge there are doubts about the magnitude of such a slowdown. In any event, the observation that the level of productivity today is higher than it was 70 years ago is (to the best of my knowledge) not in dispute.

1.2 What is inequality?

Productivity is a technical concept: in defining it for a broader audience, one must lift it from its theoretical foundations into something more accessible. The reverse journey is required in defining inequality.

Inequality is a broad, multifaceted concept. Inequality can be experienced across many dimensions of life — gender, racial background and sexual orientation among other things can all contribute to differences in the rights and opportunities of individuals in society. While these social forms of inequality can certainly

[4] The TFP estimates from the Long-Term Productivity database are derived using labour and capital as the relevant inputs to production.

have economic consequences (differences in pay between men and women, for example), the link between these factors and productivity growth is hard to conceptualise.

Other forms of inequality are more directly tied to observable economic factors. Historically, the distribution of land ownership played a major role in how countries experienced economic development — and still matters today in developing countries (Carter 2000). The distribution of wealth remains a relevant metric, including in an intergenerational sense — wealthy families bequeathing from generation to generation (Piketty 2000). One might also consider inequality in consumption expenditure as a useful applied metric, especially with regard to smoothing out transitory income effects (Atkinson and Bourguignon 2000).

The focus in this thesis is income inequality. The basic reason for this is that it is easiest to conceive of a direct relationship between productivity and income inequality than other forms. Income is a flow over time and is therefore more sensitive to changes at different points in time. By comparison, wealth is a stock measure: it grows (or falls) with changes in income. The effect of productivity on wealth is most likely to be a consequence of productivity's effect on income. Specifically, as will become apparent in the later theoretical discussion, the question here is the relationship between productivity and labour, and in turn the earnings of different types of workers. And while income can be derived from non-labour sources, it is labour income that is (in the models to be presented) affected by productivity.

Even considering inequality in terms of the distribution of income raises challenges. At what point should an individual's income be measured? A focus on gross earnings before taxes and transfer payments reveals something about the differences in how individuals are valued in the market: for example, inequality between different types of workers. But one objective of government interventions through taxes and transfer payments is specifically to redistribute income and thereby ameliorate inequality. Net earnings — the income people actual have at their disposal — give a clearer indication of the actual experienced inequality across society.[5]

The data below presents measures of inequality in both market (gross) and disposable (net) income. For the purposes of the theoretical discussion about income inequality, the relevant metric is inequality in market income — more precisely, market income from labour. The models that will be discussed do not include taxes and transfers: the role of government is not explored. This is not to suggest that government

[5] Disposable income is still an imperfect measure of inequality. For example, while it takes account of direct taxes on income, it does not account for indirect taxes — for example, on consumption. And where these indirect taxes are not neutral — that is, they affect different income groups in different ways — then disposable income will not fully reflect the distributional effects in force (Atkinson and Bourguignon 2000). But as a broad indicator of inequality, it is sound.

is irrelevant; rather, that the focus is on the market dynamics at play with respect to both productivity and inequality. In practice, government interventions may well be justified based on outcomes observed in the market — but one must first be clear about what those market outcomes are.

1.2.1 Why is income inequality relevant?

The question of income distribution has not always been at the forefront of economic discourse (Atkinson and Bourguignon 2000). At some level, the issue has attracted greater political debate than economic debate. Part of the reason for this could be that, unlike other core macroeconomic measures, there is no simple objective with respect to inequality. All else being equal, economic growth is desirable: the more, the better. Similarly, unemployment is undesirable: the lower, the better. But inequality?[6]

My own priors are that less inequality is generally preferable to greater inequality. But I cannot point to an optimal level of income distribution, or a precise model that would credibly yield one.[7] Instead, I think it prudent to consider inequality in terms of its possible consequences — in particular, the ways in which it might affect long-term economic growth.

Berg and Ostry (2017) outline two channels through which inequality can trigger adverse macroeconomic effects. The first is financial: higher levels of inequality can have implications in terms of access to credit. Specifically, all else being equal, those on the lower end of the income spectrum are more likely to seek to borrow to finance present consumption. If there are credit market imperfections — for example, if financial institutions hold off from lending money due to insufficient information about a borrower's ability to repay a loan — then this will weigh on household consumption (and investment) decisions. Alternatively, financial institutions may become more exposed to those on low incomes in an environment of high inequality. This increases the risk from financial shocks — for example, debt crises (Berg and Sachs 1988).

The second channel is political. All else being equal, the greater the level of inequality, the greater the demand for redistributive policies. To the extent those redistributive policies are poorly designed, and have the effect of deterring investment (with little concomitant gain to low income earners), the long-term effect will be depressed economic potential (Alesina and Rodrik 1994). Extreme levels of income inequality may make a country more prone to political shocks and social instability (Alesina and Perotti 1996).

[6] There is a rich literature within welfare economics on questions of distribution and how they relate to individual's utility and social welfare. It is beyond my capabilities to concisely summarise (while doing justice to) the various perspectives and their philosophical and moral underpinnings in a way that is relevant to this thesis.
Well-grounded summaries of the key themes are provided by Sen (2000) and Sandmo (2015).

[7] As Ng (2004) demonstrates, the possibility of an optimal income distribution hinges on what social welfare function one assumes.

That income distribution can affect countries' long-term growth potential — indeed, that high inequality hinders stable long-term growth — is core to the notion of 'inclusive growth'. Inclusive growth has gained traction in global economic policy circles, with (among others) the OECD, the IMF, the World Bank and the World Economic Forum all arguing in its favour (OECD 2018; Kireyev and Chen 2017; World Bank 2014; WEF 2018). In broad terms, inclusive growth is the idea that stable, long-term economic growth is incompatible with rising inequality. From a policy perspective, the focus should not simply be on lifting economic growth, but on ensuring the gains from economic growth are widespread.

It is worth noting that across much of the policy literature, inclusive growth is principally considered in the context of developing countries: raising living standards in the poorest parts of the world, such that the poorest people in the world experience gains. But this does not mean inclusive growth is irrelevant for developed countries. As the OECD's inclusive growth framework makes clear, the combination of globalisation and technological innovation in developed countries — while positive and important for productivity growth — risk displacing workers in trade- and technology-exposed industries (OECD 2018).

1.2.2 Trends in inequality

The story of the post-war trend in income inequality is, at a high level, relatively simple to tell: income inequality across the developed world on average has increased since the 1960s. A deeper look at the data though reveals marked diversity across countries and over time.

Figure 3 illustrates how income inequality has changed over time across the same set of developed countries that were considered in the earlier discussion about productivity trends. The charts (and broader discussion of the underlying trends in this section) draw on the Standardised World Income Inequality Database (Solt 2019). This database reports Gini coefficients by country, both in terms of market (gross) incomes and disposable (net) incomes.

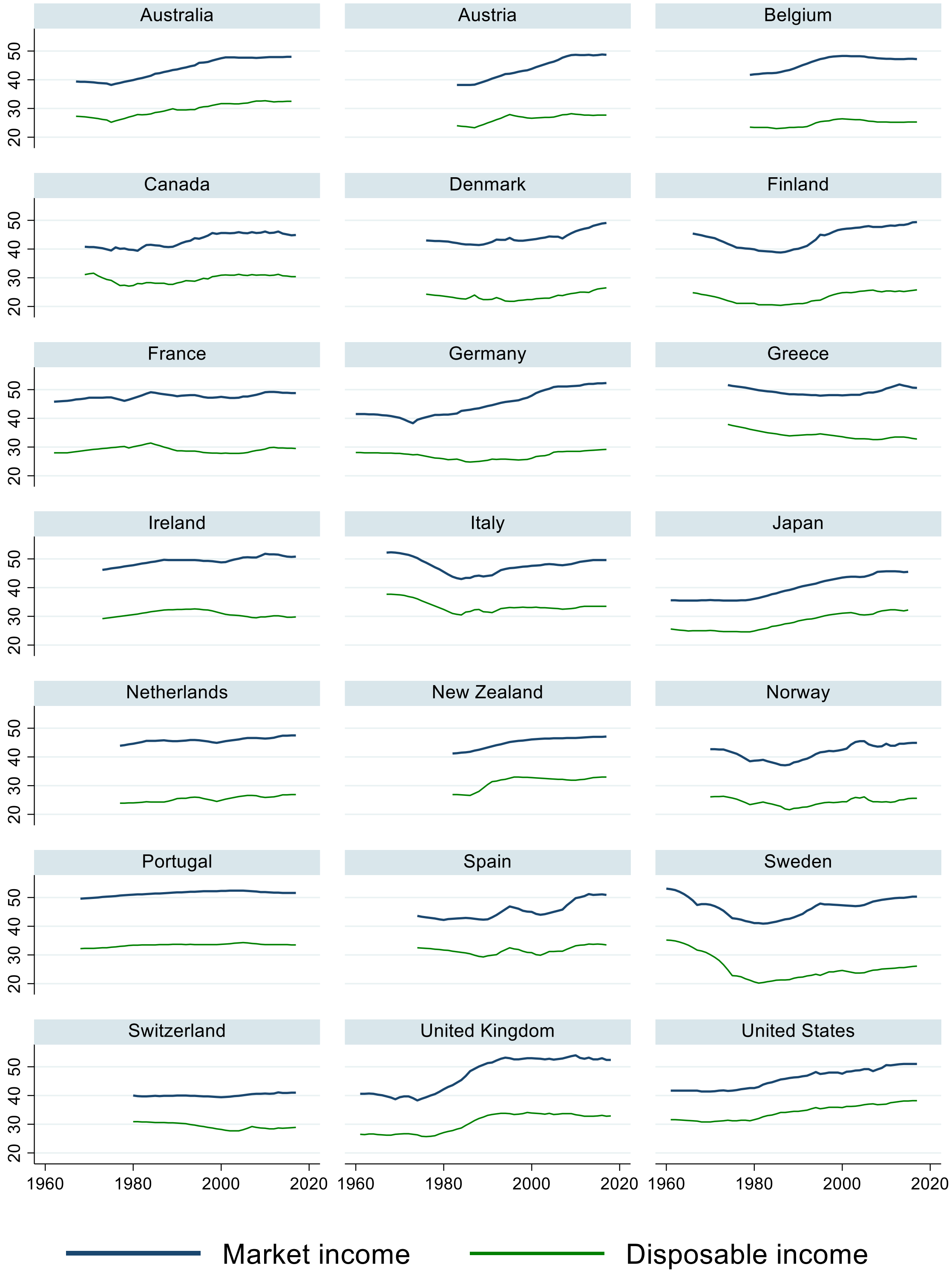
The Gini coefficient is a broad measure of inequality, capturing the full distribution of (in this case) income across a cohort.[8] The Gini coefficients here are presented as a value between 0 and 100 per cent; others choose to report Gini coefficients as decimal values between zero and one. Whichever way the values are reported, the higher is the Gini coefficient, the greater the level of inequality in the distribution.

---

[8] Specifically, the Gini coefficient is related to a device called the Lorenz curve (Gini 1921; Lorenz 1905). If one imagines a two-dimensional graph, with proportion of population along the horizontal axis and proportion of income along the vertical axis, the Lorenz curve plots the cumulative distribution of income from lowest to highest income earner across the cohort. If there is perfect equality — that is, everyone earns the same income — then the Lorenz curve will be a 45-degree line. As the curve bows in a convex fashion away from the 45-degree line, inequality is increasing. The Gini coefficient is a measure of the area between the 45-degree line and the Lorenz curve (as a proportion of the full right-angled triangle under the 45-degree line).

Figure 3 Changes in inequality: Gini coefficients by country

By market income and disposable income, selected OECD economies, 1960–2018



Notes: Data not available for all countries across the full period (1960–2018).

Data source: Standardised World Income Inequality Database (Solt 2019)

In 2015, the last year for which data are reported for the full set of countries, the disposable income Gini coefficient ranges from 25.3 (Belgium) to 38.1 (United States), while the market income Gini coefficient ranges from 40.9 (Switzerland) to 52.6 (United Kingdom). And in the latter case, Switzerland is something of an outlier — excluding that country raises the lower bound to 44.8 (Norway). The general observation is that developed countries today have comparable levels of income inequality at the market level; what makes the practical difference is distributional policies — taxes and transfers — within each country.

That said, there are clear differences in the inequality paths each country has taken to get to where they are. For example, in 1980, the market income inequality in both the United Kingdom (42.2) and the United States (42.6) was around the average for the sample (42.9).[9] Meanwhile, several continental European countries recorded Gini coefficients on market income above the average: France (47), Greece (50.1), Ireland (47.8), Italy (45.5), the Netherlands (44.6) and Portugal (51). One might infer a degree of convergence across countries over time in the distribution of market incomes — with a trend increase in market income inequality since at least the 1980s.

