Reconciling near trend-stationary growth with medium-frequency cycles.

Tom Holden[1], School of Economics, University of Surrey

Abstract: *Existing models of dynamic endogenous growth generate implausibly large trend breaks in output when augmented with standard business cycle shocks. This paper presents a model without this deficiency, yet still capable of generating large medium-frequency fluctuations around the trend. Ensuring the robustness of the trend requires that we eliminate the scale effects and knife edge assumptions that plague most growth models. In our model, medium-frequency fluctuations arise from changes in the proportion of industries producing patent protected products. However, variations in the number of firms within each industry ensure that process improvement incentives remain roughly constant.*

Keywords: *medium frequency cycles, patent protection, scale effects*

JEL Classification: *E32, E37, L16, O31, O33, O34*

[1] Contact address: School of Economics, FBEL, University of Surrey, Guildford, GU2 7XH, England
Contact email: thomas.holden@gmail.comWeb-site: http://www.tholden.org/
The author would particularly like to thank his supervisor, Simon Wren-Lewis. Further thanks are due to Guido Ascari, Eric Bartelsman, Paul Beaudry, Pedro Bento, George Bitsakakis, Beatriz de Blas, Cristiano Cantore, Diego Comin, Ana-Maria Dumitru, Martin Ellison, Fabio Ghironi, Bart Hobijn, Tim Kehoe, Tobias Kitlinski, Paul Levine, Omar Licandro, Ioana Moldovan, Peter Neary, Joaquim Oliveira Martins, Christopher Nekarda, Walter Park, Pietro Peretto, James Reade, Torsten Schmidt, Diego Restuccia, Nicole Tabasso, David Vines, Klaus Wälde and the participants in numerous seminars at which versions of this paper were presented. Financial support provided to the author by the ESRC is also greatly appreciated. The author gratefully acknowledges the use of the Oxford e-Research Centre's Windows Compute Cluster and the University of Surrey FBEL cluster for the numerical computations performed in this paper, and the use of data from the Nuffield College Data Library.

Viewed from a distance, a log-plot of the last one hundred years of US GDP looks very near linear. However, closer inspection reveals large medium frequency fluctuations around this linear trend. Generating this combination of remarkably near trend-stationary long run growth, and large cycles around the trend, is a challenge for traditional models of endogenous growth. The near linear trend requires scale effects to be removed not just in the long run, but in the short run as well. Models that remove these scale effects via knife-edge assumptions will usually fail this test, as temporary business cycle shocks will knock the model away from perfectly removing the scale effect, leading to a permanent break in the trend of the GDP. Equally, models that remove scale effects via new product creation will tend to produce such trend breaks in GDP if the stock of new products can only respond slowly following a shock. On the other hand, if the stock of products can adjust instantly following a shock, then, (in standard models) there would be no movement in productivity at all, let alone the large, persistent medium frequency cycles that Comin and Gertler (2006) document in the data. In this paper, we present a mechanism capable of reconciling this apparently contradictory low and medium frequency behaviour of output, while also matching the cyclicality of mark-ups: the key determinant of research decisions.

Our story is as follows. The returns to inventing a new product are higher in a boom due to the higher demand. As a result, during periods of expansion, the rate of creation of new products increases, in line with the evidence of Broda and Weinstein (2010). Due to a first mover advantage, patent protection, or reverse-engineering difficulties, the inventors of these new products will be able to extract rents from them, increasing the costs manufacturing firms face if they wish to produce the new product. These higher costs lead to lower competition in new industries, increasing mark-ups and thus increasing firms' incentives to perform the R&D necessary to catch-up with and surpass the frontier, for basically Schumpeterian reasons. Consequently, the higher proportion of industries that are relatively new in a boom will lead to higher aggregate productivity, lower dispersion of both productivity levels and growth rates, as well as higher mark-ups. Since the length of time for which inventors can extract rents will be determined by the effective duration of patent-protection, this effect will naturally work at medium frequencies. However, since we allow both for the creation of new industries (producing new products) and for varying numbers of firms within each industry, even in the short-run the demand faced by any given firm will be roughly constant, meaning that our model will not produce large deviations from linear growth.

Formal evidence on the small size of the unit root in output (i.e. its near trend stationarity) was presented by Cochrane (1988), and in our companion article (Holden 2013a) we present further evidence that GDP returns to trend at long lags (at least eight years after the initial shock). Evidence for the pro-cyclicality of TFP has been presented by Bils (1998) and Campbell (1998) amongst others, with Comin and Gertler (2006) showing that the evidence is particularly clear at medium-frequencies. The counter-cyclicality of productivity dispersion has been shown by Kehrig (2011), with evidence on the counter-cyclicality of the dispersion of productivity growth rates provided by e.g. Eisfeldt and Rampini (2006) and Bachmann and Bayer (2009). Evidence for the pro-cyclicality of aggregate mark-ups has been presented by Boulhol (2007) and Nekarda and Ramey (2010). Nekarda and Ramey also show that mark-ups lead output at business-cycle frequencies. In a companion article (Holden 2013a), we present further evidence that this relationship continues to hold at medium-frequencies, with mark-ups being pro-cyclical providing the data is filtered with a cut-off below sixteen years. Boulhol (2007) also shows that although aggregate mark-ups are pro-cyclical, the mark-ups in any particular industry tend to be counter-cyclical. This apparent contradiction is readily explained by our model, as the increase in competition in any particular industry will lead to a decline in mark-ups in that industry (much as in the models of Bilbiie, Ghironi, and Melitz (2012) and Jaimovich (2007)), despite the fact that aggregate mark-ups have increased due to the greater proportion of industries with relatively high mark-ups.

Direct evidence for the importance of our mechanism comes from a number of sources. Balasubramanian and Sivadasan (2011) find that firms holding patents have 17% higher TFP levels on average, and additionally find that firms that go from not holding a patent to holding one experience a 7.4% increase in a fixed effects measure of productivity, suggesting that industries producing patent-protected products are indeed significantly more productive. Serrano (2007) finds that although aggregate patenting is only weakly correlated with aggregate TFP, a measure of the number of patents whose ownership is transferred is strongly related to productivity. He argues that there is a great deal of noise in measures of total patent activity, since so many patents are never seriously commercialised. Patent transfers are usually observed though when their purchaser intends to begin exactly such a commercialisation. Thus, patent transfers provide a proxy for the commencement of production of new patented-products, one that is found to be highly pro-cyclical. Finally, in our companion article (Holden 2013a) we present new evidence that longer patent protection significantly increases the share of GDP variance attributable to cycles of medium frequency, suggesting that patent protection plays an important role in the mechanism generating medium frequency cycles in reality.

Previous papers have introduced endogenous productivity improvement into business cycle models (e.g. Comin and Gertler (2006), Comin (2009), Comin, Gertler, and Santacreu (2009), Phillips and Wrase (2006), Nuño (2008; 2009; 2011)), or looked at cycles in growth models (e.g. Bental and Peled (1996), Matsuyama (1999), Wälde (2005), Francois and Lloyd-Ellis (2008; 2009), Comin and Mulani (2009)). However, all of these papers have problems with scale effects, either in the long run, or in the short run, and thus all of them would predict counter-factually large unit roots in output in the presence of standard business cycle shocks. Furthermore, it is not obvious how these scale effects could be removed without destroying the papers' mechanisms for generating aggregate TFP movements. For example, the papers of Wälde (2005) and Phillips and Wrase (2006) rely on there being a small finite number of sectors. Removing the scale effect would mean allowing this number to grow over time with population, meaning the variance of productivity would rapidly go to zero. Indeed, this happens endogenously in the model of Horii (2011). Many models of endogenous mark-up determination (e.g. Bilbiie, Ghironi, and Melitz (2012) or Jaimovich (2007)) have a similar problem, with the presence of a small finite number of industries being crucial for explaining the observed variance of mark-ups. Indeed, Bilbiie, Ghironi, and Melitz (2011) write that "reconciling an endogenous time-varying markup with stylized growth facts (that imply constant markups and profit shares in the long run) is a challenge to growth theory". By disentangling the margins of firm entry and product creation, we will be able to answer this challenge.

