Products, patents and productivity persistence:
A DSGE model of endogenous growth

Tom Holden

CEPREMAP

CENTRE POUR LA RECHERCHE ECONOMIQUE ET SES APPLICATIONS

Products, patents and productivity persistence: A DSGE model of endogenous growth

Tom Holden[1], Department of Economics, University of Oxford

Abstract: *This paper builds a dynamic stochastic general equilibrium (DSGE) model of endogenous growth that is capable of generating substantial degrees of endogenous persistence in productivity. When products go out of patent protection, the rush of entry into their production destroys incentives for process improvements. Consequently, old production processes are enshrined in industries producing non-protected products, resulting in aggregate productivity persistence. Our model also generates sizeable delayed movements in productivity in response to preference shocks, providing a form of endogenous news shock. Finally, if we calibrate our model to match a high aggregate mark-up then we can replicate the negative response of hours to a positive technology shock, even without the inclusion of any frictions.*

Keywords: *productivity persistence, patent protection, oligopoly, research and development*

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

1.   Introduction

In this paper, we present a tractable business cycle model in which even uncorrelated shocks lead to highly persistent movements in macroeconomic aggregates. In order to match the data, standard DSGE models require their driving processes to be highly persistent.[2] However, it is difficult to justify the required persistence of these exogenous processes independently of the entire model's empirical performance. By incorporating endogenous growth features, our model produces endogenous persistence in productivity, resolving this problem. Our model does this by generating delays in the diffusion of technologies of arbitrarily long duration, in line with the slow adoption found by Mansfield (1993) and others. These delays are between one firm and another, rather than between inventors and identical manufacturing firms as in the work of Comin et al. (Comin and Gertler 2004; Comin 2009; Comin, Gertler, and Santacreu 2009). This heterogeneity across firms drives the persistence of productivity, since old technologies will remain in use in firms that choose not to invest in process improvements through research or adoption. This occurs in our model since there is so much competition to produce non-patent-protected products that they have almost no incentive to invest in productivity improvements; their production process will only be improved when it is so far behind the frontier that the adoption of a marginally better technology is almost free.

To distinguish these adoption driven incremental improvements in technology from the discrete conception of adoption used by Comin et al., in the following we refer instead to "appropriation". We assume throughout that only products are patentable,[3] and so by exerting effort firms are able to "appropriate" process innovations from other industries to aid in the production of their own product. We regard process research as equally incremental, with regular small changes rather than unpredictable jumps as in Schumpetarian models (Aghion and Howitt 1992; Wälde 2005; Phillips and Wrase 2006).

The invention of new products is also endogenous in our model. In line with the results of Broda and Weinstein (2010) our model generates pro-cyclical net product creation, which has an impact on aggregate productivity even in the absence of any preference for variety in consumption. This is because an increase in invention results in a greater proportion of products on the market being

---

[2] For example, in discussing the results of estimating their model, Del Negro et al. (2005) write that "the high persistence of many of the exogenous processes raises concerns about the ability of the DSGE model to generate endogenous propagation mechanisms".

[3] This is at least broadly in line with the law in most developed countries: ideas that are not embedded in a product (in which category we include machines) 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*.

under patent-protection, and thus having a cutting-edge production process. Unlike in the model of Bilbiie, Ghironi, and Melitz (2007), in our model an increase in net product creation will also increase the aggregate mark-up, as firms producing patent-protected products face less competition. This source of pro-cyclicality in aggregate mark-ups is dampened by fluctuations in the number of firms producing each product, following Jaimovich (2007), meaning that *within a given industry* mark-ups are counter-cyclical. In line with Nekarda and Ramey's (2010) empirical findings, this allows our model to generate mark-ups that are weakly pro-cyclical on average.

Although our model can also replicate the pro-cyclical adoption found by Comin (2009), we focus on the two consequences of firm heterogeneity already mentioned (the enshrinement of old technologies, and the weighting consequences of variations in the rate of invention), since Comin et al. have already well documented the consequences of pro-cyclical adoption. Indeed, we shall focus on a limit case in which firms producing non patent-protected products optimally choose to wait forever before performing any appropriation. This results in a low dimensional model with simple aggregation across firms, despite the heterogeneity.

Our paper is structured as follows. In section 2 we describe the core model, which we then calibrate and simulate in section 3. In order to illustrate our model's endogenous propagation mechanism, we only look at impulse responses to IID shocks. Despite this, the generated responses exhibit degrees of persistence in excess of those in models driven by correlated shock processes. We also see that preference shocks affect productivity both through the research channel and through the reweighting consequences of changes in invention rates. Since these effects on productivity occur with a lag, we show that these preference shocks may be interpreted as endogenous news shocks, a type of shock that both Beaudry and Lucke (2009) and Schmitt-Grohe and Uribe (2008) find to be key drivers of the business cycle.

In section 4 we consider an alternative calibration of our model in which rather than calibrating mark-ups to match the (low) micro-evidence, we calibrate to match the (high) macro-evidence. Under this calibration, we find that the response of hours to a positive technology shock is negative, providing an explanation for the results of Gali (1999) that avoids invoking any frictions. Finally, in section 5 we consider a small extension to our model that generates much smoother variations in the aggregate mark-up.

2.   The model

The model is a standard quarterly real business cycle (RBC) model without capital, augmented by Jaimovich's (2007) model of endogenous competition, and by the addition of models of research, appropriation and invention. The base RBC model used has no endogenous propagation mechanism, making clearer the contribution of our additions. 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 are then patent-

protected for T periods. Productivity within a firm is increased by performing research or appropriation.