On average, the gap between market and disposable income inequality has grown, suggesting that redistributive policies have increased. However, increased redistribution over time has only partly offset rising market income inequality. In an analysis of OECD countries (broader than the sample of countries presented in this thesis), Immervoll and Richardson (2011) show that much of the redistributive 'heavy lifting' since the 1980s has been the result of transfer payments — that tax reforms over time have done little to assist, particularly with changes in thresholds and entitlements that have had a more regressive effect with regard to income distribution.

1.3 Why might productivity and inequality be linked?

If productivity growth is key to rising living standards over the long term, then the naïve observer might question why productivity should influence inequality (that is, the distribution of income). The simple answer is that an average increase in living standards does not mean a uniform increase in living standards.

The obvious point here is that individuals receive different levels of income. And part of the reason for that is differences in labour-related income. Different jobs attract different wages.

Just as a supermarket includes a range of products — with different variants of even broadly similar goods available at different prices, given different properties and different underlying supply costs — the labour market is also a mix of different 'products'. More specifically, that individuals supply some portion of their

[9] The sample average for 1980 excludes Austria and New Zealand, for which data are not available.

time as a service in exchange for compensation. And the service that each individual provides will vary — in part because as demand for labour differs (different employers have different requirements), but also because the capacity for supplying different labour services differs across individuals.

This capacity for supplying labour services can be approximated as ability and/or skill. At some level, one can assume these differences as innate: different individuals have different preferences, and these preferences may shape and/or be shaped by the personality attributes and natural aptitudes of the individual. Broadly speaking, these factors are exogenous — that is, we do not determine them.[10] One can build on these innate abilities through learning and training: whether through formal programmes or as a product of life experience ('on the job'). This upgrading of one's skill level is a result of conscious choices.

Taken together, the ability and skill that individuals accumulate can be considered their 'human capital'.[11] Using this distinction, labour is a simple head count — the quantity of workers available. Human capital reflects the qualitative characteristics that add value in the production process. It is in turn differences in human capital that are relevant to differences in labour earnings, and thus income inequality.

That there is a role for human capital in this thesis about productivity and inequality is supported by the data presented of developments in both since the aftermath of World War II. As Mincer (1981) reflects with reference to history, one cannot adequately explain US economic growth since the 1950s by looking only to traditional production factors (capital and labour). Nor can one explain differences in the distribution of income simply by examining the returns to capital and labour (treated as homogeneous factors) — the distributional differences had more to do with (otherwise unexplained) differences in labour income.

However, just as productivity is not something we directly observe, so too is human capital somewhat ephemeral. But we can at least approximate investment in human capital by considering people's choices with respect to education. To be clear, education is by no means a perfect measure of human capital: it is only one factor that contributes to human capital.[12] But it is also the most obvious and objective example of how individuals choose to increase their human capital with respect to their labour activities.

---

[10] This is at least true in the short term, but changes in these basic characteristics could be influenced over the long term. Today, there is not much risk of me becoming an Olympic athlete. But had I been pushed as a child to spend more time outside playing sport rather than inside playing with plastic bricks, my 'innate' abilities *might* be different today. That counterfactual is essentially unknowable. (In any event, I have no regrets.)

[11] Shortly before I completed this thesis, Kevin Hassett, a US presidential adviser, remarked in a TV interview that the 'human capital stock' was ready to work (CNN, 24 May 2020). The comment attracted criticism and derision for treating people as production assets. I would not have used the phrase in this context — I tend not to think of individuals as human capital; rather, individuals possess human capital, which they apply in the production process.

[12] Mincer summarises some of the 'major categories' of human capital investment as 'education, job training, health, information, and migration'. These investments in turn are the choice of an individual, given their 'genetic endowment, parental wealth, and access to educational and market opportunities' (Mincer 1974, 1–2).

Across developed countries, some level of education (at least in a quantity sense) can be considered broadly constant across the population: primary and secondary schooling are nowadays mostly universal (with some variation at upper levels of secondary). But on top of this baseline, individuals can choose further 'tertiary' education — for example, universities and vocational programmes. Figure 4 shows that, averaged across the sample countries considered in the preceding discussions on productivity and inequality, there has been a strong and consistent upward trend in tertiary education (Barro and Lee 2013). The proportion of the adult population that has at least attempted some tertiary education (and though not shown here, the subset of that group that has completed a tertiary education) has increased, as has the average time invested in tertiary education.

This is by no means a full story of how productivity and inequality could be linked. But it is the presumed link between productivity and inequality through differences in labour — and specifically differences in the education-based human capital attached to labour — that will be explored in this thesis.

Figure 4 Growth in tertiary education

Average across selected OECD countries, 1950–2010

Notes: Countries included in the average measures are Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom and United States. Data based on population of individuals 25 years and over.

Data source: Barro–Lee Educational Attainment Dataset (Barro and Lee 2013)

2 The theoretical context

The previous section provided an overview of the key themes of this thesis. Across the developed world, productivity has increased over time, though productivity growth rates appear to have slowed. Inequality has increased, particularly with reference to gross market income. And human capital accumulation, which could play a role in the observed trends for both productivity and inequality, has also increased significantly.

Developing these themes further, this section discusses central planks in the theoretical architecture that will accommodate the model to be presented in section 3.

2.1 Origins of almost everything

The fundamental reference point for any exposition of long-term economic growth and its causes and consequences is the work of Solow (1956; 1957) and Swan (1956). In describing productivity as 'almost everything', Krugman echoes one of the central results of the Solow (or Solow–Swan) model: that, in the long term, GDP per capita grows at the rate of technical efficiency growth. While an elementary and widely known model, I have chosen to briefly sketch out its key properties here: it is the foundation of the discussion that follows.

Starting with a production function, where output (Y) is a function of two homogenous inputs, capital (K) and labour (L). Labour's effectiveness (A) is driven by labour-augmenting 'technology' — what I will refer to as technical efficiency. The variables Y, K, L and A are themselves functions of time[13] — that is, as variables, they are free to change over time.

$$Y = F(K, AL)$$

The production function exhibits constant returns to scale: that is, a doubling of inputs will double output. This also allows the production function to be expressed in its intensive form — output per effective worker is a function of capital per effective worker:

$$\tilde{y} = f(\tilde{k}), \qquad \tilde{y} \equiv \frac{Y}{AL}; \; \tilde{k} \equiv \frac{K}{AL}$$

This intensive form function has (among other things) the properties of positive but declining marginal returns: that is, an increase in capital per effective worker will increase output per effective worker, but

[13] That is, $Y(t) = F(K(t), A(t)L(t))$. For notational ease, the customary parenthetical t notation for models in continuous time is supressed.

each additional unit of capital per effective worker contributes proportionally less in terms of new output per effective worker.

The labour force and technical efficiency are both assumed to grow at constant rates over time (respectively, n and g). Growth in the stock of capital is a function of output (output gives rise to new investments, which build up the capital stock), but capital also requires replacement over time (depreciation, the proportion of which is denoted δ). In the intensive form, the evolution of capital per effective worker over time is thus:

$$\frac{d\tilde{k}}{dt} \equiv \dot{\tilde{k}} = sf(\tilde{k}) - (n + g + \delta)\tilde{k}$$

Where the first term on the right-hand side of the equation is actual investment (the proportion of output that is saved and invested) and the second term is break-even investment (the amount of capital per effective worker needed to replace depreciated capital and offset the constant growth in population and efficiency). Absent new actual investment, capital per effective worker ($\tilde{k} \equiv K/AL$) would fall either due to the direct fall in K from depreciation, and/or the increase in A and L over time.

The consequence of this package of conditions is that capital per effective worker converges in the long run to a stable, constant level: a steady state. In this steady state, actual investment equals break-even investment, and there is no change in capital per effective worker. Thus, output per effective worker — a function of capital per effective worker — is also constant.

And so, the 'almost everything' conclusion. Because $\tilde{y} \equiv Y/AL$, then output per worker (which is the same as output per capita; the entire population is the labour force) can be expressed as $Y/L = \tilde{y}A$. In the steady state, where $\tilde{y}$ is a constant, this means that growth in output per capita — what we might loosely consider as living standards[14] — is equal to the rate of technical efficiency growth.

$$\dot{\tilde{k}} = 0 \Longrightarrow \frac{\dot{[Y/_L]}}{Y/_L} = g$$

[14] The link from output per capita to living standards is not perfect. Living standards are broader than pure economic factors. It is not hard to imagine, for example, greater output per capita combined with higher rates of pollution per capita. It is fair to suppose that people's living standards are reduced when they do not have clear air to breathe or safe water to drink. But in this simple model, we can assume away such externalities. More generally, one could say that growth in technical efficiency over time is a *precondition* for sustaining improved living standards, but not of itself a guarantee. Especially so, when — as this thesis does — considering the living standards of specific groups, rather than just the average across the population.

This key conclusion provides a useful starting point for discussing productivity — but it is only a starting point. The Solow model does not of itself provide much insight about g — the rate of efficiency growth is constant and exogenous. Nor does this model reveal anything about inequality within a country: all workers are functionally identical, and thus earn the same wage (under the assumption of perfect competition, the wage level is equal to labour's marginal product). Nevertheless, the Solow model offers a sufficiently flexible framework, which can be expanded to explore a wider range of questions.

2.2 An efficiency–equality tradeoff?

In discussing the possible relationship between technical efficiency and income inequality, one cannot escape Okun's narrative of 'the big tradeoff' between efficiency and equality (Okun and Summers 2015). This is not so much a theory as a philosophical discussion of how markets work, and the choices that governments face in regulating otherwise free markets.[15] Elementary level economics teaches the idea that free markets, absent market failures, will yield efficient outcomes. A more nuanced take is that even with market failure in a range of fields, the decentralised decision making of economics agents in a market will still yield better (more efficient) outcomes on average than a centralised decision maker determining supply and demand.

Still, the idea of a free market is infused with a sense of equality: a free market cannot exist without equality in terms of certain rights and freedoms. For example, rule of law is a fundamental institution for well-functioning markets: among other things, by ensuring that contracts are enforceable (Haggard and Tiede 2011; Alchian 1965). But the rule of law applies (or at least should apply) to all individuals equally; punishments for law-breaking are not (or should not be) differentiated on the basis of an individual's productive capacity. And while democracy is not strictly speaking a precondition for a free market, the world's developed economies are both market-based and democratic. Democracy based on universal suffrage is an important illustration of equality but, as Okun points out, such equality does not imply that voters are equally competent. In turn, the decisions that emerge through democratic processes are not necessarily (indeed, perhaps often are not) efficient. In this sense, sacrificing some efficiency is a price we pay for equality. That price may well be — and almost certainly is — socially worthwhile, but it is a price nonetheless.

This concept of equality is plainly broader than income equality — though not wholly disconnected: income inequality can be a consequence of other forms of inequality. Moreover, focusing on income inequality

[15] For the avoidance of doubt, I mean no disrespect in comparing philosophy and theory here. The philosophical foundations of economics are evident in and central to much of economic theory. Indeed, I fear that modern economics downplays its philosophical origins to its detriment.

does not lessen the principle of a tradeoff against efficiency. At some level, the tradeoff may be greater. As Okun states:

> [While] the provision of equal political and civil rights often imposes costs on society […], the attempt to enforce equality of income would entail a much larger sacrifice. In pursuing such a goal, society would forgo any opportunity to use material rewards as incentives to production. And that would lead to inefficiencies that would be harmful to the welfare of the majority. Any insistence on carving the pie into equal slices would shrink the size of the pie. That fact poses the tradeoff between economic equality and economic efficiency. (Okun and Summers 2015, 46)

But this 'fact' of a tradeoff is not without its limits. First, because (as Okun fully acknowledges) not every economic decision invokes a tradeoff between equality and efficiency. A hypothetical low-cost intervention that substantially lifted the capabilities of low-skilled workers could concurrently improve efficiency and reduce inequality. Second, because the potential for efficiency is dynamic: it changes over time. And the way efficiency changes over time is not (at least fully) predetermined. Growth in technical efficiency can also be thought of as a product of choices, which may reduce or increase (or leave unchanged) inequality.