The paper of most relevance to our work is that of Comin and Gertler (2006), who made the important contribution of bringing the significance of medium-frequency cycles to the attention of the profession. Additionally, their theoretical model, like ours, stresses the links between mark-up variation and productivity growth. Unfortunately, however, it is a model with strong scale effects removed via a series of knife-edge assumptions, [2] with the inevitable consequence that the driving mark-up shock produces a counter-factual trend break in productivity. Furthermore, beyond this trend break, the model generates little endogenous persistence, and also counter-factually predict that increases in mark-ups lead to falls in output, contrary to the empirical evidence of Nekarda and Ramey (2010).[3] While there is room to disagree with the recent empirical work on the cyclicality of output, the fact that mark-up increases lead output increases is much more robustly established.[4] We conclude that the literature still lacks a model of productivity capable of explaining both its short run and its long run behaviour.

In this paper, we present a model capable of doing exactly this. In order to remove both the long run and the short run scale effect, as discussed above it

[2] The social value of the aggregate capital stock enters in multiple places without exponent, in order to capture the idea that "operating costs are proportional to the sophistication of the economy". Had (say) frontier technology been used in its place, the generated impulse responses would likely have been quite different, and had it entered with a non-unit exponent then the model would not have possessed a balanced growth path.
[3] Care must be taken to match measures of the aggregate mark-up. If we measure the aggregate mark-up by the inverse labour share, then holding labour supply and output constant, an increase in wage mark-ups decreases aggregate mark-ups. However, at reasonable calibrations of the Comin and Gertler (2006) model, an increase in wage mark-ups results in such a drop in labour supply that the inverse labour share increases.
[4] See footnote 8 of our companion paper (Holden 2013a) for a discussion of the evidence on whether this correlation is causal.

will feature a varying number of industries, each of which will contain a varying number of firms. We do not wish to make any exogenous assumptions on the differences between industries producing patented products versus those producing unpatented ones, so in order to match the medium-frequency behaviour of productivity and mark-ups it is important that our model allow endogenous variation in these quantities across industries. Were we to assume free transfer of technologies across industries there would be too little difference in productivity between patent-protected and un-patent-protected industries, and hence we would not be able to generate medium-frequency cycles. Equally, were we to assume technology transfer across industries was impossible, then it would be legitimate to inquire whether the difference between these industry types was implausibly large, as perhaps firms in non-protected industries would find it optimal to perform technology transfer even if they did not find it optimal to perform any research. Consequently, in modelling the endogenous productivity in each industry we will allow firms both to perform research, and to perform a costly process of catch-up to the frontier we shall term appropriation. To make clear the strength of the amplification and persistence mechanism presented here, we omit capital from the model, and we focus on the impulse responses to non-persistent shocks when we discuss our model's qualitative behaviour in section 2. In the companion article (Holden 2013a), we go on to embed our mechanism within a fully featured medium-scale model.

1. The model

Our model has a continuum of narrow industries, each of which contains finitely many firms producing a unique product. The measure of industries is increased by the invention of new products, which start their life patent-protected. However, we assume that product inventors lack the necessary human capital to produce their product at scale themselves, and so they must licence out their patent to manufacturing firms. The duration of patent-protection is given by a geometric distribution, in line with Serrano's (2010) evidence on the large proportion of patents that are allowed to expire early, perhaps because they are challenged in court or perhaps because another new product is a close substitute. An earlier working-paper version of this model (Holden 2011) considered the fixed duration case, which is somewhat less tractable. Allowing for a distribution of protection lengths also allows us to give a broader interpretation to protection within our model. Even in the absence of patent protection, the combination of contractual agreements such as NDAs, and difficulties in reverse engineering, is likely to enable the inventor of a new product to extract rents for some time.

Our model of endogenous competition within each industry is derived from Jaimovich (2007). We chose the Jaimovich model as it is a small departure from the standard Dixit-Stiglitz (1977) set-up, and leads to some particularly neat expressions. Similar results could be attained with Cournot competition, or the translog form advocated by Bilbiie, Ghironi, and Melitz (2012). One important departure from the Jaimovich model is that in our model, entry decisions take place one period in advance. This is natural as we wish to model research as taking place after entry but before production. Productivity within a firm is increased by performing research or appropriation. We regard process research as incremental, with regular small changes rather than the unpredictable jumps found in Schumpetarian models (Aghion and Howitt 1992; Wälde 2005; Phillips and Wrase 2006).

Throughout, we assume that only products are patentable,[5] and so by exerting effort firms are able to "appropriate" process innovations from other industries to aid in the production of their own product. This appropriation is costly since technologies for producing other products will not be directly applicable to producing a firm's own product. We assume that technology transfer within an industry is costless however, due to intra-industry labour flows and the fact that

[5] This is broadly in line with the law in most developed countries: ideas that are not embedded in a product (i.e. a machine) generally have at most limited patentability. In the U.S., the most recent Supreme court decision found that the following was "a useful and important clue" to the patentability of processes (*Bilski v. Kappos*, 561 U.S. ____ (2010)): "a method claim is surely patentable subject matter if (1) it is tied to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing" (*In re Bilski*, 545 F.3d 943, 88 U.S.P.Q.2d 1385 (Fed. Cir. 2008)). This "machine or transformation" test was widely believed at the time to have ended the patentability of business processes (The Associated Press 2008), and this position was only slightly softened by *Bilski v. Kappos*.

all firms in an industry are producing the same product. This is important for preserving the tractability of the model, as it means that without loss of generality we may think of all firms as just existing for two periods, in the first of which they enter and perform research, and in the second of which they produce.

The broad timing of our model is as follows. At the beginning of period t invention takes place, creating new industries. All holders of current patents (including these new inventors) then decide what level of licence fee to charge. Then, based on these licence fees and the level of overhead costs, firms choose whether to enter each industry. Next, firms perform appropriation, raising their next-period productivity towards that of the frontier, then research, further improving their productivity next period. In period $t+1$, they then produce using their newly improved production process. Meanwhile, a new batch of firms will be starting this cycle again.

We now give the detailed structure of the model.

1.1. Households

There is a unit mass of households, each of which contains N_t members in period t. The representative household maximises:

$$\mathbb{E}_t \sum_{s=0}^{\infty} \beta^s N_{t+s} \left[\log \frac{C_{t+s}}{N_{t+s}} - \frac{1}{1+\nu} \left(\frac{L_{t+s}^{\mathrm{S}}}{N_{t+s}} \right)^{1+\nu} \right]$$

where C_t is aggregate period t consumption, L_t^{S} is aggregate period t labour supply, β is the discount rate and ν is the inverse of the Frisch elasticity of labour supply to wages, subject to the aggregate budget constraint that $C_t + B_t = L_t^{\mathrm{S}} W_t + B_{t-1} R_{t-1} + \Pi_t$, where B_t is the aggregate number of (zero net supply) bonds bought by households in period t, W_t is the period t wage, R_{t-1} is the period t sale price of a (unit cost) bond bought in period $t-1$, and Π_t is the households' period t dividend income.

1.2. Aggregators

The consumption good is produced by a perfectly competitive industry from the aggregated output $Y_t(i)$ of each industry $i \in [0, I_{t-1}]$, using the following Dixit-Stiglitz-Ethier (Dixit and Stiglitz 1977; Ethier 1982) style technology:

$$Y_t = I_{t-1}^{-\lambda} \left[\int_0^{I_{t-1}} Y_t(i)^{\frac{1}{1+\lambda}} \, di \right]^{1+\lambda}$$

where $\frac{1+\lambda}{\lambda}$ is the elasticity of substitution between goods and where the exponent on the measure of industries (I_{t-1})[6] has been chosen to remove any preference for variety in consumption.[7] We normalize the price of the aggregate good to 1.

Similarly, each industry aggregate good $Y_t(i)$ is produced by a perfectly competitive industry from the intermediate goods $Y_t(i,j)$ for $j \in \{1, \dots, J_{t-1}(i)\}$,[8] using the technology:

$$Y_t(i) = J_{t-1}(i)^{-\eta\lambda} \left[\sum_{j=1}^{J_{t-1}(i)} Y_t(i,j)^{\frac{1}{1+\eta\lambda}} \right]^{1+\eta\lambda}$$

where $\eta \in (0,1)$ controls the degree of differentiation between firms, relative to that between industries.