2.1. Households

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

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

where C_t is aggregate period t consumption, L_t^S is aggregate period t labour supply, Θ_t is a demand shock, Φ_t is a labour supply shock, β is the discount rate and v is the inverse of the Frisch elasticity of labour supply to wages, subject to the aggregate budget constraint that $C_t + B_t = L_t^S W_t + B_{t-1} R_{t-1} + \Pi_t$, where B_t is the aggregate number of 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. In the following, where we refer to preference shocks we mean either a shock to Θ_t or a shock to Φ_t.

Let $\beta \Xi_{t+1}$ be the households' period t stochastic discount factor, then the households' first order conditions imply:

$$\Xi_t = \frac{\Theta_t N_t C_{t-1}}{\Theta_{t-1} N_{t-1} C_t}, \qquad \Phi_t L_t^{S^v} = N_t^{1+v} \frac{W_t}{C_t}, \qquad \beta R_t \mathbb{E}_t[\Xi_{t+1}] = 1.$$

2.2. Aggregators

2.2.1. Final good producers

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}) has been chosen to remove the preference for variety in consumption.[4]

---

[4] Incorporating a preference for variety does not change the long-run stability of our model. It does, however, provide an additional mechanism to enhance the pro-cyclicality of productivity.

Normalising the price of the aggregate consumption good to 1, and writing $P_t(i)$ for the price of the aggregate good from industry i in period t, we have that:

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

2.2.2. Industry aggregators

Similarly, each industry aggregate good $Y_t(i)$ is produced by a perfectly competitive industry from the intermediate goods $Y_t(i,j)$ for $j \in \{1, \ldots, J_{t-1}(i)\}$, 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. This means that if $P_t(i,j)$ is the price of intermediate good j in industry i:

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

2.3. Intermediate firms

Intermediate firms that enter in period t conduct research and/or appropriation in that period then produce in the following one using the production process they invented. For the sake of tractability, we assume that once a technology has been used in production within a given industry it is freely available to all other firms *within that same industry* to be used as a base from which to start their research. This may be justified by assuming knowledge leakage at trade fairs and conferences, or by assuming that there are regular labour movements between firms in the same industry—movements that carry industrial secrets with them. Under this interpretation the difficulty of taking technologies from firms producing different products (appropriation) comes from both the lower labour movements between firms in different industries, and the fact that the technology is unlikely to be perfectly transferrable to a different product.

Following Jaimovich (2007) we assume free entry within each industry. Combined with free technology transfers within each industry, this means that without loss of generality we may assume firms exist for only two periods, in the first of which they research and in the second of which they produce.

2.3.1. 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 $P_t(i,j) = \left(1 + \mu_{t-1}(i)\right)\frac{W_t}{A_t(i,j)}$, 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 $t + 1$. Aggregating this across firms implies that $P_t(i) = \left(1 + \mu_{t-1}(i)\right)\frac{W_t}{A_t(i)}$, where:

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

Finally, from aggregating across industries we have that $W_t = \frac{A_t}{1+\mu_{t-1}}$ 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:

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

gives the aggregate productivity level.[5]

2.3.2. Sunk costs: rents, appropriation and research

A firm that enters industry i in period t will generally pay four different costs that period, before they produce in period $t + 1$. Firms borrow in order to cover these upfront costs.

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. Secondly, if the product produced by

---

[5] 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.

industry i is currently patent-protected (i.e. its age is below T), then firms must pay a rent of $\Re_t(i)$ 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 $L_t^\Re(i) := \frac{\Re_t(i)}{W_t}$, e.g. 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^*. 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 $L_t^A(i,j)$ units of appropriation labour in period t then:

$$A_t^{**}(i,j) = \left[A_t^*(i)^\tau + (A_t^{*\tau} - A_t^*(i)^\tau) \frac{A_t^*(i)^{-\zeta^A} \Upsilon L_t^A(i,j)}{1 + A_t^*(i)^{-\zeta^A} \Upsilon L_t^A(i,j)} \right]^{\frac{1}{\tau}}, \tag{2.1}$$

where Υ is the productivity of appropriation labour, $\zeta^A > 0$ 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.

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

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

where Ψ is the productivity of research labour, $\zeta^R > \zeta^A$ 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. [6] 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 the productivity of research declines sharply as the firm attempts to pack more into one period. The restriction that the difficulty of research is increasing over time faster than that of appropriation is made because research is very much specific to the industry in

---

[6] 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).

which it is being conducted, whereas appropriation is a similar task across all industries attempting to appropriate the same technology, and hence is more likely to have been standardised.

In the following, we will assume that $Z_t(i,j) := Z_t$ so that all firms in all industries receive the same "idea" shock, although if they perform no research this will not contribute to the variance of their productivity. 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 $Z_t := \exp(\sigma_Z \epsilon_{Z,t})$, where $\sigma_Z > 0$ and $\epsilon_{Z,t} \sim \text{NIIID}(0,1)$.