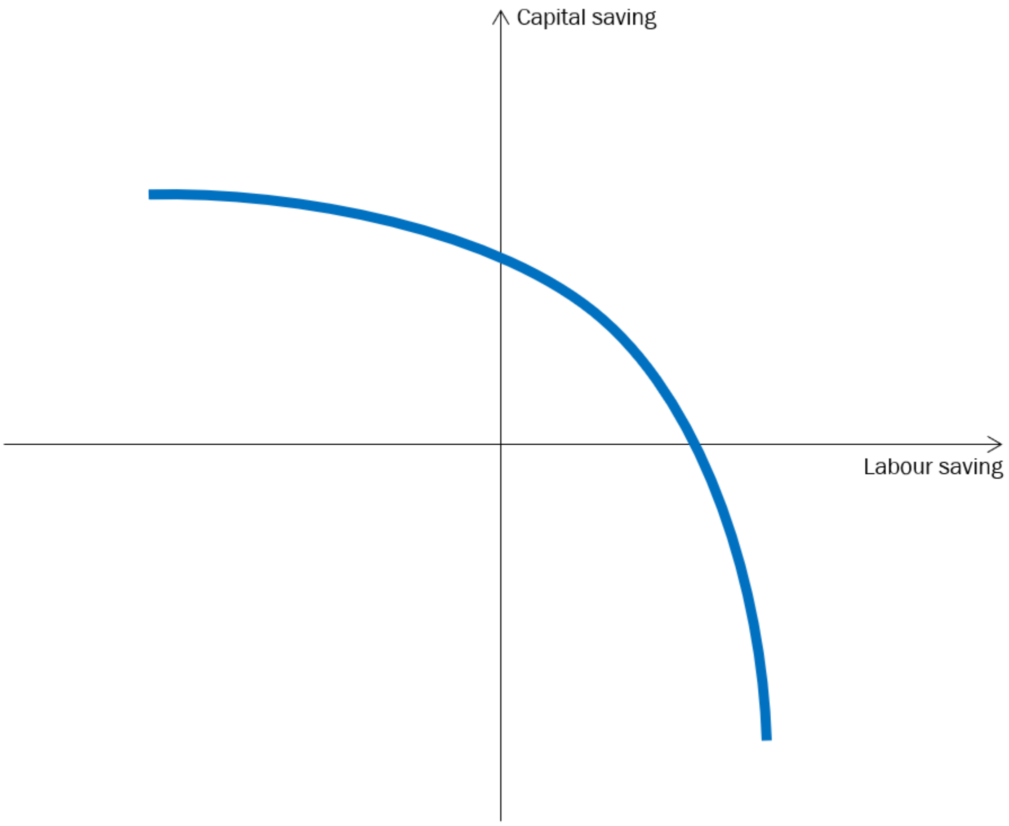
This idea of efficiency driven by choices is core to the twin concepts of induced innovation and (more recently) directed technical change. Induced innovation stems from Hicks (1963), who posits that the ratio of factor prices between capital and labour influences choices of new technical advances: for example, if labour becomes relatively more expensive, then the incentive for producers is to find ways to reduce labour costs. The fundamental rationale is a microeconomic one: the firm seeks to maximise profits, and improved technical efficiency is a means to do so. But there is no reason to assume that a given technical improvement will affect both capital and labour in equal measure. An improvement that results in an increase in the marginal product of capital relative labour would be a 'labour-saving' improvement; the opposite would be true of a 'capital-saving' improvement. Hicks (1963) suggests that there is more evidence of labour-saving than capital-saving improvements.[16] The reason for this, according to Hicks, is that some share of innovation is induced by relative factor prices; that labour-saving improvements are induced by producers' desire to reduce the use of relatively more expensive labour.

Another way of thinking about this is in terms of an innovation possibility frontier. This idea, advanced by Kennedy (1964), suggests a market for technical improvements. Output producers seek out the

[16] While Hicks offered this observation with little more than a cursory glance at the world around him, it is worth noting that this insight is not substantially different from the standard assumption of labour-augmenting technical change (that is, as in the Solow model described above, that technical efficiency enters the production function multiplicatively with labour). Indeed, this assumption — formally called Harrod neutrality — implies that technical efficiency is labour saving, as defined by Hicks (Drandakis and Phelps 2016).

improvements that deliver the greatest reduction in production costs, with respect to either or both capital and labour. Any given improvement will have different possible effects on capital and labour — it is assumed that the greater the reduction in labour-related costs from an improvement, the lower the reduction in capital-related costs. At the extreme, if substantially less labour is required in production, more capital will be required (and vice versa). This gives rise to a concave innovation possibility frontier mapping the maximum combinations of labour and capital reductions (or indeed increases). **Fejl! Henvisningskilde ikke fundet.** gives a visual overview of the frontier, where the upper-right quadrant reflects innovations that are both capital and labour saving.

Figure 5 Innovation possibility frontier



Source: Recreated from Kennedy (1964)

This is a stylised device — much like an indifference curve depicts bundles of goods that provide the same utility. And just as indifference curves reflect the utility-equivalent choices available to an individual at a point in time, so too is this innovation possibility frontier a static construct. The idea expressed here says nothing about how the frontier changes over time. Such evolution requires an assumption about the distribution of innovation payoffs.

As Binswanger (2016) notes, there is no reason to assume this distribution should be constant over time: either due to advances in basic sciences (that is, research not specifically geared to either labour- or capital-saving improvements, but which nevertheless unlock unforeseen potential for new improvements), or because the profitable opportunities for new improvements biased to one factor might be exhausted. If firms, all else being equal, prefer to choose labour-saving improvements (given the production cost share attributable to labour), one might imagine that, over time, all the low-hanging fruit of labour-saving improvements would be picked. Put another way, in the world of technical change, it might become progressively more costly to develop new labour-saving improvements relative to the stock of potential capital-saving improvements.

An alternative possibility is that, on the supply side for new improvements, past demand for labour-saving improvements induces a flow of resources into the production of labour-saving improvements. For example, in response to demand for labour-saving improvements over time, a research and development (R&D) infrastructure might be built up that is geared towards producing new labour-saving improvements. In this sense, the innovation possibility frontier would more easily shift outwards along the labour-savings axis.

This alternative scenario captures the idea of directed technical change under what Acemoglu (2002a) describes as 'state dependence'. That is, the current state of technical progress affects the future costs of different types of technical improvements. Consider an R&D process as its own form of production: just like output production, R&D draws on inputs. These inputs are scarce. Sustaining growth in R&D output over time requires that the inputs be used more productively. Over time, if the marginal productivity of R&D is not to fall, then past R&D must enable future R&D. Put more colourfully, 'spillovers imply that current researchers "stand on the shoulder of giants"' (Acemoglu 2002a, 793).

Acemoglu (2002a) formalises this specification in a model where output production is a function of two factors (what I will choose to call labour and capital here for comparability with the preceding discussion, but which Acemoglu frames as labour and an undefined Z-factor). The output of R&D directed towards the two factors (N_L and N_K) develops over time according to the following equations of motion:

$$\dot{N}_L = \eta_L N_L^{\frac{1+\delta}{2}} N_K^{\frac{1-\delta}{2}} S_L, \qquad \dot{N}_K = \eta_K N_L^{\frac{1-\delta}{2}} N_K^{\frac{1+\delta}{2}} S_K$$

Where S_L and S_K are the workforce (scientists) producing R&D directed at respectively labour and capital; η_L and η_K reflect the costs of the respective types of R&D. δ indicates the degree of state dependence: $\delta = 0$ implies no state dependence; $\delta = 1$ implies extreme state dependence.

In turn, the equilibrium condition in terms of the relative R&D output directed at the two production factors is:[17]

$$\frac{N_K}{N_L} = \left(\frac{\eta_K}{\eta_L}\right)^{\frac{\sigma}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1}{1-\delta\sigma}} \left(\frac{K}{L}\right)^{\frac{\sigma-1}{1-\delta\sigma}}$$

Where σ expresses the elasticity of substitution between factors: $\sigma < 1$ implies the factors are gross complements, $\sigma > 1$ implies the factors are gross substitutes. $0 < \gamma < 1$ reflects the relative importance of the two factors in output production: $\gamma \to 1$ implies labour is relatively more important, $\gamma \to 0$ implies capital is relatively more important.

The relative shares of each factor in terms of overall output (Ψ_L, Ψ_K) in the equilibrium state can also be expressed:

$$\frac{\Psi_K}{\Psi_L} = \left(\frac{\eta_K}{\eta_L}\right)^{\frac{\sigma-1}{1-\delta\sigma}} \left(\frac{1-\gamma}{\gamma}\right)^{\frac{1-\delta}{1-\delta\sigma}} \left(\frac{K}{L}\right)^{\frac{(\delta-1)(1-\sigma)}{1-\delta\sigma}}$$

These equations point to two parameters of interests: the substitutability of production factors, and the degree of state dependence. These parameters together determine how factors are rewarded given the direction of R&D activity (technical improvements). If, for example, the production factors are broad substitutes and there is a modest degree of state dependence (specifically, where $\sigma > 2 - \delta$), then an increase in the relative supply of one production factor (L, K) will induce an increase in the R&D investment directed at that factor (N_L, N_K), and the return to that factor as a share of output (Ψ_L, Ψ_K). That is, the (relatively) more abundant factor receives (relatively) greater rewards because of the technical improvements directed towards it, which — in the presence of a degree of state dependence — will support further technical improvements directed at that factor.

2.3 Skill-biased technical change

A specific application of directed technical change is that of *skill-biased* technical change. Acemoglu (2002b) outlines the interaction between the innovation possibility frontier and the returns to skill acquisition (that is, education). Rather than relying on labour and capital, the starting point in this model is two types of labour: low skilled and high skilled. As in the Solow model, technical efficiency augments labour. However,

[17] I have omitted other equations in Acemoglu (2002a) that lead to this expression; it cannot simply be derived from the equations of motion given earlier. Furthermore, I have omitted an additional parameter relating to the substitutability of outputs produced: the setup in terms of output production is not material to this discussion.

the difference is that each labour type has its own level of technical efficiency. Thus, the production function is expressed:

$$Y = F(A_L L, A_H H) = [(A_L L)^\rho + (A_H H)^\rho]^{\frac{1}{\rho}}$$

Where L and H denote, respectively, the total of low-skilled and high-skilled labour; A_L and A_H denote each labour type's efficiency. Both types of labour and technical efficiency are (in the first instance) exogenously determined — though endogenisation of technical efficiency is discussed below. The constant elasticity of substitution (CES) form of production function enables consideration of the relative substitutability of low- and high-skilled labour in the model. Where $0 < \rho \leq 1$, the two labour types are gross substitutes; where $\rho < 0$, the two are gross complements. Given perfect competition, each labour type earns their marginal product. That is:

$$w_L = \frac{\partial Y}{\partial L}, \qquad w_H = \frac{\partial Y}{\partial H}$$

All else being equal, the greater the share of the labour force that is low skilled, the greater is the wage to high-skilled workers — and vice versa. The relativity between the two wages — the idea of a skills premium — can also be explored. Specifically, by taking the log derivative of the ratio of wages with respect to the ratio of the labour types, one finds an expression for the substitution effect at play:

$$\frac{\partial \ln\left(\frac{w_H}{w_L}\right)}{\partial \ln\left(\frac{H}{L}\right)} = \rho - 1 = -(1-\rho)$$

As ρ cannot be greater than one this implies that, given a one per cent increase in the labour ratio between H and L, there is a $1-\rho$ per cent fall in the wage ratio between w_H and w_L. The responsiveness of the wage ratio to a change in the labour ratio is thus greater than 1 in the case of complements ($\rho < 0$). However, this holds the relative technical efficiency levels of the two labour types constant — and as these change, so too do relative wage levels change.

$$\frac{\partial \ln\left(\frac{w_H}{w_L}\right)}{\partial \ln\left(\frac{A_H}{A_L}\right)} = \rho$$

That is, improvements in the technical efficiency of high-skilled workers relative to low-skilled workers will result in an increase in the wage of high-skilled workers relative to low-skilled workers only if the two types of labour are substitutes. Thus, in this model, the degree of substitutability between labour types is material to the question of how productivity affects inequality.

Acemoglu (2002b) notes that estimating the size of ρ is difficult,[18] in part because economies incorporate multiple industries — in practice, substitution is not merely a question of changes within an industry but also across industries. Nevertheless, based on other studies, he concludes gross substitutability between low- and high-skilled workers is a plausible assumption (implying a credible ρ estimate of between zero and one half).

If low- and high-skilled labour types are broadly substitutes — and thus improvements in the relative technical efficiency of high-skilled labour contribute to higher wages for high-skilled workers relative to low-skilled workers — then the question is, what might drive such underlying changes in technical efficiency? One option, as Acemoglu (2002b) shows, is to endogenise the market's choice of technology investment in the model — incorporating directed technical change.