1.3. Intermediate firms

1.3.1. Pricing

Firm j in industry i has access to the linear production technology $Y_t(i,j) = A_t(i,j) L_t^{\mathrm{P}}(i,j)$ for production in period t. As in Jaimovich (2007), strategic profit maximisation then implies that in a symmetric equilibrium, the price of the good in industry i is given by $P_t(i) = \left(1 + \mu_{t-1}(i)\right) \frac{W_t}{A_t(i,j)} = \left(1 + \mu_{t-1}(i)\right) \frac{W_t}{A_t(i)}$, where $\mu_t(i) := \lambda \frac{\eta J_t(i)}{J_t(i) - (1-\eta)} \in (\eta\lambda, \lambda]$ is the industry i mark-up in period $t+1$ and $A_t(i) = A_t(i,j)$ is the productivity shared by all firms in industry i in symmetric equilibrium. From aggregating across industries, we then have that $W_t = \frac{A_t}{1+\mu_{t-1}}$

[6] The $t-1$ subscript here reflects the fact that industries are invented one period before their product is available to consumers.

[7] Incorporating a preference for variety would not change the long-run stability of our model.

[8] Again, the $t-1$ subscript reflects the fact that firms enter one period before production.

where $\frac{1}{1+\mu_t} = \left[\frac{1}{I_t}\int_0^{I_t}\left[\frac{1}{1+\mu_t(i)}\right]^{\frac{1}{\lambda}} di\right]^{\lambda}$ determines the aggregate mark-up μ_{t-1} and where:

$$A_t := \frac{\left[\frac{1}{I_{t-1}}\int_0^{I_{t-1}}\left[\frac{A_t(i)}{1+\mu_{t-1}(i)}\right]^{\frac{1}{\lambda}} di\right]^{\lambda}}{\left[\frac{1}{I_{t-1}}\int_0^{I_{t-1}}\left[\frac{1}{1+\mu_{t-1}(i)}\right]^{\frac{1}{\lambda}} di\right]^{\lambda}}$$

is a measure of the aggregate productivity level.[9]

1.3.2. Sunk costs: rents, appropriation and research

Following Jaimovich (2007), we assume that the number of firms in an industry is pinned down by the zero profit condition that equates pre-production costs to production period profits. Firms borrow in order to cover these upfront costs, which come from four sources.

Firstly, firms must pay a fixed operating cost L^{F} that covers things such as bureaucracy, human resources, facility maintenance, training, advertising, shop set-up and capital installation/creation. Asymptotically, the level of fixed costs will not matter, but including it here will help in our explanation of the importance of patent protection for long run growth.

Secondly, if the product produced by industry i is currently patent-protected, then firms must pay a rent of $\mathcal{R}_t(i)$ units of the consumption good to the patent-holder for the right to produce in their industry. Since all other sunk costs are paid to labour, for convenience we define $L_t^{\mathcal{R}}(i) := \frac{\mathcal{R}_t(i)}{W_t}$, i.e. the labour amount equivalent in cost to the rent.

Thirdly, firms will expand labour effort on appropriating the previous process innovations of the leading industry. We define the level of the leading technology within industry i by $A_t^*(i) := \max_{j\in\{1,\dots,J_{t-1}(i)\}} A_t(i,j)$ and the level of the best technology anywhere by $A_t^* := \sup_{i\in[0,I_{t-1}]} A_t^*(i)$. Due to free in-industry transfer, even without exerting any appropriation effort, firms in industry i may start their research from $A_t^*(i)$ in period t. By employing appropriation workers, a firm may raise this level towards A_t^*.

[9] Due to the non-linear aggregation, it will not generically be the case that aggregate output is aggregate labour input times A_t. However, the aggregation chosen here is the unique one under which aggregate mark-ups are known one period in advance, as industry mark-ups are.

We write $A_t^{**}(i,j)$ for the base from which firm $j \in \{1, \dots, J_t(i)\}$ will start research in period t, and we assume that if firm j employs $L_t^{\mathrm{A}}(i,j)$ units of appropriation labour in period t then:

$$A_t^{**}(i,j) = \left[A_t^*(i)^\tau + (A_t^{*\tau} - A_t^*(i)^\tau)\frac{A_t^*(i)^{-\zeta^{\mathrm{A}}}\Upsilon L_t^{\mathrm{A}}(i,j)}{1 + A_t^*(i)^{-\zeta^{\mathrm{A}}}\Upsilon L_t^{\mathrm{A}}(i,j)}\right]^{\frac{1}{\tau}}, \qquad (1.1)$$

where Υ is the productivity of appropriation labour, $\zeta^{\mathrm{A}} > 0$ controls the extent to which appropriation is getting harder over time (due, for example, to the increased complexity of later technologies) and where $\tau > 0$ controls whether the catch-up amount is a proportion of the technology difference in levels ($\tau = 1$), log-levels ($\tau = 0$) or anything in between or beyond. This specification captures the key idea that the further a firm is behind the frontier, the more productive will be appropriation. Allowing for appropriation (and research, and invention) to get harder over time is both realistic, and essential for the tractability of our model, since it will lead our model to have a finite dimensional state vector asymptotically, despite all the heterogeneity across industries.

Fourthly and finally, firms will employ labour in research. If firm $j \in \{1, \dots, J_t(i)\}$ employs $L_t^{\mathrm{R}}(i,j)$ units of research labour in period t, its productivity level in period $t+1$ will be given by:

$$A_{t+1}(i,j) = A_t^{**}(i,j)\left(1 + \gamma Z_{t+1}(i,j) A_t^{**}(i,j)^{-\zeta^{\mathrm{R}}} \Psi L_t^{\mathrm{R}}(i,j)\right)^{\frac{1}{\gamma}},$$

where Ψ is the productivity of research labour, $\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$ controls the extent to which research is getting harder over time, $Z_{t+1}(i,j) > 0$ is a shock representing the luck component of research, and $\gamma > 0$ controls the "parallelizability" of research. [10] If $\gamma = 1$, research may be perfectly parallelized, so arbitrarily large quantities may be performed within a given period without loss of productivity, but if γ is large, then, in line with the evidence of Siliverstovs and Kancs (2012), the returns to research decline as the firm attempts to pack more into one period. The restriction that $\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$ means that the difficulty of research is increasing over time faster than the difficulty of appropriation. This is made because research is very much specific to the industry in which it is being conducted,

[10] Peretto (1999) also looks at research that drives incremental improvements in productivity, and chooses a similar specification. The particular one used here is inspired by Groth, Koch, and Steger (2009).

whereas appropriation is a similar task across all industries attempting to appropriate the same technology, and hence is more likely to have been standardised, or to benefit from other positive spillovers.

In the following, we will assume that $Z_t(i,j) \coloneqq Z_t$ so that all firms in all industries receive the same "idea" shock. We make this assumption chiefly for simplicity, but it may be justified by appeal to common inputs to private research, such as university research output or the availability of new tools, or by appeal to in-period labour market movements carrying ideas with them. We will see in the following that allowing for industry-specific shocks has minimal impact on our results, providing there are at least correlations across industries (plausible if they are producing similar products). For concreteness, we assume that $Z_t \coloneqq \exp(\sigma_Z \epsilon_{Z,t})$, where $\sigma_Z > 0$ and $\epsilon_{Z,t}$~NIIID(0,1).

1.3.3. Research and appropriation effort decisions

Firms are owned by households and so they choose research and appropriation to maximize:

$$\beta \mathbb{E}_t \left[\Xi_{t+1} \left(P_t(i,j) - \frac{W_{t+1}}{A_{t+1}(i,j)} \right) Y_t(i,j) \right] \\ - [L_t^R(i,j) + L_t^A(i,j) + L_t^{\mathcal{R}}(i) + L^F] W_t,$$

where $\Xi_t = \frac{N_t C_{t-1}}{N_{t-1} C_t}$. It may be shown that, for firms in frontier industries (those for which $A_t^*(i) = A_t^*$), if an equilibrium exists, then it is unique and symmetric within an industry; but we cannot rule out the possibility of asymmetric equilibria more generally.[11] However, since the coordination requirements of asymmetric equilibria render them somewhat implausible, we restrict ourselves to the unique equilibrium in which all firms within an industry choose the same