2.3.3. Research and appropriation effort decisions

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

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

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. However, since the coordination requirements of asymmetric equilibria render them somewhat implausible, we restrict ourselves to ones in which all firms within an industry choose the same levels of research and appropriation. Let us then define effective research performed by firms in industry i by $\mathfrak{L}_t^{\text{R}}(i) := A_t^{**}(i)^{-\zeta^{\text{R}}} \Psi L_t^{\text{R}}(i,j)$ (valid for any $j \in \{1, \dots, J_{t-1}(i)\}$) and effective appropriation performed by firms in that industry by $\mathfrak{L}_t^{\text{A}}(i) := A_t^*(i)^{-\zeta^{\text{A}}} \Upsilon L_t^{\text{A}}(i,j)$ (again, valid for any $j \in \{1, \dots, J_{t-1}(i)\}$). Providing $\frac{1}{\mu_t(i)} < \min\{\gamma, \tau\}$ and $\gamma > \zeta^{\text{R}}$ (for the second order conditions[7] and for uniqueness), combining the first order and free entry conditions then gives us that, in the limit as $\sigma_Z \to 0$:[8]

$$\mathfrak{L}_t^{\text{R}}(i) = \max\left\{ 0, \frac{A_t^{**}(i)^{-\zeta^{\text{R}}} \Psi \left[L_t^{\text{A}}(i) + L_t^{\Re}(i) + L^{\text{F}} \right] - \mu_t(i)}{\gamma \mu_t(i) - 1} \right\} \tag{2.2}$$

and: $$\mathfrak{L}_t^{\text{A}}(i) = \max\left\{ 0, -\mathfrak{K}_t(i) + \sqrt{\max\{0, \mathfrak{K}_t(i)^2 + \mathfrak{A}_t(i)\}} \right\}, \tag{2.3}$$

---

[7] The second order condition for research may be derived most readily by noting that 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 $\frac{1+\mu_t(i)}{\mu_t(i)}$. 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.

[8] The first order and zero profit conditions are reported in an appendix, section 8.1, 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.

where:

$$\mathfrak{A}_t(i) \coloneqq \frac{1}{\tau\mu_t(i)}\frac{1+(\gamma-\zeta^{\mathrm{R}})\mathfrak{L}_t^{\mathrm{R}}(i)}{1+\gamma\mathfrak{L}_t^{\mathrm{R}}(i)}A_t^*(i)^{-\zeta^{\mathrm{A}}}\Upsilon\big[L_t^{\mathrm{R}}(i)+L_t^{\mathfrak{R}}(i)+L^{\mathrm{F}}\big]\left[1-\left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}\right]-\left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}$$

and:
$$\mathfrak{K}_t(i) \coloneqq 1-\frac{1}{2}\left[1+\frac{1}{\tau\mu_t(i)}\frac{1+(\gamma-\zeta^{\mathrm{R}})\mathfrak{L}_t^{\mathrm{R}}(i)}{1+\gamma\mathfrak{L}_t^{\mathrm{R}}(i)}\right]\left[1-\left(\frac{A_t^*(i)}{A_t^*}\right)^{\tau}\right].$$

The most important thing to note about this solution is that research and appropriation are independent of the level of demand, except insomuch as demand affects mark-ups. This is because variations in the number of firms in each industry absorb the fluctuations in demand. Holding appropriation and rents constant, research is decreasing in the mark-up, since high mark-ups lessen the degree to which market share may be expanded through productivity improvements. Likewise, holding research and rents constant, appropriation is decreasing in mark-ups. It is also increasing in distance from the frontier, since the further behind a firm is the greater are the returns to appropriation. Since both rents and fixed costs are sunk when research and appropriation decisions are made, neither enter into the first order conditions. They do enter into (2.2) and (2.3) though: low fixed costs result in high entry, decreasing post-entry profits and hence discouraging firms from performing research and appropriation.

In industries older than T, rents will be zero (i.e. $L_t^{\mathfrak{R}}(i) \equiv 0$). 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 $A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi\big[L_t^{\mathrm{A}}(i)+L_t^{\mathfrak{R}}(i)+L^{\mathrm{F}}\big]-\mu_t(i)$ is asymptotically negative since $\mu_t(i) \in (\eta\lambda, \lambda]$. Furthermore, appropriation is performed if and only if $\mathfrak{A}_t(i) > 0$, which in this case is true if and only if:

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

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 an appendix, section 8.2, 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,[9] and is important

---

[9] 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).

for the tractability of our model since it enables us to focus on the asymptotic case in which non-protected firms never perform appropriation.

2.4. Inventors

Each new industry is controlled by an inventor who owns the patent rights to the product the industry produces. In each of the T periods for which patent-protection lasts, the inventor optimally chooses the rent $\Re_t(i)$ (or equivalently $L_t^{\Re}(i)$) to charge all the firms performing research within their industry. We suppose inventors lack the necessary human capital to produce their product at scale themselves.

2.4.1. Optimal rent decisions

Inventor's businesses are also owned by households; hence, an inventor's problem is to choose $L_{t+s}^{\Re}(i)$ for $s = 0 \dots T - 1$ to maximise:

$$\mathbb{E}_t \sum_{s=0}^{T-1} \beta^s \left[\prod_{k=1}^{s} \Xi_{t+k} \right] L_{t+s}^{\Re}(i) W_{t+s} J_{t+s}(i),$$

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*". [10,11] The established legal definition of a "reasonable royalty" is set at the outcome of a hypothetical bargaining process that took place immediately before production, [12] so patent-holders may just as well undertake precisely this bargaining process before production begins. [13]

---

[10] 35 U.S.C. § 284 Damages.

[11] 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).