Consider a production process that mirrors the two labour-type production function outlined above, but now applies the same logic to two types of output that consumers may buy (denoted Y_L and Y_H, where the subscripts match to low- and high-skilled-produced goods). One type of product is produced by low-skilled workers using machinery that matches their low skills (N_L); the other product is produced by high-skilled workers using machinery that matches their high skills (N_H). In essence, production is fully segregated between low- and high-skilled activities, where the machinery is the embodiment of technical efficiency — increasing investment in high-skilled workers' machinery implies increased technical efficiency to high-skilled workers.

$$Y = \left(Y_L^{\rho} + Y_H^{\rho}\right)^{\frac{1}{\rho}}, \qquad \text{where } Y_L = N_L L;\ Y_H = N_H H$$

The prices of the two final goods are their respective marginal products (delivering matching equations to the identities given above for the respective wages of low- and high-skilled workers). The ratio of the prices is such that:

$$\frac{p_H}{p_L} = \left(\frac{N_H H}{N_L L}\right)^{\rho-1}$$

By contrast, the prices of the two types of machinery are given by (final good) producers' marginal willingness to pay for an additional unit: machinery is supplied by a profit-maximising monopolist, who incurs marginal costs of zero in producing additional units of existing machinery. On the face of it, this assumption seems implausible. However, it should be understood as the translation into physical terms of

[18] Acemoglu (2002b) focuses on the elasticity of substitution between labour types (the production factors), given by $\sigma \equiv \frac{1}{1-\rho}$. One can, in the alternative, thus express $\rho = \frac{\sigma-1}{\sigma}$.

the inherent non-rival nature of technical efficiency: once an idea is developed, any number of people can use it without imposing costs at the margin.

The profit-maximising monopolist will continue to produce each type of machine until (final good) producers' respective willingnesses to pay for each of the machines is equal. The marginal willingness to pay for each machine type is given by the partial derivative of the value of the relevant final good with respect to machinery.

$$\frac{\partial}{\partial N_L}(p_L Y_L) = p_L L, \qquad \frac{\partial}{\partial N_H}(p_H Y_H) = p_H H$$

The machinery monopolist's profit is maximised where:

$$\frac{p_H H}{p_L L} = 1$$

Note that the quantities of each labour type are fixed; it is only price that adjusts in this mechanism. And the price level can only adjust if the relevant supply of machinery adjusts. Recall that machinery, in this model, is a proxy for the availability of technical improvements. Therefore, in the broader interpretation, the balancing mechanism that enables equilibrium to be reached is the relative balance of technical efficiency directed at the two factors:

$$\frac{A_H}{A_L} \equiv \frac{N_H}{N_L} = \left(\frac{H}{L}\right)^{\frac{\rho}{1-\rho}}$$

Where $\rho > 0$, this implies the two goods are gross substitutes. In this case, as the relative supply of high-skilled labour increases, so too will the supply of high-skilled workers' machinery: that is, high-skilled technical efficiency.

While skill-biased technical change (and directed technical change more broadly) provides a useful framework for considering the interaction between productivity and inequality, there are two key shortcomings to this story. The first is that it (in practical application) requires some knowledge of how technical improvements are directed. Acemoglu (2002a) concedes this is hard: even when looking at patents as a proxy for technical improvements, he notes that we can only observe where patents are applied (that is, by industry). But we do not know at which production input (or labour skill type) the technical improvements was directed.

The broader point here is that, though one can reasonably confidently separate labour and capital, or distinguish between low- and high-skilled labour, productivity is rather less amenable to such clean distinctions.[19] Even in the two-input world, technical changes that principally augment one input may have spillover effects that serendipitously benefit the other input (either directly, or indirectly by inspiring new technical advances). Similarly, to borrow from the language of Mokyr (1990), there can be substantial technical leaps ('macroinventions') that of themselves need not be economically significant, but which open the door to new specific innovations ('microinventions') that can be directed at particular input factors. Correctly attributing these effects in any empirical sense is fundamentally more complicated than assuming one technical efficiency variable.

The second problem is that the model described above leaves unanswered the question of how workers become skilled. That is, there is no function to describe the choice made by individuals to 'upgrade' their human capital. One might well presume that a growing wage differential between low- and high-skilled labour should induce more people to pursue education.[20] This is not explicitly defined in the model presented here. To consider that issue, I turn now to another model.

2.4 The human capital – productivity nexus

As with Acemoglu (2002b), Galor and Moav (2000) also consider the relationship between productivity and inequality. They take skill-biased technical change as given; that is, technical efficiency is assumed to deliver greater relative rewards to high-skilled workers than low-skilled workers. Their model also seeks to endogenise technical efficiency growth. But the central driver in this case is human capital investment: as the returns to high-skilled labour grows, more individuals choose to become high skilled. In turn, increases in the level of human capital contribute to increases in the technical efficiency growth rate.

Consider a neoclassical production function in discrete time with constant returns to scale. In this model, output is a function of physical capital and human capital, and technical efficiency augments human capital. Human capital is the combination of raw labour with any education that augments ability.

$$Y_t = F(K_t, A_t H_t), \qquad \tilde{y}_t \equiv \frac{Y_t}{A_t H_t} = f(\tilde{k}_t), \qquad \tilde{k}_t \equiv \frac{K_t}{A_t H_t}$$

[19] Of course, one can flip this on its head. Maybe the flawed assumption is that there exists some meaningful aggregate measure of productivity — that the more 'natural' approach is to consider a separate efficiency factor for each input. But one can easily extend the thought to question the underlying wisdom of growth accounting: that an entire economy, with multiple types of output and inputs, can be modelled with one production function with homogeneous factors. And that would be a rather larger question than this thesis considers.

[20] Acemoglu (1998) considers the scenario of technical improvements that complement skills. Using a model with endogenous supply of skills, he shows how skill-augmenting improvements could increase the returns to skilled labour, inducing more individuals to become skilled, increasing the market for skill-augmenting improvements, spurring new skill-augmenting improvements, which increase the returns to skilled labour and so on.

In this model, the definition of 'per effective worker' that gives the intensive form of variables implies division by AH (rather than AL, as discussed previously).

The model assumes a small open economy. The small open economy can borrow or lend capital at the world rental rate, which is held constant; the small open economy is 'small' in the sense that it cannot influence world prices. But there is perfect competition (such that the price of each factor equals its marginal product), and producers are profit maximisers. Hence, the level of capital per effective worker equilibrates to the factor price for capital: the world rental rate. As this is constant, so too is the level of capital per effective worker in the economy.[21]

$$r_t = f'(\tilde{k}_t) \implies \bar{r} = f'(\bar{k})$$

This constant rate of $\tilde{k}_t = \bar{k}$ also affects the factor price for human capital.

$$w_t = A_t\big[f(\tilde{k}_t) - f'(\tilde{k}_t)\tilde{k}_t\big] = A_t\big[f(\bar{k}) - f'(\bar{k})\bar{k}\big] \equiv A_t\bar{w}$$

Note that this expression is not the same as the wage paid to workers, as will be shown. Specifically, workers are heterogeneous, with differing levels of innate ability on a spectrum from zero to one. Those of higher ability (above some threshold) choose education and become high-skilled workers. The remainder (of ability zero up to the threshold level) are low-skilled workers.

Aggregate human capital — comprised of units of high-skilled labour (h_t) and low-skilled labour (l_t) — is given by:

$$H_t = \beta h_t + l_t(1 - \delta g_t), \qquad \beta > 1;\ 0 < \delta g_t < 1;\ g_t \equiv \frac{A_t - A_{t-1}}{A_t}$$

This expression implies a human-capital premium from high-skilled workers (β), while the human-capital contribution of the low-skilled workers is depreciated by some factor (δ) of the efficiency growth rate. The higher is efficiency growth, the greater is the decline in the human-capital weighting of low-skilled workers — though the parameter restrictions ensure that their human capital contribution cannot be completely wiped out.

The factor prices for units of low- and high-skilled labour are given by their marginal products (note that one *unit* is not the same as one *individual*; this point is discussed below). That is, the derivative of the

[21] This assumption offers a useful analytical simplification: the model is focused on the interaction between A and H, and thus holding k constant takes this variable out of play. Moreover, as Galor and Moav (2000) observe, neither the decision to invest in human capital nor the effects in terms of income inequality are influenced by r.

output function with the embedded aggregate human capital function, with respect to both l_t and h_t. That is:

$$w_{l,t} = (1 - \delta g_t)A_t\overline{w}, \qquad w_{h,t} = \beta A_t\overline{w}$$

As noted, ability helps determine whether an individual chooses education and therefore become a high-skilled worker. But as the above equations suggest, it is not the sole factor. As the efficiency growth rate increases, the relative return to high-skilled labour rises. But there is a cost to education. Specifically, there is an opportunity cost: for the time one is studying, one does not earn any income. The underlying tradeoff gives rise to an optimisation problem.

Galor and Moav (2000) assume a two-period, overlapping-generations model. In this model, an individual lives for two periods. During the first period, the individual works — and if she chooses to study, she studies for some time share of the period, τ. During the second period, the individual is retired. All individuals have the same utility function; they are identical except for differences in ability. An individual's utility is derived from consumption. Thus, to maximise utility, one must maximise lifetime consumption;[22] to maximise one's consumption, one must maximise income.

Galor and Moav (2000) provide two different expressions for the income of (respectively) low- and high-skilled workers:

$$I_{i,l,t} = w_{l,t}l_{i,t} = (1 - \delta g_t)A_t\overline{w}[1 - (1 - a_{i,t})g_t]$$

$$I_{i,h,t} = w_{h,t}h_{i,t} = \beta A_t\overline{w}(1 - \tau)[a_{i,t} - (1 - a_{i,t})g_t] = A_t\overline{w}[a_{i,t} - (1 - a_{i,t})g_t], \qquad \beta = \frac{1}{1 - \tau}$$

In both cases, the expressions contained with the square brackets are the individual's effective labour contribution given their ability type and education — that is, how much of a unit of labour one individual contributes given their ability and the efficiency growth rate. In both cases, efficiency growth has a depreciating effect on the contribution of labour — an effect that falls away the higher one's ability is. Furthermore, for individuals that choose to become high-skilled workers, education unlocks a direct benefit of their ability. But this expression also means that low ability types have little incentive to become high-skilled workers: their $a_{i,t}$ will be lower, while their $(1 - a_{i,t})$ will be greater.

[22] This abstracts from any consideration of the intertemporal choice of relative consumption between the two periods. The point is simply that one derives greater utility from being able to consume more across *both* periods: this underpins the typical optimisation problem assumption that an individual fully exhausts their budget constraint.

As a simplifying assumption, Galor and Moav (2000) define $\beta(1-\tau)=1$; the premium to high-skilled workers must compensate for the time cost of education. This condition also ensures that at least some individuals in the economy will choose education.

Taken together, the individual's choice of education simplifies to a comparison of two possible income functions. If the individual, given her ability type and the economy's efficiency growth rate, has greater earning potential from becoming high skilled than remaining low skilled, she will choose education. As a tiebreaking condition, Galor and Moav (2000) further assume that an individual will choose education if their income is equal under both functions.

$$I_{i,h,t} \geq I_{i,l,t}$$

The point of equality between the two functions also reveals the ability threshold, given the efficiency growth rate. The higher the efficiency growth rate is, the lower is the ability threshold at which an individual will choose to become high skilled.

$$a_t^* = \frac{1-\delta g_t + g_t^2}{1+\delta g_t^2}$$

Given that all individuals face the same optimisation problem with respect to education, all individuals with ability greater than or equal to the applicable threshold will become high-skilled workers. Thus, the ability threshold also defines the proportion of the workforce in a given period that will invest in human capital. In turn, this proportion of high-skilled workers $(1-a_t^*)$ influences the efficiency growth rate in the next period. Galor and Moav (2000) assume the following function:

$$g_{t+1} = \gamma(1-a_t^*) = \frac{\gamma\delta g_t}{1+\delta g_t^2}, \qquad 0<\delta<1; \frac{1}{\delta}<\gamma<1+\frac{1}{\delta}$$

Key conditions here are that g_{t+1} is a strictly increasing, strictly concave function with respect to g_t. The parameter assumption for γ is necessary for achieving a stable steady state, where $g_{t+1}=g_t \equiv g_{ss}$. (There is also a steady state at $g_{t+1}=g_t=0$, but this is trivial. For any positive value of g_t, the efficiency growth rate converges on g_{ss}, subject to the parameter restriction on γ.) The steady state is determined by the parameter values of δ and γ (bounded by their restrictions).

$$g_{ss} = \left(\frac{\gamma\delta-1}{\delta}\right)^{\frac{1}{2}}, \qquad 0<g_{ss}<1$$

What are the implications of this? If a starting (positive) value g_t is less than the steady state (that is, $0<g_0<g_{ss}$), then the efficiency growth rate will rise over time. As the efficiency growth rate rises, the

ability threshold for undertaking an education falls: the proportion of high-skilled workers in the economy increases. Furthermore, despite this supply increase, the average wage of high-skilled workers will rise — with high-skilled workers of the highest levels of ability enjoying the greatest income gains.