[11] The equilibrium concept we use is that of pure-strategy subgame-perfect *local* Nash equilibria (SP*L*NE) (i.e. only profitable local deviations are ruled out). We have no reason to believe the equilibrium we find is not in fact a subgame-perfect Nash equilibria (SPNE). Indeed, if there is a pure-strategy symmetric SPNE then it will be identical to the unique pure-strategy symmetric SP*L*NE that we find. Furthermore, our numerical investigations suggest that at least in steady-state, at our calibrated parameters, the equilibrium we describe is indeed an SPNE. (Code available on request.) However, due to the analytic intractability of the second stage pricing game when productivities are asymmetric, we cannot guarantee that it remains an equilibrium away from the steady-state, or for other possible calibrations. However, SPLNE's are independently plausible since they only require firms to know the demand curve they face in the local vicinity of an equilibrium, which reduces the riskiness of the experimentation they must perform to find this demand curve (Bonanno 1988). It is arguable that the coordination required to sustain asymmetric equilibria and the computational demands of mixed strategy equilibria render either of these less plausible than our SPLNE.

levels of research and appropriation. Let us then define effective research performed by firms in industry i by $\mathcal{L}_t^{\mathrm{R}}(i) \coloneqq A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi L_t^{\mathrm{R}}(i,j)$ (valid for any $j \in \{1, \dots, J_{t-1}(i)\}$) and effective appropriation performed by firms in that industry by $\mathcal{L}_t^{\mathrm{A}}(i) \coloneqq A_t^{*}(i)^{-\zeta^{\mathrm{A}}}\Upsilon L_t^{\mathrm{A}}(i,j)$ (again, valid for any $j \in \{1, \dots, J_{t-1}(i)\}$).

Providing $\frac{1}{\mu_t(i)} < \min\{\gamma, \tau\}$, $\gamma > \zeta^{\mathrm{R}}$ and $\lambda < 1$ (for the second order conditions [12] and for uniqueness), combining the first order and free entry conditions then gives us that, in the limit as $\sigma_Z \to 0$:[13]

$$\mathcal{L}_t^{\mathrm{R}}(i) = \max\left\{0, \frac{d_t(i) A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi\left(L_t^{\mathrm{A}}(i,j) + L_t^{\mathcal{R}}(i) + L_t^{\mathrm{F}}\right) - \mu_t(i)}{\gamma\mu_t(i) - d_t(i)}\right\} \tag{1.2}$$

and:

$$\mathcal{L}_t^{\mathrm{A}}(i) = \max\left\{0, f_t(i) + \sqrt{\max\{0, f_t(i)^2 + g_t(i)\}}\right\}, \tag{1.3}$$

where $d_t(i) \in (0,1)$[14] is small when firm behaviour is highly distorted by firms' incentives to deviate from choosing the same price as the other firms in their industry, off the equilibrium path (so $d_t(i) \to 1$ as $J_t(i) \to \infty$), and $f_t(i)$ and $g_t(i)$ are increasing in an industry's distance from the frontier,[15] as the further behind a firm is, the greater are the returns to appropriation.

Equations (1.2) and (1.3) mean that research and appropriation levels are increasing in the other sunk costs a firm must pay prior to production, but decreasing in mark-ups. They also mean that the strategic distortions caused by there being a small number of firms within an industry tend to reduce research and appropriation levels. Other sunk costs matter for research levels because when other sunk costs are high, entry into the industry is lower, meaning that

---

[12] The second order condition for research may be derived most readily by noting that when $d_t(i) \to 1$, (i.e. $J_t(i) \to \infty$) the first order condition for research is identical to the one that would have been derived had there been a continuum of firms in each industry with exogenous elasticity of substitution $\frac{1+\mu_t(i)}{\mu_t(i)}$. That it holds more generally follows by continuity. Since $A_t^{**}(i,j)$ is bounded above, no matter how much appropriation is performed the highest solution of the appropriation first order condition must be at least a local maximum.

[13] The first order and zero profit conditions are reported in the online appendix (Holden 2013b sec. 1), where we also derive these solutions. We do not assume $\sigma_Z = 0$ when simulating, but it leads here to expressions that are easier to interpret.

[14] Defined in the online appendix (Holden 2013b sec. 1).

[15] $f_t(i) \coloneqq \frac{1}{2}\left[1 + \frac{d_t(i)}{\tau\mu_t(i)}\frac{1+(\gamma-\zeta^{\mathrm{R}})\mathcal{L}_t^{\mathrm{R}}(i)}{1+\gamma\mathcal{L}_t^{\mathrm{R}}(i)}\right]\left[1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}\right] - 1$, $g_t(i) \coloneqq \frac{d_t(i)}{\tau\mu_t(i)}\frac{1+(\gamma-\zeta^{\mathrm{R}})\mathcal{L}_t^{\mathrm{R}}(i)}{1+\gamma\mathcal{L}_t^{\mathrm{R}}(i)} A_t^*(i)^{-\zeta^{\mathrm{A}}}\Upsilon[L_t^{\mathrm{R}}(i) + L_t^{\mathcal{R}}(i) + L^{\mathrm{F}}]\left[1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}\right] - \left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}$.

each firm receives a greater slice of production-period profits, and so has correspondingly amplified research incentives.

Why mark-up increases decrease research incentives is clearest when those mark-up increases are driven by exogenous decreases in the elasticity of substitution. When products are close substitutes, then by performing research (and cutting its price) a firm may significantly expand its market-share, something that will not happen when the firm's good is a poor substitute for its rivals. When $d_t(i) \approx 1$ (i.e. there are a lot of firms in the industry) firms act as if they faced an exogenous elasticity of substitution $\frac{1+\mu_t(i)}{\mu_t(i)}$, and so when mark-ups are high they will want to perform little research. When $d_t(i)$ is small (i.e. there are only a few firms[16]) then firms' behaviour is distorted by strategic considerations. Each firm realises that if they perform extra research today then their competitors will accept lower mark-ups the next period. This reduces the extent to which research allows market-share expansion, depressing research incentives.

The key thing to note about (1.2) and (1.3) is that research and appropriation are independent of the level of demand, except insomuch as demand affects mark-ups or the level of the strategic distortion. This is because when demand is high there is greater entry, so each firm still faces roughly the same demand. This is essential for removing the short-run scale effect.

In industries that are no longer patent-protected, rents will be zero (i.e. $L_t^{\mathcal{R}}(i) \equiv 0$). Since research is getting harder at a faster rate than appropriation ($\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$), at least asymptotically, no research will be performed in these industries. This is because $A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi[L_t^{\mathrm{A}}(i) + L^{\mathrm{F}}] - \mu_t(i)$ is asymptotically negative since $\mu_t(i) \in (\eta\lambda, \lambda]$. For growth to continue forever in the absence of patent protection, we would require that the overhead cost (L^{F}) was growing over time at exactly the right rate to offset the increasing difficulty of research.

[16] The minimum value of $d_t(i)$ occurs when there is more than one firm in the industry. If there is a single firm in an industry, then, as you would expect, very little research will be performed (because the firm's only incentive to cut prices comes from competition from other industries, competition which is very weak, since those industries are producing poor substitutes to its own good). However, this drop in research incentives works entirely through the mark-up channel, and $d_t(i) \to 1$ as $J_t(i) \to 1$. One intuition for this is that there can be no strategic behaviour when there is only a single firm.

This does not seem particularly plausible. However, it will turn out that optimal patent rents grow at exactly this rate, so with patent protection we will be able to sustain long run growth even when overhead costs are asymptotically dominated by the costs of research. In the presence of sufficiently severe financial frictions of the "pledgibility constraint" form (Hart and Moore 1994), it may be shown that long run growth is sustainable even without patent protection. We leave the details of this for future work.

Appropriation is performed in an industry if and only if $g_t(i) > 0$, which, for a non-patent protected industry no longer performing research, is true if and only if:

$$\frac{A_t^*(i)}{A_t^*} < \left(\frac{A_t^*(i)^{-\zeta^{\mathrm{A}}} \Upsilon L^{\mathrm{F}}}{A_t^*(i)^{-\zeta^{\mathrm{A}}} \Upsilon L^{\mathrm{F}} + \tau \frac{\mu_t(i)}{d_t(i)}} \right)^{\frac{1}{\tau}}.$$

The left hand side of this equation is the relative productivity of the industry compared to the frontier. The right hand side of this equation will be shrinking over time at roughly $\frac{\zeta^{\mathrm{A}}}{\tau}$ times the growth rate of the frontier, meaning the no-appropriation cut-off point is also declining over time. Indeed, we show in the online appendix (Holden 2013b sec. 2) that asymptotically the relative productivity of non-protected firms shrinks at $\frac{\zeta^{\mathrm{A}}}{\tau}\left[1 + \frac{\zeta^{\mathrm{A}}}{\tau}\right]^{-1}$ times the growth rate of the frontier. This is plausible since productivity differences across industries have been steadily increasing over time,[17] and is important for the tractability of our model since it enables us to focus on the asymptotic case in which non-protected firms never perform appropriation. It is also in line with the long delays in the diffusion of technology found by Mansfield (1993) amongst others.