[12] *Georgia-Pacific,* 318 F. Supp. at 1120 (S.D.N.Y. 1970), *modified on other grounds,* 446 F.2d 295 (2d Cir.), *cert. denied,* 404 U.S. 870 (1971), cited in Pincus (1991), defines a reasonable royalty as "*the amount that a licensor (such as the patentee) and a licensee (such as the infringer) would have agreed upon (at the time the infringement began) if both had been reasonably and voluntarily trying to reach an agreement; that is, the amount which a prudent licensee—who desired, as a business proposition, to obtain the license to manufacture and sell a particular article embodying the patented invention—would have been willing to pay as a royalty and yet be able to make a reasonable profit and which amount would have been acceptable by a prudent patentee who was willing to grant a license.*"

[13] 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

This leads patent-holders to set:

$$L_t^{\Re}(i) = \frac{1 - \wp}{\wp}[L_t^{R}(i) + L_t^{A}(i) + L^{F}], \tag{2.4}$$

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. A full description of the legally motivated bargaining process is contained in an appendix, section 8.3, along with a discussion of some technical complications pertaining to off equilibrium play.

From combining (2.2) and (2.4) then, at least for sufficiently large t, in the limit as as $\sigma_Z \to 0$:

$$\mathfrak{L}_t^{R}(i) = \frac{\wp\mu_t(i) - A_t^{**}(i)^{-\zeta^R}\Psi(L_t^{A}(i) + L^{F})}{1 - \gamma\wp\mu_t(i)}.$$

For there to be growth in the long run then, we require $1 > \gamma\wp\mu_t(i)$, which together with the second order and appropriation uniqueness conditions means that $\wp\gamma < \frac{1}{\mu_t(i)} < \min\{\gamma, \tau\}$. For this to hold it is sufficient that $\wp\gamma < \frac{1}{\lambda} \leq \eta \min\{\gamma, \tau\}$, since $\frac{1}{\mu_t(i)} \in \left[\frac{1}{\lambda}, \frac{1}{\eta\lambda}\right)$. We see that, once optimal rents are allowed for, research is no longer decreasing in monopoly power within an industry, at least for leading firms. Instead, the patent-holder effectively controls how much research is performed by firms, and takes most of the rewards from their research, making it unsurprising that we reach these Schumpeterian conclusions.

2.4.2. 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. For simplicity, we assume that new products are available for firms in the period in which they are invented, so a product invented in period t will (potentially) have appropriation and research performed on its production process that period, and will be produced in period $t + 1$. New products appear at the end of the product spectrum. Additionally, once a product has been invented, it cannot be "uninvented". 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

---

because the more sunk costs the firms have already expended before bargaining begins, the greater the size of the "pie" they are bargaining over.

technology has certainly improved over time;[14] 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. However, it is equally possible that due to the greater scope of human activity in the modern world, there are in fact more possible products today. In addition, the necessity of actually finding a way to produce a prototype will result in the cost of product invention also being increasing in $A_t^*(i)$, the initial productivity level of the process for producing the new product. As a result of these considrations, we assume that the labour cost is given by $L_t^I I_{t-1}^\chi A_t^*(i)^{\zeta^I}$, where $L_t^I > 0$ is a stationary process representing the fluctuations in the difficulty of invention and where $\chi \in \mathbb{R}$ and $\zeta^I > 0$ 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:

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

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

It may be shown that, in the long run, $g_I = \frac{1}{1+\chi}(g_N - \zeta^I g_{A^*})$ (where g_V is the asymptotic growth rate of the variable V_t). Therefore, if $\chi = \zeta^I = 0$ 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 asymptotically, eventually there would be no protected industries, and hence no productivity growth. Therefore, for long-run growth, we either require that $g_N \geq \zeta^I g_{A^*}$ (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.[15] 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.

---

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

[15] Bilbiie, Ghironi, and Melitz (2007) include such product depreciation in their model. We have chosen not to model it here.

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 $\max\left\{\zeta^{\mathrm{I}}, \zeta^{\mathrm{R}} + \frac{1}{\chi}(\zeta^{\mathrm{R}} - \zeta^{\mathrm{I}})\right\} < \frac{g_N}{g_{A^*}}$, which will hold for a positive measure of parameter values providing population growth is strictly positive.[16]

Assuming this condition holds, we may show[17] 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[18]), 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.

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 the number of firms is asymptotically finite. 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

---

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

[17] 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 $\lim_{t \to \infty} A_t^{**}(i_t)^{-\zeta^{\mathrm{R}}} \Psi L_t^A(i_t) = 0$ and verify. [17] This assumption implies that effective research is asymptotically bounded, since mark-ups are. Hence from (2.3), 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 $A_t^{**}(i)^{-\zeta^{\mathrm{R}}} \Psi L_t^A(i_t)$ 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 $\zeta^{\mathrm{R}} > \zeta^{\mathrm{A}}$.

[18] 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.

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 $0 = \chi < \zeta^A < \zeta^R = \zeta^I$. 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.

3. Simulations

With $0 = \chi < \zeta^A < \zeta^R = \zeta^I$, as $t \to \infty$ the behaviour our model tends towards stationarity in the key variables. It is this asymptotically stationary model that we simulate. For convenience we define $\zeta := \zeta^R = \zeta^I$.

3.1. The de-trended model

We begin by describing the de-trended model that we simulate, which is the stationary model to which the model described in section 2 converges as $t \to \infty$.

3.1.1. Households

- **Stochastic discount factor:** $\Xi_t = \frac{\Theta_t \hat{C}_{t-1}}{\Theta_{t-1} \hat{C}_t G_{A,t}}$, where $\hat{C}_t := \frac{C_t}{N_t A_t}$ is consumption per person in labour supply units and $G_{V,t}$ is the exponent of the growth rate of the variable V_t at t.