It is also relevant to observe an alternative scenario: consider an economy in steady state. A shock lifts the efficiency growth rate above its long-term rate (that is, g_{ss}). The efficiency growth rate will converge to its steady state, but from a higher level: that is, the efficiency growth rate must fall over time. The long-term effect here is unremarkable: the post-shock economy, converging back to steady state, will over time exhibit the same conditions as the pre-shock economy in steady state.

The short-term effect is rather more interesting. Recall that income is a function of the efficiency growth rate; low-skilled workers are more adversely affected by a rise in the efficiency growth rate than high-skilled workers. The higher efficiency growth rate induced by the shock will deliver greater returns to high-skilled labour. That is, inequality will increase following a (positive) productivity shock — but only temporarily. Inequality will return to its earlier (pre-shock) level, as the efficiency growth rate returns toward its steady state level over time.

An additional feature of Galor and Moav's (2000) model is that the differences in individual's ability levels also provide for variation within the groups of low- and high-skilled labour. These within-group differences can be exploited to comment on inequality between different workers with the same education level. The headline finding is that, as efficiency growth rises, so too does within-group inequality — just as it does between the groups.

As with any (useful) model, Galor and Moav (2000) rely on considerable simplifying assumptions. But some assumptions have a greater effect than others.

One assumption that merits discussion is the restriction on g_t: the efficiency growth rate must not exceed one. At first pass, this might not seem like a troubling restriction. Given the observed facts of productivity, the developed world is a long way off annual TFP growth rates of 100 per cent. However, annual growth rates are not what matters in this model. This model considers the decisions of a cohort of workers in a two-period overlapping generations model. There is no explicit definition of what the length of time period is, but if one conservatively assumes a 30-year working lifespan, it is far from inconceivable that the level of technical efficiency could more than double within a period. Indeed, based on the data presented earlier in this thesis, that is precisely what occurred during the 30-year period to 1980: the observed level of average productivity across the sample countries by 1980 was 2.1 times the level in 1950. While productivity growth is lower on average today than in the post-war era, the possibility of stronger growth again in future cannot be excluded.

A further simplification that Galor and Moav (2000) introduce relates to the human capital of high-skilled workers: specifically, that the human capital premium must fully offset the opportunity cost of education (mathematically, that $\beta(1-\tau)=1$). Furthermore, β and τ are constants. This leads to an asymmetry in the model: ability is allowed to vary between 0 and 1, but ability affects only whether one chooses to pursue education — a binary choice between 0 and 1. What this loses, is the choice of education level: how much human capital an individual chooses to invest in.

This question features in Galor's unified growth theory (Galor 2011; 2005). The full scope of that theory, which proposes a model that explains humanity's economic development from the Malthusian era to today, is not material to this thesis. But the model's human capital function is relevant. Here, an individual's human capital is a function of both their education level and the efficiency growth rate, such that (consistent with the notation above):

$$h_{i,t} = h(\tau_{i,t}, g_t)$$

Human capital is concave with respect to education level and convex with respect to efficiency growth. But there is complementarity as well: higher efficiency growth induces greater investment in education. What is missing, in the context of describing the choice between being a low-skilled or high-skilled worker, is ability. I propose to address this in an augmented model of Galor and Moav (2000), which explores the effect of education level on productivity and inequality.

3 The model

The preceding discussion has identified some useful theoretical insights about the relationship between inequality and productivity. A factor that is central to this relationship is human capital: different workers have different levels of human capital, whether due to their innate abilities or a conscious decision to upgrade skills through education (a choice which is partly influenced by innate ability). The greater an individual's human capital is, the greater are their potential earnings as a worker — differences in human capital across the labour force are a contributing factor in observed income inequality. Moreover, human capital can be hypothesised as contributing to efficiency growth: higher levels of education enable greater production of new ideas, or better adaptation of ideas in the production of output.

But the discussion also points to further opportunities for exploration. Specifically, whereas Galor and Moav (2000) show that the share of the workforce that chooses to become high skilled matters for both inequality and productivity, I propose to demonstrate that the amount of education individuals choose also has relevant effects.

The proposed model here takes as its starting point the model of Galor and Moav (2000). It is further supplemented by insights from Galor (2011; 2005) and Acemoglu (2002b). I start by sketching out the model in terms of the key functions and their properties, before proceeding to define and apply explicit (closed form) functions.

3.1 Functional outline

The model starts with a neoclassical production function with constant returns to scale. Output is a function of physical capital and human capital; the latter is augmented by technical efficiency.

$$Y_t = F(K_t, A_t H_t)$$

The assumption of a small open economy ensures that capital (and therefore output) per effective worker is constant.

$$\frac{K_t}{A_t H_t} \equiv \tilde{k}_t = \bar{k}, \qquad \frac{Y_t}{A_t H_t} \equiv \tilde{y}_t = f(\bar{k}) = \bar{y}$$

Human capital is embedded in workers — once those workers retire, their human capital ceases to contribute to production. However, the knowledge and ideas those workers generate as a product of their human capital accumulation contributes to improvements in technical efficiency over time. Hence, the average level of the workforce's human capital in one period contributes to the next period's efficiency growth:

$$\frac{A_{t+1} - A_t}{A_t} \equiv g_{t+1} = \mathrm{M}\left(\frac{H_t}{L}\right)$$

Where M is some function of average human capital that gives the properties of a positive, concave function with respect to g_t,[23] such that:

$$\mathrm{M}(g_t = 0) = 0, \qquad \mathrm{M}'(g_t) > 0, \qquad \mathrm{M}''(g_t) < 0$$

It is worth reflecting on the role M plays, and through it, the contribution of average human capital. It need not be assumed that average human capital is the *only* factor that contributes to growth in technical efficiency over time. Indeed, in reality, such an assumption would be wholly implausible. For the purposes of this thesis, M is simply a function that controls how much human capital accumulation contributes to technical efficiency improvements.

Human capital is a function of raw labour and technical efficiency growth. Labour is assumed to be of a constant size — that is, the labour force in every period consists of L workers;[24] there is no population growth. This assumption does not have any material effect on the results.[25]

$$H_t = G(L, g_t)$$

This aggregate human capital expression is a sum of each individual's human capital. That is:

$$H_t = \sum_{i=1}^{L} h_{i,t}(\tau_{i,t}, a_i, g_t)$$

This individual human capital function is integral to the model. Individuals choose how much human capital to acquire ($\tau_{i,t}$, a proportion of the working time period the individual devotes to education), given their

[23] The equivalent function in Galor and Moav (2000) is denoted Φ, with the same properties identified here.
[24] Strictly speaking, the designation L refers to the total number of students and workers in a given period. But all individuals in this cohort will be workers at some point during the period.
[25] As noted, efficiency growth is influenced by average human capital. Moreover, while aggregate human capital will be greater the more workers there are, the overall mix of education levels does not depend on the size of the labour force — only the relative shares of different ability types.

level of ability (a_i). This choice, and in turn the individual's human capital level, also depends on the rate of technical efficiency.

$$h_{i,t} = h_{i,t}(\tau_i, a_i, g_t), \qquad \tau_{i,t} = \tau_{i,t}(a_i, g_t)$$

$$0 \leq \tau_{i,t} \leq 1;\ 0 \leq a_i \leq 1$$

Ability is a one-dimensional parameter, taking a value along the range 0 to 1. Ability here is both narrowly and broadly defined. Ability is narrow in the sense that this term is an expression of how receptive an individual is to education — that some people are more likely to thrive in, and get something out of, a learning environment than others. Ability is therefore broad in the sense that there are many contributing factors — innate qualities of the individual, including their preferences in life. In that sense, the word 'ability' should be thought of as shorthand for an ability to realise economic gains from human capital accumulation.

The higher the ability level is, the greater are the returns from investing in education. Moreover, an individual of ability type 0 has no incentive to invest in education — their maximum individual human capital level is one (their unit of raw labour).[26]

$$h'_{i,t}(a_i) > 0$$

$$h_{i,t}(a_i = 0) = 1$$

There are some important features of the relationship between education, efficiency growth and the human capital level that should be observed. Consistent with the function adopted in unified growth theory (Galor 2011; 2005), the human capital level is concave with respect to education, and convex with respect to efficiency growth. That is, there are positive but diminishing marginal returns from education to human capital — as the share of time devoted to education increases, the human capital gain from each incremental increase is progressively smaller. Convexity with respect to efficiency growth owes to an erosion effect: as technical efficiency rises, accumulated knowledge becomes redundant, thus diminishing the value of human capital holding all else constant. A further condition, however, is that the interaction between changes in the level of education and the rate of efficiency growth is such that, as both increase,

[26] Note that this is a different assumption from Galor and Moav (2000), who depreciate the effective value of unskilled labour by some function of the efficiency growth rate. But their model also assumes that there must be some skilled labour in the economy. By contrast, the model proposed here allows the flexibility of no skilled labour — if all workers remain unskilled, then there is no human capital beyond raw labour (that is, $H_t = L$), and the production function collapses to a basic Solow-model setting with a fully exogenous efficiency growth rate ($g = \mu$).

the effect on human capital accumulation is positive. Expressed another way, an individual should have an incentive to choose more education when efficiency growth is increasing.

$$h'_{i,t}(\tau_{i,t}) > 0, \qquad h''_{i,t}(\tau_{i,t}) < 0$$

$$h'_{i,t}(g_t) < 0, \qquad h''_{i,t}(g_t) > 0$$

$$\frac{\partial^2 h_{i,t}}{\partial \tau_{i,t} \partial g_t} > 0$$

In addition, the presence of ability in the model does not change these core identities: ability merely reinforces the existing direction of the first-order effects from education level and efficiency growth. That is, in a more technical sense, the second mixed partial derivatives of human capital with respect to ability and (respectively) education level and efficiency growth are (respectively) positive and negative.

$$\frac{\partial^2 h_{i,t}}{\partial a_i \partial \tau_{i,t}} > 0, \qquad \frac{\partial^2 h_{i,t}}{\partial a_i \partial g_t} < 0$$

The individual's choice implies an optimisation problem. The mechanism underpinning this is a standard two-period overlapping generations model. Each individual lives for two periods. In the first period, they work and — if they choose to — pursue an education. In the second period, they retire; their only income in the second period comes from their savings in the first period. All individuals have an identical utility function, where their utility comes from consumption over the two time periods. To facilitate this consumption, they seek to maximise their lifetime income. By design, this means they must maximise their labour income in the first period. Hence, they choose what level of education will deliver the highest income — and this choice will depend on the factors outlined above. Individuals of a given ability type are homogeneous, and thus have the same optimal choice.

$$\max_{\tau_{i,t}} y_{i,t} = (1 - \tau_{i,t}) w_t h_{i,t}$$

The model does not include any direct financial costs of pursuing an education.[27] The only cost is the opportunity cost: for the period an individual is in education, they do not work and therefore earn no income.

[27] Other models, including the supplementary three-period model discussed in Galor and Moav (2000), do consider financial costs, which allow for relevant discussions about the effect of (for example) credit market imperfections. However, as an Australian (where university students have default access to a public loan system, where students start to repay their loan only once they earn an income above a defined threshold) now living in Denmark (where there is no user payment for education, and students receive a monthly allowance for the duration of their studies), I am relaxed about omitting financial costs from this model.

Factor prices are, consistent with standard models, equal to their marginal product. Given the small open economy assumption, the level of capital employed in the domestic economy automatically adjusts in line with the global rental rate ($r_t = \bar{r}$), which is constant. Each worker's wage is a function of their own human capital and the level of technical efficiency.

$$w_{i,t} = \bar{w} A_t h_{i,t}, \qquad \bar{w} A_t \equiv w_t, \qquad \bar{w} = \frac{\partial Y_t}{\partial (A_t H_t)}$$

The notation here allows for a general baseline wage in the economy (w_t) for the given time period. This baseline wage is what workers without any human capital accumulation earn. Note also that $\bar{w}$ is a constant, again due to the small open economy assumption (as this price is exclusively a function of capital per effective worker, which is a constant).

While it is no way material, to aid in understanding the model, it is worth recalling that there is a two-period OLG model in the background — this is what gives rise to the individual's optimal choice of education. It also means therefore that factor prices are expressed in terms of earnings across a single period. Furthermore, as individuals choose how much to study (and therefore not work) during the period, their actual earnings for the period must be adjusted by a factor of $\left(1 - \tau_{i,t}\right)$.

3.2 Application

Converting the broad contours of the model into something specific requires choices. Central to this exercise is the human capital function, where an explicit functional form must satisfy multiple first- and second-order conditions.