1.4. Inventors

Each new industry is controlled by an inventor who owns the patent rights to the product the industry produces. Until the inventor's product goes on sale, the patent holder can successfully protect their revenue stream through contractual

[17] Some indirect evidence for this is provided by the increase in wage inequality, documented in e.g. Autor, Katz, and Kearney (2008). Further evidence is provided by the much higher productivity growth rates experienced in manufacturing, compared to those in services (mostly unpatented and unpatentable), documented in e.g. Duarte and Restuccia (2009).

arrangements, such as non-disclosure agreements. This means that even in the absence of patent-protection a patent holder will receive one period of revenues. In this period, and each subsequent one for which they have a patent, the inventor optimally chooses the rent $\mathcal{R}_t(i)$ (or equivalently $L_t^{\mathcal{R}}(i)$) to charge all the firms that wish to produce their product. We are supposing inventors lack the necessary human capital to produce their product at scale themselves.

The inventor of a new product has a probability of $1-q$ of being granted a patent to enable them to extract rents for a second period. After this, if they have a patent at t, then they face a constant probability of $1-q$ of having a patent at $t+1$.

The reader should have a firm such as Apple in mind when thinking about these inventors. Apple has no manufacturing plants and instead maintains its profits by product innovation and tough bargaining with suppliers.

1.4.1. Optimal rent decisions

Inventors' businesses are also owned by households; hence, an inventors' problem is to choose $L_{t+s}^{\mathcal{R}}(i)$ for $s \in \mathbb{N}$ to maximise their expected profits, which are given by:

$$\pi_t := \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s (1-q)^s \left[\prod_{k=1}^{s} \Xi_{t+k}\right] L_{t+s}^{\mathcal{R}}(i) W_{t+s} J_{t+s}(i),$$

subject to an enforceability constraint on rents. If the rents charged by a patent-holder go too high, a firm is likely to ignore them completely in the hope that either they will be lucky, and escape having their profits confiscated from them by the courts (since proving patent infringement is often difficult), or that the courts will award damages less than the licence fee. This is plausible since the relevant U.S. statute states that "*upon finding for the claimant the court shall award the claimant damages adequate to compensate for the infringement but in no event less than a reasonable royalty for the use made of the invention by the* *infringer, together with interest and costs as fixed by the court*".[18, 19] The established legal definition of a "reasonable royalty" is set at the outcome of a hypothetical bargaining process that took place immediately before production,[20] so patent-holders may just as well undertake precisely this bargaining process before production begins.[21]

This leads patent-holders to set:

$$L_t^{\mathcal{R}}(i) = \frac{1-\wp}{\wp}[L_t^{\mathrm{R}}(i) + L_t^{\mathrm{A}}(i) + L^{\mathrm{F}}], \tag{1.4}$$

at least for sufficiently large t, where $\wp \in (0,1)$ is the bargaining power of the firm, in the sense of the generalized Nash bargaining solution. The simple form of this expression comes from the fact that a firm's production period profits (which is what is being bargained over) are precisely equal to the costs they face prior to production, thanks to the free entry condition. A full description of the legally motivated bargaining process is contained in the online appendix (Holden 2013b sec. 3), along with a discussion of some technical complications pertaining to off equilibrium play.

From combining (1.2) and (1.4) then, at least for sufficiently large t, in the limit as as $\sigma_Z \to 0$, we have that:

$$\mathcal{L}_t^{\mathrm{R}}(i) = \frac{\wp\mu_t(i) - d_t(i)A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi(L_t^{\mathrm{A}}(i) + L^{\mathrm{F}})}{d_t(i) - \gamma\wp\mu_t(i)}.$$

For there to be growth in the long run then, we now require $d_t(i) > \gamma\wp\mu_t(i)$, which together with the second order and appropriation uniqueness conditions means that it must at least be true that $\gamma\wp < \frac{1}{\mu_t(i)} < \min\{\gamma, \tau\}$.[22] We see that, once optimal rents are allowed for, research is no longer decreasing in mark-ups

[18] 35 U.S.C. § 284 Damages.

[19] The reasonable royalty condition is indeed the relevant one for us since our assumption that the patent-holder lacks the necessary human capital to produce at scale themself means it would be legally debatable if they had truly "lost profits" following an infringement (Pincus 1991).

[20] See the online appendix (Holden 2013b sec. 3) for evidence supporting this interpretation.

[21] In any case, if we allow for idiosyncratic "idea shocks" firms will wish to delay bargaining until this point anyway, since with a bad shock they will be less inclined to accept high rents. Patent-holders also wish to delay till this point because the more sunk costs the firms have already expended before bargaining begins, the greater the size of the "pie" they are bargaining over.

[22] If the number of firms in protected industries is growing over time then $d_t(i) \to 1$, so asymptotically these conditions are equivalent.

within an industry, at least for firms at the frontier. Mathematically, this is because the patent-holder sets rents as such a steeply increasing function of research levels. More intuitively, you may think of the patent-holder as effectively controlling how much research is performed by firms in their industry, and as taking most of the rewards from this research. It is then unsurprising that we reach these Schumpeterian conclusions.[23]

1.4.2. Invention and long-run stability

We consider invention as a costly process undertaken by inventors until the expected profits from inventing a new product fall to zero. New products appear at the end of the product spectrum. Additionally, once a product has been invented, it cannot be "un-invented". Therefore, the product index i always refers to the same product, once it has been invented.

There is, however, no reason to think that newly invented products will start off with a competitive production process. A newly invented product may be thought of as akin to a prototype: yes, identical prototypes could be produced by the same method, but doing this is highly unlikely to be commercially viable. Instead, there will be rapid investment in improving the product's production process until it may be produced as efficiently as its rivals can be. In our model, this investment in the production process is performed not by the inventor but by the manufacturers. Prototyping technology has certainly improved over

[23] The empirical evidence (Scott 1984; Levin, Cohen, and Mowery 1985; Aghion et al. 2005; Tingvall and Poldahl 2006) suggests that the cross-industry relationship between competition and research takes the form of an inverted-U. Based on the fact that strategic distortions are maximised (i.e. $d_t(i)$ is minimised) when there is a small finite number of firms, one might perhaps hope that this holds in our model too. Unfortunately, the maximum of $\frac{\mu_t(i)}{d_t(i)}$ (and hence of research) as a function of $J_t(i)$ may be shown to always occur at some $J_t(i) < 1$. While fractional entry may be a legitimate way of modelling niche products that are never fully commercialised, we prefer to explain the inverted-U in the data with reference to the cross-sectional distribution of industries. New industries will start with a production process behind that of the frontier, and thus firms in them will wish to perform large amounts of appropriation and relatively small amounts of research, since appropriation is a cheaper means of increasing productivity for a firm behind the frontier. In the presence of a luck component to appropriation (not included above, for simplicity) this leads new industries to have the highest degree of productivity dispersion, as older industries remain close to the frontier. As a result of this high productivity dispersion, there will be firms in new industries setting both very high, and very low mark-ups, which, combined with the fact they are performing less research than more mature patent-protected industries, would generate an inverted-U.

time;[24] in light of this, we assume that a new product i is invented with a production process of level $A_t^*(i) = E_t A_t^*$, where $E_t \in (0,1)$ controls initial relative productivity.

Just as we expect process research to be getting harder over time, as all the obvious process innovations have already been discovered, so too we may expect product invention to be getting harder over time, as all the obvious products have already been invented. In addition, the necessity of actually finding a way to produce a prototype will result in the cost of product invention also increasing in $A_t^*(i)$, the initial productivity level of the process for producing the new product. As a result of these considerations, we assume that the labour cost is given by $\mathcal{L}^{\mathrm{I}} I_{t-1}^{\chi} A_t^*(i)^{\zeta^{\mathrm{I}}}$, where $\mathcal{L}^{\mathrm{I}} > 0$ determines the difficulty of invention and where $\chi \in \mathbb{R}$ and $\zeta^{\mathrm{I}} > 0$ control the rate at which inventing a new product gets more difficult because of, respectively, an increased number of existing products or an increased level of productivity.