- **Labour supply:** $\Phi_t \hat{L}_t^{S^{\nu}} = \frac{\widehat{W}_t}{\hat{C}_t}$, where $\hat{L}_t^S := \frac{L_t^S}{N_t}$ is labour supply per person and $\widehat{W}_t := \frac{W_t}{A_t}$ is the wage per effective unit of labour supply.

- **Euler equation:** $\beta R_t \mathbb{E}_t[\Xi_{t+1}] = 1$, where R_t is the real interest rate.

3.1.2. Aggregate relationships

- **Aggregate mark-up pricing:** $\widehat{W}_t = \frac{1}{1+\mu_{t-1}}$ where μ_{t-1} is the aggregate mark-up in period t.

- **Mark-up aggregation:** $\left(\frac{1}{1+\mu_t}\right)^{\frac{1}{\lambda}} = \left(\frac{1}{1+\mu_t^P}\right)^{\frac{1}{\lambda}} [1 - \mathfrak{w}_t] + \left(\frac{1}{1+\eta\lambda}\right)^{\frac{1}{\lambda}} \mathfrak{w}_t$, where $\mu_t^P = \mu_t(I_t)$ is the mark-up in any protected industry at $t + 1$, and $\mathfrak{w}_t := \prod_{k=0}^{T-1} G_{I,t-k}^{-1}$ is the proportion of industries that will produce an unprotected product that period.

- **Productivity aggregation:** $\left(\frac{\hat{A}_t}{1+\mu_{t-1}}\right)^{\frac{1}{\lambda}} = \left(\frac{1}{1+\mu_{t-1}^{P}}\right)^{\frac{1}{\lambda}}[1-\mathfrak{w}_{t-1}] + \left(\frac{\hat{A}_t^{N}}{1+\eta\lambda}\right)^{\frac{1}{\lambda}}\mathfrak{w}_{t-1}$, where $\hat{A}_t := \frac{A_t}{A_t^*}$ is aggregate productivity relative to the frontier[19] and

$$\hat{A}_t^{N} := \left[\left(\frac{1}{G_{A^*,t}}\right)^{\frac{1}{\lambda}}\left(1 - G_{I,t-T-1}^{-1}\right) + \left(\frac{\hat{A}_{t-1}^{N}}{G_{A^*,t}}\right)^{\frac{1}{\lambda}}G_{I,t-T-1}^{-1}\right]^{\lambda}$$ is the aggregate relative productivity of non-protected industries.

3.1.3. Firm decisions

- **Strategic in-industry pricing:** $\mu_t^{P} = \lambda\frac{\eta \hat{J}_t^{P}}{\hat{J}_t^{P}-(1-\eta)}$, where $\hat{J}_t^{P} := J_t(I_t)$ is the number of firms in a protected industry performing research at t.

- **Firm research decisions:** $\frac{1}{\wp\mu_t^{P}}\mathbb{E}_t\Xi_{t+1}G_{Y,t+1}\hat{A}_{t+1}^{-\frac{1}{\lambda}}\frac{Z_{t+1}\widehat{\mathfrak{L}}_t^{R}}{1+\gamma Z_{t+1}\widehat{\mathfrak{L}}_t^{R}} = \left(1 - m_t^{R}(i)\right)\mathbb{E}_t\Xi_{t+1}G_{Y,t+1}\hat{A}_{t+1}^{-\frac{1}{\lambda}}$, where $\widehat{\mathfrak{L}}_t^{R} := A_t^{**}(i)^{-\zeta^{R}}\Psi L_t^{R}(i)$ is the amount of effective research conducted by firms in protected industries and Z_t is the aggregate research-return shock. (This equation means that $\widehat{\mathfrak{L}}_t^{R} \approx \frac{\wp\mu_t^{P}}{1-\wp\gamma\mu_t^{P}}$.)

- **Research and appropriation payoff:** $G_{A^*,t} = \left(1 + \gamma Z_t\widehat{\mathfrak{L}}_{t-1}^{R}\right)^{\frac{1}{\gamma}}$.

- **Free entry of firms:** $\beta\frac{1}{\hat{I}_t\hat{J}_t^{P}}\frac{\mu_t^{P}}{1+\mu_t^{P}}\left(\frac{1+\mu_t}{1+\mu_t^{P}}\right)^{\frac{1}{\lambda}}\mathbb{E}_t\Xi_{t+1}G_{Y,t+1}\hat{A}_{t+1}^{-\frac{1}{\lambda}} = \frac{1}{\wp}\widehat{\mathfrak{L}}_t^{R}\frac{\widehat{W}_t}{\widehat{Y}_t}$, where $\hat{I}_t := \frac{I_t}{N_t A_t^{-\zeta}\Psi}$ is the measure of products relative to its trend,[20] and $\widehat{Y}_t := \frac{Y_t}{N_t A_t}$ is output per person in labour supply units.

3.1.4. Inventor decisions

- **Free entry of inventors:** Either $G_{I,t} \geq 1$ binds or

$$\Psi E^{\zeta}L_t^{I}\widehat{W}_t \geq \frac{1-\wp}{\wp}\mathbb{E}_t\sum_{s=0}^{T-1}\beta^{s}\left[\prod_{k=1}^{s}\Xi_{t+k}G_{A,t+k}G_{A^*,t+k}^{\zeta}\right]\widehat{\mathfrak{L}}_{t+s}^{R}\widehat{W}_{t+s}\hat{J}_{t+s}^{P} \text{ does .}[21]$$

---

[19] As a consequence, we have that $G_{A,t} = \frac{\hat{A}_t}{\hat{A}_{t-1}}G_{A^*,t}$.