3.2.1 The individual's human capital function

Let an individual's human capital be expressed as follows:

$$h_{i,t} = h_{i,t}\left(\tau_{i,t}, a_i, g_t\right) = \left(1 + \frac{a_i}{1 + g_t} + (1 + a_i)\tau_{i,t}\right)^{a_i}$$

What does this expression say? Working through each of the function's three factors:

- Human capital rises with education: If $\tau_{i,t} = 0$ — that is, the individual chooses no education — then only the relationship between the individual's ability level and efficiency growth matters for their overall human capital.
- Human capital rises with ability. The positive effect of ability on human capital enters the model in three ways: the exponent, which scales the overall return to the individual of human capital accumulation (where $a_i = 0$, $h_{i,t} = 1$); the numerator of the second term, where a_i is subject to

the erosion effect of efficiency growth; and the third term, where a_i augments the gains from each increment of education.

- Efficiency growth is responsible for the erosion effect with respect to ability: higher efficiency growth depreciates ability. Note as well, in contrast to Galor and Moav (2000), that there is no upper bound on the growth rate: the only requirement here is that $g_t \neq -1$ (which is to say, the entire productive capacity cannot be wiped out). In any event, I will consider only $g_t > 0$.

The combination of these elements gives rise to a human capital function that adheres to the first- and second-order conditions assumed in the preceding discussion of the open form model. The function is positive with respect to ability, concave with respect to education level, and convex with respect to efficiency growth. The key mixed partial derivative (the cross effect of education and efficiency growth) is positive.

It follows from this expression that the optimal education level for a given ability type is given by the expression:

$$\tau_{i,t}^* = \begin{cases} \dfrac{a_i(1+a_i) - \left(1 + \dfrac{a_i}{1+g_t}\right)}{(1+a_i)^2} & a_i(1+a_i) > 1 + \dfrac{a_i}{1+g_t} \\ 0 & a_i(1+a_i) \leq 1 + \dfrac{a_i}{1+g_t} \end{cases}$$

The inequality expressions are not immediately easy to interpret. Rewriting the condition for a positive education level, the growth rate must be sufficiently high (for a given ability level) that:

$$g_{min}(g_t > 0) > \frac{1-a_i^2}{a_i(1+a_i)-1}, \qquad \frac{\partial g_t}{\partial a_i} > 0$$

For any positive growth rate, an individual with ability $a_i = 1$ will always choose some level of education. As the growth rate continues to increase, lower ability types will begin to choose some level of education. The minimum level of ability, given any growth rate, translates to approximately 0.62 — or precisely $(\sqrt{5}-1)/2$.

In turn, the human capital level for a given ability type, given their optimal education level, is:

$$h_{i,t}^*(\tau_{i,t}^*, a_i, g_t) = \begin{cases} \left(1 + \dfrac{a_i}{1+g_t} + \dfrac{a_i(1+a_i) - \left(1 + \dfrac{a_i}{1+g_t}\right)}{1+a_i}\right)^{a_i} & a_i(1+a_i) > 1 + \dfrac{a_i}{1+g_t} \\ \left(1 + \dfrac{a_i}{1+g_t}\right)^{a_i} & \tau_{i,t}^* = 0 \end{cases}$$

3.2.2 Aggregate human capital: two ability types

To make the problem more tractable, a significant simplifying assumption is applied with regard to ability. I assume (at least in the first instance) that there are two ability types: $a_0 = 0$ and $a_1 = 1$. There are βL individuals of ability type a_0 and $(1-\beta)L$ individuals of ability type a_1. Thus, the economy's human capital can be defined as:

$$H_t = \beta L\, h_{0,t} + (1-\beta)L\, h_{1,t}$$

Where the i notation is replaced by 0 and 1 depending on the ability type. (To reiterate, all individuals of the same ability type are identical, and thus make the same optimising choice with respect to education.) Furthermore, a consequence of setting $a_0 = 0$ is that $h_{0,t} = 1$: an individual of ability type a_0 will not choose any education.

Thus, the interesting question is how much education an individual of ability type a_1 will choose — and how this choice will change over time, with changes in g_t. With $a_1 = 1$, the optimisation problem takes the form:

$$h_{1,t} = h_{i,t}(\tau_{1,t}, a_1, g_t) = \left(1 + \frac{(1)}{1+g_t} + (1+(1))\tau_{1,t}\right)^{(1)} = 1 + \frac{1}{1+g_t} + 2\tau_{1,t}$$

$$\max_{\tau_1,t} y_{1,t} = (1-\tau_1)w_t\left(1 + \frac{1}{1+g_t} + 2\tau_{1,t}\right)$$

Taking the first derivative of $y_{1,t}$ with respect to $\tau_{1,t}$, and isolating for the optimal $\tau_{1,t}$ yields:

$$\tau_{1,t}^* = \frac{g_t}{4(1+g_t)}$$

As noted earlier, an individual of ability $a_1 = 1$ will always choose some amount of education given a positive efficiency growth rate. The individual's human capital, given their education, is thus:

$$h_{1,t}^* = 1 + \frac{1}{1+g_t} + 2\frac{g_t}{4(1+g_t)} = 1 + \frac{2+g_t}{2(1+g_t)}$$

Consequently, the economy's aggregate human capital is:

$$H_t = \beta L + (1-\beta)L\left(1 + \frac{2+g_t}{2(1+g_t)}\right)$$

With human capital per worker:

$$\frac{H_t}{L} = \beta + (1-\beta)\left(1 + \frac{2+g_t}{2(1+g_t)}\right) = (1-\beta)\left(\frac{2+g_t}{2(1+g_t)}\right) + 1$$

3.2.3 Efficiency growth

With the human capital side of the model determined, the next function to formalise is the efficiency growth function, denoted M. Let efficiency growth develop in a Cobb–Douglas function, such that:

$$g_{t+1} = \mathrm{M}\left(\frac{H_t}{L}\right) = g_t^{\eta}\left(\mu\frac{H_t}{L}\right)^{1-\eta} = g_t^{\eta}\left(\mu(1-\beta)\left(\frac{2+g_t}{2(1+g_t)}\right)+\mu\right)^{1-\eta}, \qquad \mu > 0;\ 0 < \eta < 1$$

The transformation of average human capital through the efficiency growth function ensures that the conditions of concavity are met. This equation can be interpreted as saying that human capital contributes *some* productivity gain (adjusted by a factor of μ). But it is not the sole driver of efficiency growth — g_t directly enters the function external to human capital's contribution. The relative contributions of the internal (human capital) and external factors is regulated by the parameter η.

The efficiency growth function results in a quadratic expression for steady state — that is, where $g_{t+1} = g_t \equiv g_{ss}$. This implies two steady states (of a rather inelegant form):

$$g_{ss} = \frac{(\mu(3-\beta)-2) \pm \sqrt{(\mu(3-\beta)-2)^2 + 16\mu(2-\beta)}}{4}$$

In fact, there is only one steady state that we are interested in — a steady state greater than zero is only positive through the positive sign of the root. (And the positive sign of the root cannot produce a negative steady state.) The parameter assumptions for β and μ prevent the root from being smaller than the first term in numerator. Here, I will use 'steady state' in the singular, focusing only on the steady state emerging from the positive sign of the root.

It is apparent that the value of the steady state depends on the values of β and μ. The β parameter is relatively straightforward to consider: it is restricted to the range (between and including) zero to one. If $\beta = 1$, the workforce is comprised solely of low-skilled workers, who do not accumulate human capital. It follows that $g_{ss}(\beta = 1) = \mu$. That is, efficiency growth is fully exogenous.

The result for any value of β depends on μ. Without an assumed value for μ, the most that can be said is that $g_{ss} > 0$ and g_{ss} depends positively on μ. As μ goes to zero at the limit, g_{ss} also converges on zero.

Any positive steady state will be stable. Starting from the equation for g_{t+1}, differentiating with respect to g_t and then rearranging leads to the following expression:

$$\frac{\partial g_{t+1}}{\partial g_t}\frac{g_t}{g_{t+1}} = \eta - \frac{g_t}{1+g_t}\frac{(1-\eta)(1-\beta)}{(2+g_t)(1-\beta)+2(1+g_t)}$$

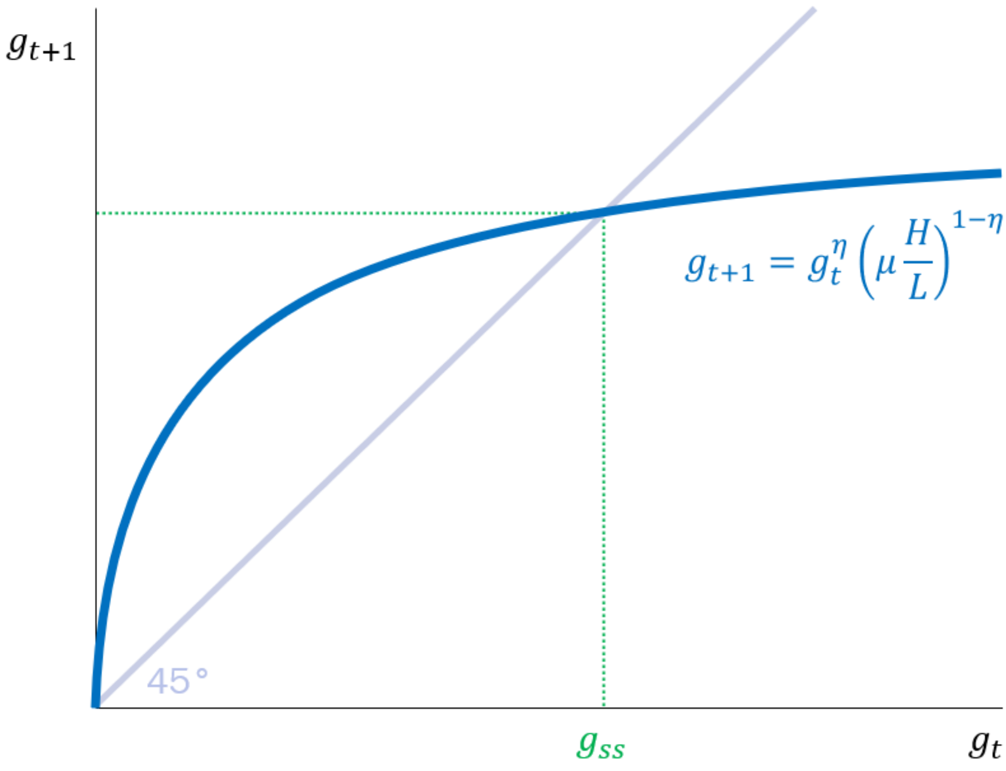
Due to the parameter restrictions, the right-hand side of the equation will have an absolute value less than one. Specifically:

- $g_t > 0 \Longrightarrow 0 < \frac{g_t}{1+g_t} < 1$
- $0 < \eta < 1, 0 \leq \beta \leq 1 \Longrightarrow 0 \leq (1-\eta)(1-\beta) \leq 1$
- $(2+g_t)(1-\beta) + 2(1+g_t) > 2$

Hence stability requires only that the ratio between g_t and g_{t+1} (set on the left-hand side of this equation above) not be sufficiently smaller than one such that (for the equality to hold) $\partial g_{t+1}/\partial g_t > 1$. Evaluating the derivative in steady state means that this ratio is equal to one ($g_t = g_{t+1}$). Thus, the absolute value of the first partial derivative is less than one; the condition for steady state stability is met. Figure 6 illustrates the concave efficiency growth function.

Figure 6 Efficiency growth function

Transition diagram



3.2.4 Returns to labour

One consequence of the assumption of two ability types in this model (a_0 and a_1) is that there are low-skilled workers who will never invest in human capital accumulation, and there are high-skilled workers who (for any $g_t > 0$) will always allocate some of their time to human capital accumulation. This has implications for their earnings — and income inequality within the economy. Specifically, the wages for the two ability types are given by:

$$w_{0,t} = \overline{w}A_t, \qquad w_{1,t} = \overline{w}A_t\left(1 + \frac{2 + g_t}{2(1 + g_t)}\right)$$

Where high-skilled workers choose their optimal education level. In turn, the ratio of high- to low-skilled wages is equal to the ratio of their respective human capital (which is $h_{0,t} = 1$ for low-skilled workers):

$$\frac{w_{1,t}}{w_{0,t}} = 1 + \frac{2 + g_t}{2(1 + g_t)} = h_{1,t}$$

This ratio decreases as the efficiency growth rate rises, with the erosion effect reducing the gains to high-skilled workers from human capital. Moreover, higher efficiency growth also means that the overall technical efficiency level (A_t) rises faster: with gains accruing to both low- and high-skilled workers in equal measure. In short, higher efficiency growth reduces inequality and raises labour incomes — at least in this highly stylised model.

3.3 Extending the model

The fundamental results of the model are in line with Galor and Moav (2000). This is unsurprising: the broad functional design, particularly in relation to the interaction between human capital and efficiency growth, is unchanged. In the remained of this section, I open the door on possible changes and extensions of the model that can add further detail on some of the dynamics — but also reveal some of the model's limitations.