We are assuming there is free entry of new inventions, so the marginal entrant must not make a positive profit from entering. That is, $I_t \geq I_{t-1}$ must be as small as possible such that:

$$\mathcal{L}^{\mathrm{I}} I_{t-1}^{\chi} A_t^*(i)^{\zeta^{\mathrm{I}}} W_t \geq \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s (1-\mathcal{q})^s \left[\prod_{k=1}^{s} \Xi_{t+k} \right] L_{t+s}^{\mathcal{R}}(I_t) W_{t+s} J_{t+s}(I_t).$$

If, after a shock, invention can satisfy this equation with equality without the growth rate of the stock of products turning negative, then the number of firms per industry will not have to adjust significantly. However, if the $I_t \geq I_{t-1}$ constraint binds, then the number of firms per industry will have to adjust instead, meaning there may be an asymmetry in the response of mark-ups to certain shocks.

It may be shown that, in the long run, $g_I = \frac{1}{1+\chi}(g_N - \zeta^{\mathrm{I}} g_{A^*})$ (where g_V is the asymptotic growth rate of the variable V_t). Therefore, if $\chi = \zeta^{\mathrm{I}} = 0$ the stock of products will grow at exactly the same rate as population, and away from this special case it will be growing more slowly. If invention were to stop

[24] Examples of recent technologies that have raised the efficiency of prototype production include 3D printing and computer scripting languages such as Python.

asymptotically, eventually there would be no protected industries, and hence no productivity growth. Therefore, for long-run growth, we either require that $g_N \geq \zeta^{\mathrm{I}} g_{A^*}$ (which will hold providing research is getting more difficult sufficiently slowly, as long as population growth continues), or that there is sufficiently fast depreciation of the stock of products.[25] Even without product depreciation, productivity growth may be sustained indefinitely in the presence of a declining population if the government offers infinitely renewable patent-protection.

The existence of a solution for our model, at all time periods, requires the number of firms in a protected industry to be bounded below asymptotically. The previous result on the growth rate of the stock of products implies it is sufficient that $\left(\zeta^{\mathrm{R}} - \frac{\zeta^{\mathrm{I}}}{1+\chi}\right) g_{A^*} \leq \frac{\chi}{1+\chi} g_N$ for this to hold. This inequality is guaranteed to be satisfied providing $\zeta^{\mathrm{R}} - \frac{\zeta^{\mathrm{I}}}{1+\chi}$ is sufficiently small. To do this while also ensuring that $g_I > 0$ requires that $\max\left\{\zeta^{\mathrm{I}}, \zeta^{\mathrm{R}} + \frac{1}{\chi}\left(\zeta^{\mathrm{R}} - \zeta^{\mathrm{I}}\right)\right\} < \frac{g_N}{g_{A^*}}$, which will hold for a positive measure of parameter values providing population growth is strictly positive.[26]

Assuming this condition holds, we may show[27] that providing the growth rate of the productivity of newly invented products is sufficiently close to the frontier growth rate (i.e. E_t does not decline too quickly[28]), asymptotically catch-up to the frontier is instantaneous in protected industries, and the frontier growth rate is stationary. This instantaneous catch-up to the frontier means that, had we allowed for industry-specific shocks, all other protected industries would "inherit" the best industry shock, the period after it arrived. This justifies our focus on aggregate "idea" shocks. Additionally, instantaneous catch-up to the frontier

[25] Bilbiie, Ghironi, and Melitz (2012) include such product depreciation in their model. We have chosen not to model it here.

[26] More generally, when population is stable, providing there is sufficiently fast (proportional) depreciation of the stock of products, we just require that $\zeta^{\mathrm{R}} < \frac{\zeta^{\mathrm{I}}}{1+\chi}$.

[27] Suppose $(i_t)_{t=0}^{\infty}$ is a sequence of industries, all protected at t, whose productivity grows at rate $\tilde{g} \leq g_{A^*}$ asymptotically. We conjecture that $\lim_{t\to\infty} A_t^{**}(i_t)^{-\zeta^{\mathrm{R}}} \Psi L_t^{\mathrm{A}}(i_t) = 0$ and verify. This assumption implies that effective research is asymptotically bounded, since mark-ups are. Hence from (1.3), since $\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$, effective appropriation is growing at a rate in the interval $\left(\frac{\zeta^{\mathrm{R}}\tilde{g} - \zeta^{\mathrm{A}}\tilde{g}}{2}, \frac{\zeta^{\mathrm{R}} g_{A^*} - \zeta^{\mathrm{A}}\tilde{g}}{2}\right) \subseteq (0, \infty)$. Therefore $A_t^{**}(i)^{-\zeta^{\mathrm{R}}} \Psi L_t^{\mathrm{A}}(i_t)$ is growing at a rate in the interval $\left(-\zeta^{\mathrm{R}} g_{A^*} + \zeta^{\mathrm{A}}\tilde{g} + \frac{\zeta^{\mathrm{R}}\tilde{g} - \zeta^{\mathrm{A}}\tilde{g}}{2}, -\zeta^{\mathrm{R}}\tilde{g} + \zeta^{\mathrm{A}}\tilde{g} + \frac{\zeta^{\mathrm{R}} g_{A^*} - \zeta^{\mathrm{A}}\tilde{g}}{2}\right)$. For our claim to be verified we then just need that $\frac{\zeta^{\mathrm{R}}}{2\zeta^{\mathrm{R}} - \zeta^{\mathrm{A}}} g_{A^*} < \tilde{g}$, which certainly holds when $\tilde{g} = g_{A^*}$ as $\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$.

[28] As $\zeta^{\mathrm{A}} \to 0$ it is sufficient that E_t is declining at less than half the rate that A_t^* is growing.

means that providing there is population growth or product depreciation, asymptotically, long-run growth may be sustained even in the absence of patent-protection (i.e. when $q = 1$), as the one period in which the inventor has a first mover advantage is sufficient for their industry to surpass the existing frontier.

If the number of firms in protected industries were asymptotically infinite, then our simulations would tell us nothing about the consequences of the variations in this number that we might see non-asymptotically. Therefore, it will be helpful if it is additionally the case that this number is finite asymptotically. To guarantee this will, unfortunately, require a knife-edge assumption, namely that $\left(\zeta^{\mathrm{R}} - \frac{\zeta^{\mathrm{I}}}{1+\chi}\right) g_{A^*} = \frac{\chi}{1+\chi} g_N$. To satisfy this without restricting population growth rates means $\chi = 0$ (so invention is not made more difficult by the number of existing products) and $\zeta^{\mathrm{R}} = \zeta^{\mathrm{I}}$ (so prototype production is increasing in difficulty at the same rate as research). The former assumption may be justified by noting that many situations in which invention is apparently getting harder over time because of congestion effects may equally well by explained by production-process-difficulty effects. The latter assumption is immediately plausible, since both parameters are measuring the complexity of working with a given production process. However, unlike with knife-edge growth models whereby relatively slight departures from the stable parameter values results in growth that could not possibly explain our observed stable exponential growth, here, away from the knife-edge case we will have slowly decreasing mark-ups, consistent with Ellis's (2006) evidence of a persistent decline in UK whole economy mark-ups over the last thirty years and Kim's (2010) evidence of non-stationarity in mark-ups.

We assume then that $0 = \chi < \zeta^{\mathrm{A}} < \zeta^{\mathrm{R}} = \zeta^{\mathrm{I}}$. Since asymptotically non-protected industries perform no research or appropriation under these assumptions, their entry cost to post-entry industry profits ratio is tending to zero, meaning their number of firms will tend to infinity as $t \to \infty$. This is in line with our motivating intuition that excess entry in non-protected industries kills research and appropriation incentives.

2. Simulations

With $0 = \chi < \zeta^{\mathrm{A}} < \zeta^{\mathrm{R}} = \zeta^{\mathrm{I}}$, as $t \to \infty$ the behaviour of our model tends towards stationarity in the key variables. We simulate this asymptotically stationary model. For convenience we define $\zeta := \zeta^{\mathrm{R}} = \zeta^{\mathrm{I}}$. The full set of equations of the de-trended model is given in the online appendix (Holden 2013b sec. 4). The definition of equilibrium here is entirely standard.