[20] This means $G_{I,t} = G_{N,t}G_{A^*,t}^{-\zeta}\frac{\hat{I}_t}{\hat{I}_{t-1}}$.

[21] If we define $\mathfrak{I}_{t,s}$ recursively by $\mathfrak{I}_{t,T} := 0$, $\mathfrak{I}_{t,s} := \beta^{s}\left[\prod_{k=1}^{s}\Xi_{t-s+k}G_{A,t-s+k}G_{A^*,t-s+k}^{\zeta}\right]\widehat{\mathfrak{L}}_t^{R}\widehat{W}_t\hat{J}_t^{P} + \mathbb{E}_t\mathfrak{I}_{t+1,s+1}$, then by the law of iterated expectations we may rewrite the first condition as $\Psi E^{\zeta}L_t^{I}\widehat{W}_t \geq \frac{1-\wp}{\wp}\mathfrak{I}_{t,0}$. This is useful since a naïve translation of the inventor entry equation into first order form would result in on the order of T^2 additional auxiliary state variables, whereas this only requires on the order of T.

3.1.5. Market clearing

- **Labour market clearing:** $\hat{L}_t^S = \Psi E^\zeta L_t^I \hat{I}_t \left(1 - \frac{1}{G_{I,t}}\right) + \hat{I}_t[1 - \mathfrak{w}_t]\hat{J}_t^P\widehat{\mathfrak{L}}_t^R + \hat{Y}_t \left[\left(\frac{1}{\hat{A}_t}\right)^{\frac{1}{\lambda_t}} \left(\frac{1+\mu_{t-1}}{1+\mu_{t-1}^P}\right)^{\frac{1+\lambda}{\lambda}} [1 -$

$\mathfrak{w}_{t-1}] + \left(\frac{\hat{A}_t^N}{\hat{A}_t}\right)^{\frac{1}{\lambda_t}} \left(\frac{1+\mu_{t-1}}{1+\eta\lambda}\right)^{\frac{1+\lambda}{\lambda}} \mathfrak{w}_{t-1}\Bigg]$.

- **Goods market clearing:** $\hat{Y}_t = \hat{C}_t$.

3.2. Calibration

Since $\Psi E^\zeta L_t^I$ always occurs as a group, without loss of generality we may make the normalization $\Psi \coloneqq E \coloneqq 1$. Additionally we set $\Phi = 1$ in steady state, which amounts to defining labour supply units. We calibrate our model in order to match the following facts:

- the US economy's steady state per capita growth rate is around 0.46% per quarter;[22]

- the US's population growth rate is around 0.31% per quarter; [23]

- patent protection lasts twenty years (as guaranteed by the WTO); [24]

- aggregate R&D expenditure equals about 2.6% of US GDP; [25]

- the aggregate mark-up is around 9.6%.[26]

We calibrate subject to the constraint that $\wp\gamma < \frac{1}{\lambda} \leq \eta\gamma$ and subject to the equations having a unique equilibrium with $J_t(i) > 1$ in all industries. With the residual degrees of freedom, we maximize both invention growth rates and the difference in mark-ups between protected and non-

---

[22] Real GDP data from the U.S. Department of Commerce: Bureau of Economic Analysis. Civilian Noninstitutional Population data from the U.S. Department of Labor: Bureau of Labor Statistics. 1948Q1-2010Q2.

[23] Average growth rate of US civilian non-institutional population data over the period 1976Q1-2010Q2. Data from the Bureau of Labor Statistics, U.S. Department of Labor. Series LNS10000000Q.

[24] In the "Agreement on Trade-Related Aspects of Intellectual Property Rights" (World Trade Organisation 1994) that all WTO members must sign.

[25] Using data from the Bureau of Economic Analysis, U.S. Department of Commerce. 1998-2007.

[26] The DSGE literature has traditionally found higher mark-ups, for example Del Negro et al. (2007a) find a value of 35%. However, the micro-evidence does not support this. Indeed, recent research (Boulhol 2008) has showed that even mark-ups estimated with an established micro-method (that of Roeger (1995)) are substantially biased upwards. In US data from 1970-2000, Boulhol (2008) finds an average mark-up over total variable costs of 5.6% and an average mark-up over short-run variable costs of at most 9.6%. (The 9.6% figure comes from assuming capital is fixed in the short-run; an intermediate figure is produced if the fixity of capital is measured.) In order to partially reconcile Boulhol's low figures with the high figures traditionally used in macro, we calibrate our aggregate mark-up to this highest estimate of 9.6%. In the alternative calibration examined in section 4 we instead calibrate mark-ups to 35%.

protected industries. [27] The former enables us to bring invention rates towards the high rates of net product creation found in the retail sector by Broda and Weinstein (2010), and the latter maximizes the responsiveness of the frontier growth rate to business cycles. Our calibration results in: [28]

Variable	ζ	λ	η	γ	$\wp$	L^I	g_N	β	ν
Value to 3 s.f.[29]	≈ 0	1.24	0.0763	10.5	0.0453	161	0.00311	0.99	2

3.3. Simulation method

We now simulate our model. In order to do this, we first express the model entirely in terms of variables in logs[30]. We then take a first-order perturbation approximation around the non-stochastic steady state, perturbing in the variance of shocks, and solve for the rational expectations solution of the linearized model.[31] As we have previously argued, the zero lower bound on net product creation (i.e. on $g_{I,t}$) means there may be a substantial asymmetry in the response to some shocks, and hence we impose this bound while generating impulse responses.[32]

3.4. Impulse responses

In this section, we present the impulse responses that result from IID (hence non-persistent) shocks to "ideas" (Z_t), labour supply (Φ_t), demand (Θ_t), population growth ($G_{N,t}$) and invention difficulty (L_t^I). Each set of graphs shows the impulse responses for one shock at magnitudes from 2% to 10% of their steady state level, given in terms of per cent deviations from the value the variable would have taken had the shock never arrived. For any variable V_t, we denote this by $V_t^{\%}$. Where there is significant asymmetry between positive and negative shocks we show both. In each graph, the horizontal axis shows time in quarters.