3.3.1 Multiple ability types

Restricting the model to two ability types is analogous to Galor and Moav's (2000) restriction on the education level. I have simply replaced one binary choice with another. This is mathematically convenient but sacrifices some richness in the story.

The model, however, can accommodate multiple ability types: the individual's human capital function allows for any ability level to be chosen between 0 and 1. Changing the assumption of the distribution of ability only affects how the aggregate human capital function is defined.

Consider an expanded model with five ability types. In addition to $a_0 = 0$ and $a_1 = 1$, three further ability types are introduced: $a_{0.7} = 0.7$, $a_{0.8} = 0.8$ and $a_{0.9} = 0.9$. (As noted earlier, there is a minimum ability threshold given any growth rate — approximately 0.62. The three values here have been deliberately chosen as comfortably above this threshold.) To aid understanding of the model, assume now that the five ability types are equally distributed: they each account for 20 per cent of the labour force. The aggregate human capital function now takes the form:

$$H_t = 0.2L(h_{0,t} + h_{0.7,t} + h_{0.8,t} + h_{0.9,t} + h_{1,t})$$

As before, individuals of ability type a_0 will not choose any education. And individuals of ability type a_1 will always choose some level of education. The question is, how much education will the additional three ability types choose? In the first instance, the choice of education level (given one's ability) depends on the efficiency growth rate. Given the expression for a threshold efficiency growth rate, expressed in terms of ability type, it is relatively straightforward to determine the level at which education becomes attractive to each of the additional ability types.

$$g_{min} > \frac{1 - a_i^2}{a_i(1 + a_i) - 1} = \begin{cases} 2.69 & a_{0.7} = 0.7 \\ 0.82 & a_{0.8} = 0.8 \\ 0.27 & a_{0.9} = 0.9 \end{cases}$$

If the initial growth rate (g_0, where $t = 0$) is close to but greater than zero, then in the first period only individuals of ability type a_1 will choose education. Nevertheless, this human capital accumulation contributes sufficiently to the efficiency growth rate that the next period's efficiency growth rate will be higher, inducing an increase in the amount of education. The parameter settings in the efficiency growth function determine the results here: η controls the speed at which the increase over successive generations occurs (the lower is η, the faster); μ controls the extent to which ability types lower than a_1 begin to accumulate human capital (the higher is μ, the greater the incentive).

In summary, as the efficiency growth rate rises, a greater proportion of the population chooses education — up to a threshold ability level. And those that choose education will choose progressively higher amounts of education.

3.3.2 Directed technical change

The model as described above is based around a single technical efficiency variable. And this technical efficiency variable augments human capital in the production function. The design of the human capital function means that the average level of education in the economy increases with efficiency growth — both in the two-ability-type setup, where high ability types invest more time in education; and in the extended setup with multiple ability types, where a greater proportion of the labour force begins to acquire education as the efficiency growth rate increases. In other words, the model assumes skill-biased technical change as a starting point.

If one were to try and establish skill-biased technical change as a function of the model, the starting point would be to amend the individual human capital function, such that the relevant augmenting efficiency growth rate for a given ability level was applied ($g_{i,t}$ rather than g_t):[28]

$$h_{i,t} = h_{i,t}(\tau_{i,t}, a_i, g_{i,t}) = \left(1 + \frac{a_i}{1 + g_{i,t}} + (1 + a_i)\tau_{i,t}\right)^{a_i}$$

The question would then be, what determines the respective efficiency growth rates? As outlined earlier, the endogenous model of Acemoglu (2002b) proposes a market mechanism in the supply of new technical improvements. This makes sense in explaining choices about whether to supply technical improvements directed at one ability/skill type over another. But it would also represent a fundamentally different approach from Galor and Moav (2000) and the model presented in this section, where g_{t+1} is a function of average human capital. (That is, efficiency growth is partly the result of positive spillover effects from increased knowledge accumulation, rather than a specific innovation production process.)

Alternatively, one might simply assume different parameter settings (μ_i rather than μ), such that average human capital has a greater effect on the efficiency growth augmenting higher skill levels. But then this would essentially reduce the question to an exogenous one. Consequently, the bias in technical change would still be an assumption rather than a feature of the model.

3.3.3 The production function

It is notable that, in presenting this model, I have had little to say about the production function. The 'action' in the model is contained to the labour side of the function (the combination of technical efficiency and human capital). This is a direct consequence of the assumption of a small open economy: the key results emerge without any interesting consequences in terms of capital accumulation. It is sufficient to say that, whatever explicit production function one assumes (within the boundaries defined by the model), aggregate output increases as both the technical efficiency level and aggregate human capital level increase.

[28] In the case that one employs the extended, multiple-ability-type specification of the model, one could restrict the analysis to two types of efficiency growth: one associated with workers that have any level of education, the other attached to workers that have chosen no education.

There is a gap that this model leaves unexplored: the relationship between different labour types and capital and the extent to which they are substitutable with one another. Why might this be interesting in terms of inequality? One concern might be technical changes, such as robotics and automation, which enable labour — and especially low-skilled labour — to be replaced by capital (Acemoglu and Restrepo 2019). Viewed through the purely theoretical prism, this is not necessarily a flaw. Given a long-term model assuming perfect competition and full employment, the intrinsic expectation is that no workers are ever permanently displaced. Thus, if the concern were about workers being pushed out of the labour force by machines, one would not start with this model.

But even with that qualification, one might still expect that the degree of substitution between capital and different types of labour would be reflected in factor prices, and likely also (in the microeconomic sense) the investment incentives with regard to technical improvements.

The natural approach to test this question would be, in line with Acemoglu (2002b), to adopt a CES production function. However, there are two related problems with this approach. The first is that, with more than two inputs — capital, and low-skilled and high-skilled labour (in the case of the two-ability-type model, which is what I discuss here) — the CES function becomes more complicated.[29] A nested approach, with one CES function inside another CES function, would be feasible: allowing, for example, two variables to be complements (the nested function) and those two variables to be substitutes with a third variable (the main function).

This, however, leads to the second problem. The model already envisages, in effect, nested functions. The production function is a function of physical capital and human capital (with human capital augmented by technical efficiency). And as has been shown, aggregate human capital includes different types of labour. The problem is, that the expected relations between the variables would be that capital and low-skilled labour are gross substitutes, that capital and high-skilled labour are gross complements, and that the two labour types are gross substitutes. This would require high-skilled labour and capital together in the nested function, and low-skilled labour in the main function — it would break the model's aggregate human capital expression. In short, this approach does not appear fruitful.

[29] The Constant Elasticity of Substitution function is, as the name suggests, about elasticity. And elasticity relates to the sensitivity of one variable to a change in another variable. The problem in a nutshell here is: if there are more than two variables in the function, then what should elasticity be measured with respect to? Especially where different pairs of inputs have different elasticity relationships, the applicability of the basic multi-input CES functional form is constrained. Nested CES functions offer a path through (Henningsen and Henningsen 2011).

4 Discussion

The preceding section has outlined a model of how choices in the amount of education affect both productivity and inequality. It builds on, and usefully complements, Galor and Moav (2000).

A key question for any model is how well it holds up against the observed facts. Looking at the developed world since the 1950s/60s, productivity growth has slowed from its post-war high and inequality has increased. This model is able to reflect that story: assuming an exogenous shock which temporarily pushes efficiency growth above its long-term (steady state) growth rate, efficiency growth will steadily decline over successive generations, converging back to its initial level. As it does so, the reduction in the erosion effect from the (falling) efficiency growth rate will increase income inequality — the returns to high-income workers will increase.

Where the model falls over, with regard to the observed facts, is human capital accumulation. Continuing the example of an exogenous shock pushing growth above its long-term rate, the model would predict that high ability types would choose *less* education as efficiency growth falls. This is the opposite of what has occurred. This suggests that the dynamics at play are more complicated than what the model reflects: that either (or both) the choice of education is driven by other factors, or that the relationship between human capital and productivity is not as clear cut as the model implies.

It must be acknowledged that the model here does not attempt to tell the full story of how productivity emerges, or explain all the origins of inequality. There are various exogenous parameters that might not be truly exogenous in reality. Ability, for example, is likely affected by inequality: those who are less well off in life may not have access to the same opportunities for education as those with greater resources. Even the idea that ability can be represented on a spectrum from zero to one is, frankly, fanciful.

However, the model provides a base from which further effects can be tested — either by adjusting some of the key functions (as discussed at the end of section 3) or by augmenting the model with additional elements. Some options for the latter are considered below.

4.1 Inclusive growth

At one level, the model is consistent with the idea of 'inclusive growth' discussed earlier in this thesis. The model shows an inverse relationship between productivity growth and inequality: that the income gap between low- and high-skilled workers falls as the efficiency growth rate rises. The erosion effect associated with rising efficiency growth reduces the return to human capital (while simultaneously inducing an increase in education) — but at the same time, the higher efficiency growth rate necessarily implies that the overall level of technical efficiency is rising more quickly, which increases the returns to all forms of labour.

However, at another level, the model is lacking: it is based on the unidirectional effect of productivity on inequality. It does not contain any form of feedback loop — that rising inequality might dampen long-term productivity growth. The obvious question that follows from this is, what might such an effect in the opposite direction be caused by?

One candidate is demography. If one assumes that the population growth rate is constant over time and that the (exogenous) distribution of ability across the population is unchanged from one generation to the next, then neither the size of the population nor its growth rate changes the conclusions presented earlier: the model's results hinge on the relative population shares of each ability type, and these shares do not change over time.

But what if the population growth rate differed by ability type? For simplicity here, I sketch out a scenario based on the two-ability-type model, but the point holds more generally. Suppose that ability is transmitted across generations: that a low-ability parent (in the narrow sense of 'ability' defined in this model) gives birth to a low-ability child.[30] Suppose further that the population growth rate for the low ability type was greater than for the high ability type. If one expects the opportunity cost of raising a child to be greater for high-skilled workers than for low-skilled workers (new parents exit the workforce for a period of time, and forgo their income; high-skilled workers have higher incomes), then it follows that low-skilled workers would have more children on average than high-skilled workers.

This is not wholly implausible: La Croix and Doepke (2003) present a model along the above lines, where there is interaction between fertility and education in the context of income inequality. As income inequality grows, so too does the difference in fertility rates between rich and poor. The average level of human capital falls, as subsequent generations are comprised more heavily of those from poorer households (which do not invest as much in human capital due to a quantity–quality tradeoff[31]).

[30] This abstracts away from the reality that it takes two parents (in the biological sense) to bring a child into the world. And the two parents need not have the same ability type. This wrinkle could be ironed out by assuming (restrictively) that all couples have matching ability levels. A more complicated approach would be to assume that some proportion of the population of couples have differing ability levels, and that their offspring will be of low or high ability type with a probability distribution of β and $1-\beta$ (matching the initial exogenous distribution).

[31] This quantity–quality tradeoff was discussed by Hanushek (1992) and also features in unified growth theory (Galor 2011; 2005). The basic idea here is that as both child-rearing and education are costly, households (given their budget constraint) must choose between having many children, but not educating them (quantity); or having fewer children, but investing in their education (quality). As the returns from education increase, it becomes more attractive to switch from quantity to quality.

Consequently, according to their model, economic growth will be lower where there is greater income inequality.

A second likely channel, following the background discussion on inequality earlier, is political institutions. This model does not take account of any redistributive policies: the tax and transfer system — indeed government more generally — does not feature. One could augment the model with an income redistribution mechanism, which would lessen the income gap between low- and high-skilled income earners (creating a wedge between market income and disposable income). The precise effect would depend on the design of the mechanism, but an anticipated effect might be a reduction in human capital accumulation (as high ability types would be recover less of a return from their education), in turn slowing the pace of efficiency growth. In the context of this model, such an intervention could — counter-intuitively — induce a further increase in inequality.

A third possibility relates to the cost of education. This model only includes the opportunity cost of not earning an income while one is a student; there is no direct financial cost to students of education. If there are student fees, and students have limited scope to borrow against future earnings, then this will constrain who can study. In particular, those individuals from poorer households will be less likely to receive an education (regardless of ability). This would leave human capital potential untapped, thereby weakening productivity growth.

5 Conclusion

My intent with this thesis was to explore the possible relationship between productivity and inequality. Both fields are backed by a wealth of literature, but the links between them are not fully clear. Even after studying the material in depth, I am not sure I have a convincing answer as to what the relationship between the two is. The most I can confidently offer is, 'it depends'.