When $\lambda = \nu = \gamma = 1$, it may be shown analytically that the equations determining the model's steady-state have at most two solutions with more than one firm in each industry. However, only one of these two solutions exists for large values of $\mathcal{L}^{\mathrm{I}}$, i.e. when invention is costly. Since we think that in reality invention is getting harder over time due to congestion effects (i.e. $\chi > 0$), any solution that only exists for small values of $\mathcal{L}^{\mathrm{I}}$ is non-feasible. Our numerical investigations suggest that the model always has at most these two equilibria, and that always at most one of them exists for large values of $\mathcal{L}^{\mathrm{I}}$.[29] However, at the chosen parameters, the model has a unique solution, which will exist for arbitrarily high values of $\mathcal{L}^{\mathrm{I}}$.

Since $\Psi E^{\zeta} \mathcal{L}^{\mathrm{I}}$ always occurs as a group, without loss of generality we may make the normalization $\Psi := E := 1$. We fix all of the model's other parameters, except $\mathcal{L}^{\mathrm{I}}$, to the values we estimated in a medium-scale version of the model in Holden (2013a). $\mathcal{L}^{\mathrm{I}}$ is set such that the number of firms in patent-protected industries in this model is equal to that in the estimated extended model. The full parameterisation is reported in the online appendix (Holden 2013b sec. 7). We note that β is set to 0.99 consistent with an interpretation as a quarterly model, given our focus on both the business cycle as well as lower frequency phenomena.

In Figure 1 we present the nonlinear perfect foresight impulse responses that result from IID (hence non-persistent) shocks to population growth ($G_{N,t}$) and "ideas" (Z_t), in the fully nonlinear model. [30] We set the magnitude of the idea shock to 1%, and choose the magnitude of the population shock to give a similar

[29] It may be shown analytically that the complete model may always be solved by solving a single nonlinear equation, which was always concave for all the parameters we examined.
[30] This was performed using Dynare (Adjemian et al. 2011).

productivity response after 5 years.[31] Each graph is given in terms of per cent deviations from the value the variable would have taken had the shock never arrived, and the horizontal axis shows time in years, though this remains a quarterly model.

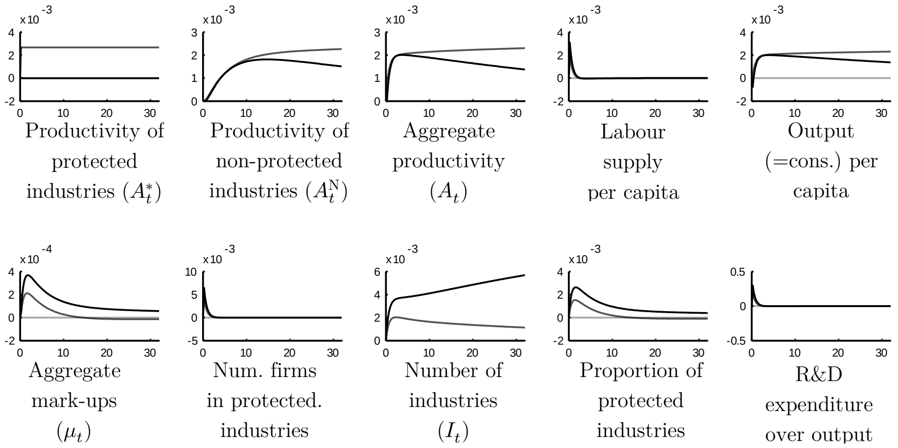

Figure 1: Impulse responses from population (solid) and idea (dashed) shocks.
(Vertical axes are in percent, horizontal axes are in years.)

The principle mechanism of our paper is illustrated most clearly by the population growth rate shock, shown by a solid line in each graph. (We do not wish to advance population shocks as a key driver of business cycles though, since real rigidities will significantly reduce their impact.) Following a permanent increase in population, demand is permanently higher, so, in the long run, the number of industries must grow to balance this out. Given sufficiently inelastic labour supply, this long run increase in the measure of industries requires a short-run substitution of labour from production to invention, pushing down consumption and pushing up wages, and so moderating the rate at which invention will grow. Consequently, in the short run some of the additional demand is absorbed by fluctuations in the number of firms in each industry. Without this additional margin of adjustment, this shock would have led to a large increase in average firm sizes, with a consequent increase in the frontier growth rate and counter-factually large unit root in output.

[31] This required a 0.01% shock to $G_{N,t}$.

Despite the tiny movement in frontier productivity (less than 0.000001%), there is still however a substantial movement in aggregate productivity in the medium-term. Following the shock, more new products are being invented each period, meaning that a greater proportion of industries are relatively new, and so a greater proportion are patent-protected. But because patent-protected industries have such strong incentives to catch-up to the frontier, patent-protected industries are more productive than non-protected ones, so an increase in the proportion of industries that are patent protected means an increase in aggregate productivity. Patent-protected industries also have higher mark-ups due to the cost of paying licence fees, enabling our model to generate pro-cyclical mark-ups.[32] Indeed, since there is so much persistence in productivity coming from patent protection, as in the data, movements in mark-ups lead movements in output, bringing us close to the observed cross correlation even in this toy model.

Fluctuating invention rates also drive the model's response to any other shock that might be considered, not least the idea shock, shown by the dashed line in each graph. Initially, an idea shock just increases the productivity of patent-protected industries. This also makes them relatively more profitable, enabling patent holders to extract higher rents, and leading to an increase in invention with a corresponding further increase in aggregate productivity. Over time, patent protected industries fall out of patent-protection, carrying their higher productivity with them, and thus increasing the average productivity of non-protected firms too. Consequently, aggregate productivity slowly rises towards its permanently higher long run level. However, even with this reasonably large research productivity shock (1%), frontier productivity still only rises by less than 0.005%, consistent with a very small unit root.

[32] Pavlov and Weder (2012) also develop a business cycle model capable of generating pro-cyclical mark-ups, via the changing importance of different types of *buyers* over the business cycle. The properties of these buyers are exogenous in their model however, whereas the properties of the different types of *sellers* that drive our results are endogenous.

3. Conclusion

Many have expressed the worry that "the apparent fit of the DSGE model [has] more to do with the inclusion of suitable exogenous driving processes than with the realism of the model structure itself"[33]. In this paper, we have demonstrated that if productivity is endogenized through research, appropriation and invention then even a frictionless RBC model is capable of generating rich persistent dynamics from uncorrelated shocks, thanks to fluctuations in the proportion of industries that are producing patent-protected products. Furthermore, this improvement in the model's propagation mechanism does not come at the expense of implausibly large unit roots in output, counter-factual movements in mark-ups, or the use of a growth model that we can reject thanks to the absence of strong scale effects in the data. In all of these respects, then, our model presents a substantial advance on the prior literature. In Holden (2013a) we go further still, embedding our core model within a fully featured medium-scale model, and showing that this enables just a few shocks to explain much of the data at both business and medium frequencies.

Our model suggests that a switch to indefinite patent protection would result in significant welfare improvements. Such a switch would both permanently increase the level of aggregate productivity, and substantially lessen its variance and persistence, while only slightly increasing mark-ups and efficiency losses due to research duplication. Indeed, it may be shown that in our model increasing patent protection even slightly increases growth rates, as industry profits are decreasing in aggregate productivity, and so with indefinite patent protection each (protected) industry has fewer firms meaning higher mark-ups and higher research. However, it is clear that the structure of our model has "stacked-the-deck" in favour of finding a beneficial role for patent protection. Patents in our model are less broad than in the real world, and they do not hinder future research or invention. One minimal conclusion we can draw on patent protection is that product patents should at least be long enough that by the end of patent

[33] Del Negro et al. (2007) paraphrasing Kilian (2007).

protection, production process have reached frontier productivity. In our model, this time goes to zero asymptotically. A less radical policy change might be to grant temporary extensions to patents that would otherwise expire during a recession. We intend to explore the full policy implications of this model in future work.

4. References

Adjemian, Stéphane, Houtan Bastani, Michel Juillard, Ferhat Mihoubi, George Perendia, Marco Ratto, and Sébastien Villemot. 2011. *Dynare: Reference Manual, Version 4*. Dynare Working Papers. CEPREMAP.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: An Inverted-U Relationship." *Quarterly Journal of Economics* 120 (2) (May): 701–728.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2) (March): 323–351.