For each shock, the graphs are arranged in three rows of five. The top row (from right to left) shows frontier productivity, mark-ups in protected industries, the number of firms in these industries, the amount of research they each conduct and aggregate research expenditure ($\mathcal{E}_t^R$). The next row down shows aggregate productivity in non-protected industries, aggregate mark-ups, the number of products, labour supply and invention expenditure ($\mathcal{E}_t^I$). The final row shows aggregate productivity, wages, the proportion of industries that are non-protected, output and interest rates.

---

[27] In particular, our objective function is $\mu_t - \mu_t^P - G_{I,t}$. The global optimization was performed using the CMA-ES algorithm (Hansen et al. 2009).
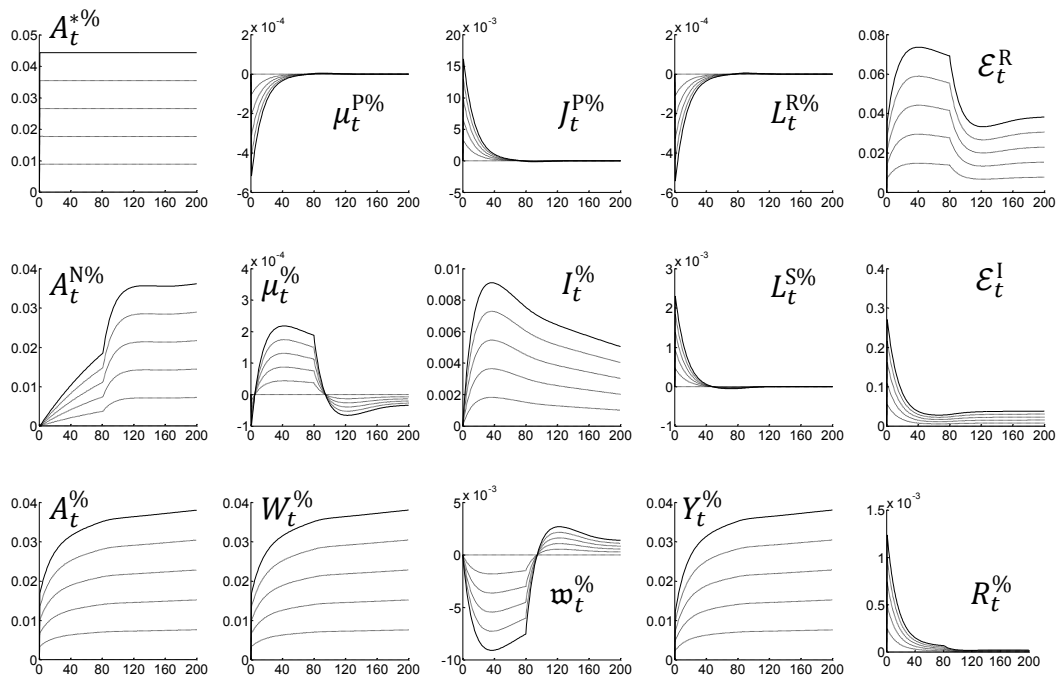
[28] Highlighted cells are standard values and are not calibrated.

[29] In steady state, for stochastic variables.

[30] With the exception of mark-ups, where we work in terms of logs of gross mark-ups instead.

[31] This was performed using Dynare (Adjemian et al. 2011).

[32] Our algorithm for doing this is described.

Figure 1: Response to a 2%,…,10% positive "idea" shock (i.e. a positive shock to Z_t)