Productivity is the product of technical changes: new ideas and knowledge that improve how inputs are used in the production of outputs. Inadvertent consequences of natural phenomena aside, technical progress is a result of the actions of economic agents: individuals, firms and governments. People make choices that, at some level, shape productivity. Those choices will in turn have a bearing on the level of inequality. However, it is not implausible that this is a two-way street: that the level of inequality — through its effect on the distribution of resources throughout society — might also influence the choices that shape technical progress over the long term.

If one assumes that a given country's actual productivity is close to (or at) its maximum potential technical efficiency, then, in the short term, any reallocation of resources on equality grounds will — as per Okun — incur a cost in terms of efficiency. (More precisely, given technical efficiency, any reallocation on equality grounds could not result in a higher level of efficiency.) But what happens at a point in time, where technical efficiency is held constant, does not automatically hold over time — technical changes increase productive potential, providing new opportunities both in a growth sense and in an allocative sense. That is, in the long term, productivity growth need not be inconsistent with reducing inequality: Okun's efficiency–equality tradeoff is a short-term constraint, not necessarily a long-term one.

The language here is cautious: the terms 'need not' and 'necessarily' bear a great deak of weight in that caveated statement. There is no ex ante reason to assume that productivity growth must lead to either greater or lower inequality over the long term. Any distributional consequences partly hinge on the source and nature of the technical change. The evidence, as per Acemoglu, would suggest that the direction of technical change is to augment skills (and thus high-income earners).

I introduced this thesis by posing two questions: first, can productivity growth drive long-term economic growth without contributing to long-term inequality growth? And second, can constraining or reducing inequality support long-term economic growth without hampering productivity growth?

The twin questions of this thesis have been approached in different ways. The primary question, of how productivity might influence inequality, has been explored theoretically with reference to existing models

(Acemoglu 2002b; Galor and Moav 2000). In addition, I have presented a model whereby education choice affects both productivity growth and inequality growth over time. The model predicts that as productivity growth rises, income inequality should fall and human capital accumulation should rise. This would be encouraging in the context of inclusive growth; the problem is, this result does not accord with the observed evidence.

The secondary question of how inequality might affect productivity has been explored in less depth. In part, the story of inclusive growth may simply have less relevance in the context of developed economies: the dynamics of productivity and inequality are likely different for developing countries (whether due to credit market imperfections, political and institutional factors, or something else entirely). My larger concern is that it is not immediately clear through which channel inequality might adversely affect productivity growth in developed countries over the long term. The most compelling candidate (in my view) is through demographic effects: that the composition of the population will change over time, if low income earners have on average more children than high income earners. If I were to continue researching this field, this is where I would choose to start.

References

Abramovitz, Moses. 1956. "Resource and Output Trends in the United States since 1870." National Bureau of Economic Research (NBER).

Acemoglu, Daron. 1998. "Why Do New Technologies Complement Skills ? Directed Technical Change and Wage Inequality." *The Quarterly Journal of Economics* 113 (4): 1055–89. https://doi.org/10.1162/003355398555838.

———. 2002a. "Directed Technical Change." *Review of Economic Studies* 69 (4): 781–809.

———. 2002b. "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature* 40 (1): 7–72. https://doi.org/10.1257/jel.40.1.7.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." IZA DP No. 12293. Discussion Paper.

Alchian, Armen A. 1965. "Some Economics of Property Rights." *Il Politico* 30 (4): 816–29.

Alesina, Alberto, and Roberto Perotti. 1996. "Income Distribution, Political Instability, and Investment." *European Economic Review* 40 (6): 1203–28. https://doi.org/10.1016/0014-2921(95)00030-5.

Alesina, Alberto, and Dani Rodrik. 1994. "Distributive Politics and Economic Growth." *The Quarterly Journal of Economics* 109 (2): 465–90. https://doi.org/10.2307/2118470.

Atkinson, A.B., and F. Bourguignon. 2000. "Introduction: Income Distribution and Economics." *Handbook of Income Distribution* 1: 1–58. https://doi.org/10.1016/S1574-0056(00)80003-2.

Barro, Robert J., and Jong-Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics* 104: 184–98.

Berg, Andrew G., and Jonathan D. Ostry. 2017. "Inequality and Unsustainable Growth: Two Sides of the Same Coin?" *IMF Economic Review* 65 (4): 792–815. https://doi.org/10.1057/s41308-017-0030-8.

Berg, Andrew G., and Jeffrey Sachs. 1988. "The Debt Crisis: Structural Explanations of Country Performance." Cambridge, MA. https://doi.org/10.3386/w2607.

Bergeaud, A., G. Cette, and R Lecat. 2016. "Productivity Trends in Advanced Countries between 1890 and 2012." *Review of Income and Wealth* 62 (3): 420–44.

Binswanger, Hans P . 2016. “A Microeconomic Approach to Induced Innovation.” *The Economic Journal* 84 (336): 940–58.

Blinder, Alan S., and Richard E. Quandt. 1997. “The Computer and the Economy.” *The Atlantic*, December 1997.

Carter, Michael R. 2000. “Land Ownership Inequality and the Income Distribution Consequences of Economic Growth.” 295533. WIDER Working Papers.

David, Paul A. 2000. “Understanding Digital Technology’s Evolution and the Path of Measured Productivity Growth: Present and Future in the Mirror of the Past.” In *Understanding the Digital Economy: Data, Tools, and Research*, edited by Erik Brynjolfsson and Brian Kahin. Cambridge: MIT Press.

Diewert, Erwin, and Kevin J . Fox. 1999. “Can Measurement Error Explain the Productivity Paradox?” *The Canadian Journal of Economics (Revue Canadienne d’Economique)* 32 (2): 251–80.

Drandakis, E . M ., and E . S . Phelps. 2016. “A Model of Induced Invention, Growth and Distribution.” *The Economic Journal* 76 (304): 823–40.

Easterly, William, and Ross Levine. 2016. “It’s Not Factor Accumulation: Stylized Facts and Growth Models.” *The World Bank Economic Review* 15 (2): 177–219.

Galor, Oded. 2005. “From Stagnation to Growth: Unified Growth Theory.” *Handbook of Economic Growth* 1 (SUPPL. PART A): 171–293. https://doi.org/10.1016/S1574-0684(05)01004-X.

———. 2011. *Unified Growth Theory*. Princeton University Press. https://doi.org/10.2307/j.ctvcm4h7m.8.

Galor, Oded, and Omer Moav. 2000. “Ability-Biased Technological Transition, Wage Inequality, and Economy Growth.” *The Quarterly Journal of Economics*, no. May: 469–97. https://academic.oup.com/qje/article-abstract/115/2/469/1840455.

Gini, Corrado. 1921. “Measurement of Inequality of Incomes.” *The Economic Journal* 31 (121): 124. https://doi.org/10.2307/2223319.

Gordon, Jenny, Shiji Zhao, and Paul Gretton. 2015. “On Productivity: Concepts and Measurement.” Australian Government Productivity Commission. Canberra.

Gordon, Robert J. 2012. “Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds.” 18315. NBER Working Paper. Washington DC. http://www.nber.org/papers/w18315.

———. 2014. “The Turtle’s Progress: Secular Stagnation Meets the Headwinds.” In *Secular Stagnation:*

Facts, Causes, and Cures, edited by Coen Teulings and Richard Baldwin, 47–60. Centre for Economic Policy Research.

Griliches, Zvi. 1979. “Issues in Assessing the Contribution of Research and Development to Productivity Growth.” *The Bell Journal of Economics* 10 (1): 92–116. https://about.jstor.org/terms.

Grubb, Michael, Jean-Charles Hourcade, and Karsten Neuhoff. 2014. *Planetary Economics Energy, Climate Change and the Three Domains of Sustainable Development*. Routledge.

Haggard, Stephan, and Lydia Tiede. 2011. “The Rule of Law and Economic Growth: Where Are We?” *World Development* 39 (5): 673–85. https://doi.org/10.1016/j.worlddev.2010.10.007.

Hanushek, Eric A. 1992. “The Trade-off between Child Quantity.” *Journal of Political Economy*. Vol. 100.

Henningsen, Arne, and Géraldine Henningsen. 2011. “Econometric Estimation of the ‘Constant Elasticity of Substitution’ Function in R: Package MicEconCES.” 9. FOI Working Paper. Copenhagen. www.foi.life.ku.dk.

Hicks, J. R. 1963. *The Theory of Wages*. https://doi.org/10.1007/978-1-349-00189-7.

IMF. 2019. “World Economic Outlook: Global Manufacturing Downturn, Rising Trade Barriers.” Washington DC. https://www.imf.org/~/media/Files/Publications/WEO/2019/October/English/text.ashx?la=en.

Immervoll, Herwig, and Linda Richardson. 2011. “Redistribution Policy and Inequality Reduction in OECD Countries: What Has Changed in Two Decades?” IZA DP 6030. Discussion Paper.

Islam, Nazrul. 2008. “Determinants of Productivity across Countries: An Exploratory Analysis.” *The Journal of Developing Areas* 42 (1): 201–42.

Kennedy, Charles. 1964. “Induced Bias in Innovation and the Theory of Distribution.” *The Economic Journal* 74 (295): 541–47.

Kireyev, Alexei, and Jingyang Chen. 2017. “Inclusive Growth Framework.” WP/17/127. IMF Working Papers.

Krugman, Paul. 1994. *The Age of Diminished Expectations: U.S. Economic Policy in the 1990s*. MIT Press.

La Croix, David De, and Matthias Doepke. 2003. “American Economic Association Inequality and Growth: Why Differential Fertility Matters.” *The American Economic Review*. Vol. 93.

Lorenz, M. O. 1905. “Methods of Measuring the Concentration of Wealth.” *Publications of the American Statistical Association* 9 (70): 209. https://doi.org/10.2307/2276207.

Mankiw, Gregory, David N. Weil, and David Romer. 1992. “A Contribution to the Empirics of Economic Growth.” *The Quarterly Journal of Economics* 107 (2): 407–37. https://doi.org/10.2307/2118477.

Mincer, Jacob. 1974. “Progress in Human Capital Analyses of the Distribution of Earnings.” 53. Working Paper.

———. 1981. “Human Capital and Economic Growth.” Cambridge, MA. https://doi.org/10.3386/w0803.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195074772.001.0001.

———. 2014. “Secular Stagnation? Not in Your Life.” In *Secular Stagnation: Facts, Causes, and Cures*, edited by Coen Teulings and Richard Baldwin, 83–90. Centre for Economic Policy Research.

Ng, Yew-Kwang. 2004. *Welfare Economics: Towards a More Complete Analysis*. 1st ed. Hampshire: Palgrave Macmillan.

OECD. 2001. “Measuring Productivity - OECD Manual.” *Measuring Productivity - OECD Manual*. https://doi.org/10.1787/9789264194519-en.

———. 2018. “The Framework for Policy Action on Inclusive Growth.”

Okun, Arthur M., and Lawrence Summers. 2015. *Equality and Efficiency: The Big Tradeoff*. https://doi.org/10.2307/1978438.

Parham, Dean. 2012. “Australia’s Productivity Growth Slump: Signs of Crisis, Adjustment or Both?” Visiting Researcher Paper. Canberra. www.pc.gov.au.

Piketty, Thomas. 2000. “Theories of Persistent Inequality and Intergenerational Mobility.” In *Handbook of Income Distribution*, 1:429–76. Elsevier. https://doi.org/10.1016/S1574-0056(00)80011-1.

Romer, Paul M. 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98 (5).

Sandmo, Agnar. 2015. “The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought.” In *Handbook of Income Distribution*, 2:3–65. Elsevier. https://doi.org/10.1016/B978-0-444-59428-0.00002-3.

Sen, Amartya. 2000. “Social Justice and the Distribution of Income.” *Handbook of Income Distribution*. https://doi.org/10.1016/S1574-0056(00)80004-4.

Solow, Robert M. 1956. “A Contribution to the Theory of Economic Growth.” *Source: The Quarterly Journal of Economics*. Vol. 70.

———. 1957. “Technical Change and the Aggregate Production Function.” *The Review of Economics and Statistics* 39 (3): 312–20.

———. 1987. “We’d Better Watch Out.” *New York Times Book Review*, no. 12 July: 36.

———. 2001. “Applying Growth Theory across Countries.” *The World Bank Economic Review* 15 (2): 283–88. https://www.jstor.org/stable/3990266.

Solt, Frederick. 2019. “The Standardized World Income Inequality Database, Version 8.” Harvard Dataverse, V4.

Swan, T. W. 1956. “Economic Growth and Capital Accumulation.” *Economic Record* 32 (2): 334–61. https://doi.org/10.1111/j.1475-4932.1956.tb00434.x.

WEF. 2018. “The Inclusive Development Index 2018.” Geneva. http://fsolt.org/swiid/.

World Bank. 2014. “An Overview of the World Bank Group Strategy.” Washington DC.