Autor, David H, Lawrence F Katz, and Melissa S Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90 (2): 300–323.

Bachmann, Ruediger, and Christian Bayer. 2009. *The Cross-section of Firms over the Business Cycle: New Facts and a DSGE Exploration.* CESifo Group Munich.

Balasubramanian, Natarajan, and Jagadeesh Sivadasan. 2011. "What Happens When Firms Patent? New Evidence from U.S. Economic Census Data." *Review of Economics and Statistics* 93 (1) (July 7): 126–146.

Bental, Benjamin, and Dan Peled. 1996. "The Accumulation of Wealth and the Cyclical Generation of New Technologies: A Search Theoretic Approach." *International Economic Review* 37 (3) (August): 687–718.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz. 2011. *Endogenous Entry, Product Variety, and Business Cycles.* CEPR Discussion Papers. C.E.P.R. Discussion Papers.

———. 2012. “Endogenous Entry, Product Variety, and Business Cycles.” *Journal of Political Economy* 120 (2) (April 1): 304–345.

Bils, Mark. 1998. “Discussion of Basu (1998).” *Conference Series ; [Proceedings]* (Jun).

Bonanno, Giacomo. 1988. “Oligopoly Equilibria When Firms Have Local Knowledge of Demand.” *International Economic Review* 29 (1) (February 1): 45–55.

Boulhol, Hervé. 2007. “The Convergence of Price–cost Margins.” *Open Economies Review* 19 (2) (May): 221–240.

Broda, Christian, and David E Weinstein. 2010. “Product Creation and Destruction: Evidence and Price Implications.” *American Economic Review* 100 (3) (June): 691–723.

Campbell, Jeffrey R. 1998. “Entry, Exit, Embodied Technology, and Business Cycles.” *Review of Economic Dynamics* 1 (2) (April): 371–408.

Cochrane, John H. 1988. “How Big Is the Random Walk in GNP?” *Journal of Political Economy* 96 (5) (October 1): 893–920.

Comin, Diego. 2009. “On the Integration of Growth and Business Cycles.” *Empirica* 36 (2) (May 1): 165–176.

Comin, Diego, and Mark Gertler. 2006. “Medium-Term Business Cycles.” *American Economic Review* 96 (3) (June): 523–551.

Comin, Diego, Mark Gertler, and Ana Maria Santacreu. 2009. *Technology Innovation and Diffusion as Sources of Output and Asset Price Fluctuations*. National Bureau of Economic Research, Inc.

Comin, Diego, and Sunil Mulani. 2009. “A Theory of Growth and Volatility at the Aggregate and Firm Level.” *Journal of Monetary Economics* 56 (8) (November): 1023–1042.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. “Monopolistic Competition and Optimum Product Diversity.” *The American Economic Review* 67 (3) (June): 297–308.

Duarte, Margarida, and Diego Restuccia. 2009. “The Role of the Structural Transformation in Aggregate Productivity”. Working Papers tecipa-348. University of Toronto, Department of Economics.

Eisfeldt, Andrea L., and Adriano A. Rampini. 2006. “Capital Reallocation and Liquidity.” *Journal of Monetary Economics* 53 (3) (April): 369–399.

Ellis, Colin. 2006. *Elasticities, Markups and Technical Progress: Evidence from a State-Space Approach*. SSRN.

Ethier, Wilfred J. 1982. “National and International Returns to Scale in the Modern Theory of International Trade.” *The American Economic Review* 72 (3) (June): 389–405.

Francois, Patrick, and Huw Lloyd-Ellis. 2008. “Implementation Cycles, Investment, and Growth.” *International Economic Review* 49 (3) (August 1): 901–942.

———. 2009. “Schumpeterian Cycles with Pro-cyclical R&D.” *Review of Economic Dynamics* 12 (4) (October): 567–591.

Groth, Christian, Karl-Josef Koch, and Thomas Steger. 2009. “When Economic Growth Is Less Than Exponential.” *Economic Theory* (July 4).

Hart, Oliver, and John Moore. 1994. “A Theory of Debt Based on the Inalienability of Human Capital.” *The Quarterly Journal of Economics* 109 (4) (November 1): 841–879.

Holden, Tom. 2011. *Products, Patents and Productivity Persistence: A DSGE Model of Endogenous Growth*. Dynare Working Papers. CEPREMAP.

———. 2013a. “Data Consistent Modelling of Medium-frequency Cycles and Their Origins.”

———. 2013b. “Online Appendices to ‘Reconciling Near Trend-stationary Growth with Medium-frequency Cycles’ and ‘Data Consistent Modelling of Medium-frequency Cycles and Their Origins’.”

Horii, Ryo. 2011. “Wants and Past Knowledge: Growth Cycles with Emerging Industries.” *Journal of Economic Dynamics and Control* In Press, Accepted Manuscript.

Jaimovich, Nir. 2007. “Firm Dynamics and Markup Variations: Implications for Sunspot Equilibria and Endogenous Economic Fluctuations.” *Journal of Economic Theory* 137 (1) (November): 300–325.

Kehrig, Matthias. 2011. *The Cyclicality of Productivity Dispersion*. Center for Economic Studies, U.S. Census Bureau.

Kilian, Lutz. 2007. “Comment.” *Journal of Business & Economic Statistics* 25. Journal of Business & Economic Statistics: 156–159.

Kim, Bae-Geun. 2010. “Identifying a Permanent Markup Shock and Its Implications for Macroeconomic Dynamics.” *Journal of Economic Dynamics and Control* (April 14).

Levin, Richard C., Wesley M. Cohen, and David C. Mowery. 1985. “R & D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses.” *The American Economic Review* 75 (2) (May): 20–24.

Mansfield, Edwin. 1993. “The Diffusion of Industrial Robots in Japan and the United States.” *Research Policy* 22 (2) (April): 105.

Matsuyama, Kiminori. 1999. “Growing Through Cycles.” *Econometrica* 67 (2) (March): 335–347.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters. 2007. “Rejoinder.” *Journal of Business and Economic Statistics* 25 (2): 159.

Nekarda, Christopher, and Valerie Ramey. 2010. “The Cyclical Behavior of the Price-Cost Markup.”

Nuño, Galo. 2008. *Schumpeterian Foundations of Real Business Cycles*. Working Papers. International Economics Institute, University of Valencia.

———. 2009. *Technology, Convergence and Business Cycles*. Banco de España.

———. 2011. “Optimal Research and Development and the Cost of Business Cycles.” *Journal of Economic Growth* 16 (3) (April 26): 257–283.

Pavlov, Oscar, and Mark Weder. 2012. “Variety Matters.” *Journal of Economic Dynamics and Control* 36 (4) (April): 629–641.

Peretto, Pietro F. 1999. "Cost Reduction, Entry, and the Interdependence of Market Structure and Economic Growth." *Journal of Monetary Economics* 43 (1) (February 19): 173–195.

Phillips, Kerk L., and Jeff Wrase. 2006. "Is Schumpeterian `creative Destruction' a Plausible Source of Endogenous Real Business Cycle Shocks?" *Journal of Economic Dynamics and Control* 30 (11) (November): 1885–1913.

Pincus, Laura B. 1991. "Computation of Damages in Patent Infringement Actions." *Harvard Journal of Law & Technology* 5: 95.

Scott, John. 1984. "Firm Versus Industry Variability in R&D Intensity." In *R & D, Patents, and Productivity*. NBER.

Serrano, C. J. 2007. "Technology Transfer and the Business Cycle: Evidence from Patent Reassignments."

Serrano, Carlos J. 2010. "The Dynamics of the Transfer and Renewal of Patents." *The RAND Journal of Economics* 41 (4) (December 1): 686–708.

Siliverstovs, Boriss, and D'Artis Kancs. 2012. *R&D and Non-Linear Productivity Growth of Heterogeneous Firms*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

The Associated Press. 2008. "Court Rules Business Concept Cannot Be Patented." *The New York Times*, October 31, sec. Business.

Tingvall, Patrik Gustavsson, and Andreas Poldahl. 2006. "Is There Really an Inverted U-shaped Relation Between Competition and R&D?" *Economics of Innovation and New Technology* 15 (3): 101.

Wälde, Klaus. 2005. "Endogenous Growth Cycles." *International Economic Review* 46 (3): 867–894.