Figure 1 plots the impulse response to a positive "idea" shock. This results in an immediate jump in frontier productivity ($A_t^{*\%}$) and a consequent initial jump in aggregate productivity ($A_t^{\%}$). In each subsequent period (quarter), a protected industry that benefited from this initial shock becomes non-protected, raising the aggregate productivity of non-protected industries ($A_t^{N\%}$). This leads to the slow increase in aggregate productivity ($A_t^{\%}$) we see over time. Indeed, aggregate productivity is still increasing at the end of our simulation period, fifty years after the initial shock, although most of its movement happens within twenty years. Mark-ups in protected industries ($\mu_t^{P\%}$) fall, counter-cyclically, but after an initial drop aggregate mark-ups ($\mu_t^{\%}$) increase. This is because the higher relative productivity of protected industries increases the profitability of invention, leading to more products ($I_t^{\%}$) and a fall in the proportion of industries that are non-protected ($\mathfrak{w}_t^{\%}$). This increase in invention rates also causes the lagged bump we see in the productivity of non-protected industries ($A_t^{N\%}$). The fact that invention was higher twenty years ago means there is a greater weight on relatively new industries in the aggregate productivity of non-protected industries, as the newest non-protected industries are producing products invented twenty years ago. In the real world, there are often substantial delays between patenting a product and that product going into production, which would tend to smooth out this kink. Additionally, in many countries, patents require regular maintenance fees to be paid in order for them to last the full twenty years, which would again smooth out this kink given idiosyncratic industry "idea" shocks. Finally, we see that in line with the evidence presented by Comin and Gertler (2006), both research expenditure ($\mathcal{E}_t^R$) and invention expenditure ($\mathcal{E}_t^I$) are pro-cyclical, the former because the increase in the total number of firms more than compensates for the drop in research.
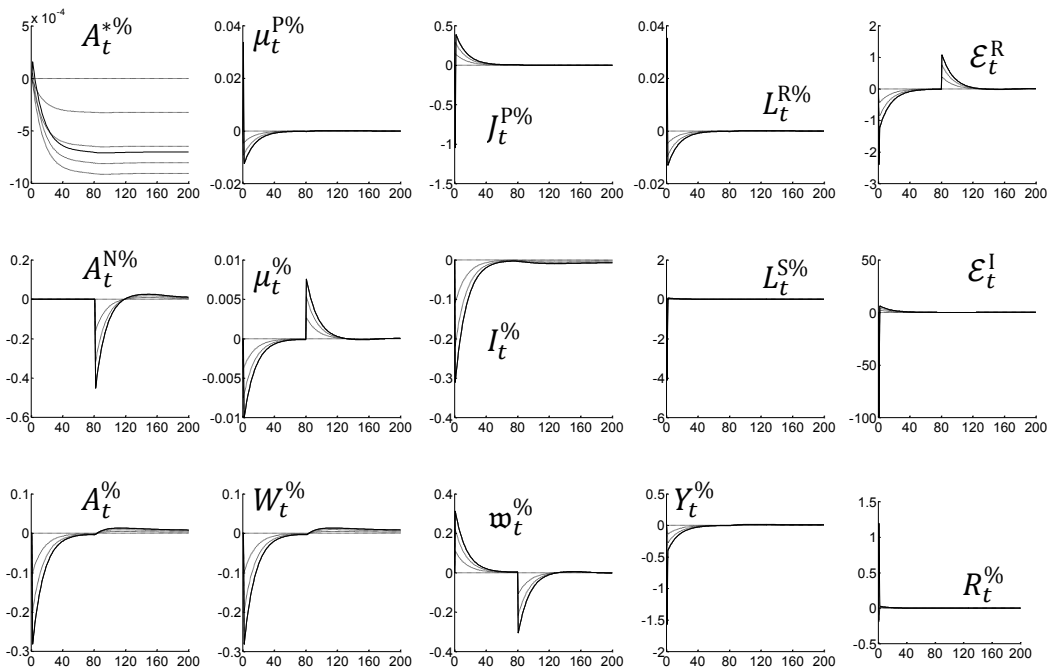
Figure 2: Response to a 2%,…,10% negative labour supply shock (i.e. a positive shock to Φ_t)
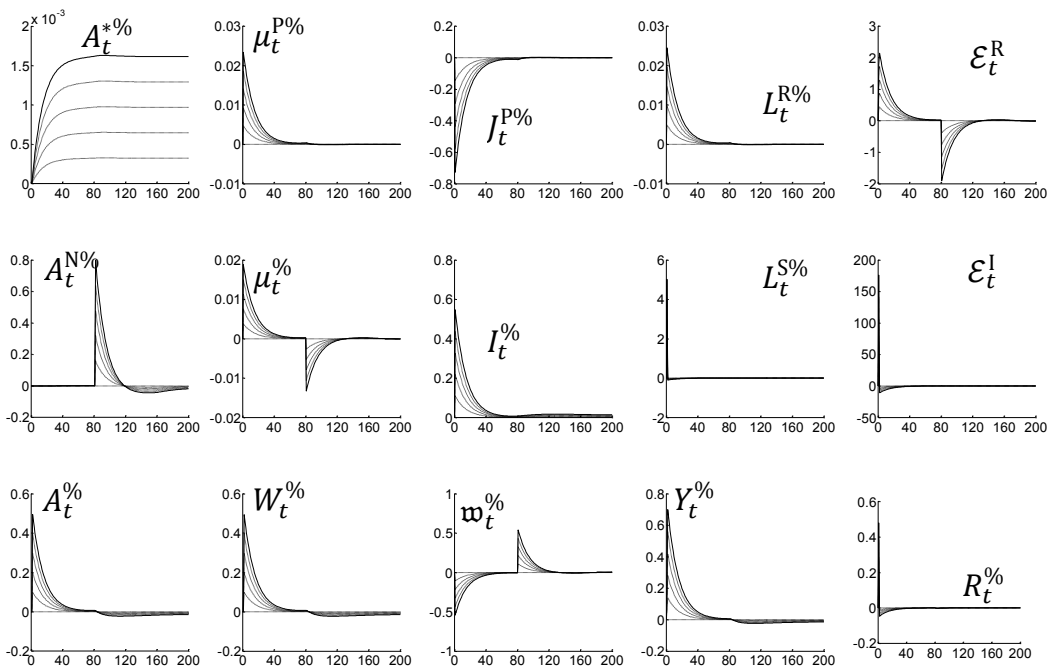
Figure 3: Response to a 2%,…,10% positive labour supply shock (i.e. a negative shock to Φ_t)

In Figure 2 and Figure 3, we plot the impulse responses to negative and positive labour supply shocks respectively. A large negative labour supply shock causes invention to drop so much that the stock of products ($I_t^\%$) hits its lower bound, which results in some asymmetry between positive and negative shocks. Invention falls because the labour supply shock means research and rent costs are higher compared to industry size, which must result in a drop in $I_t J_t(i)$ due to free entry. Were the number of firms per protected industry ($J_t^{P\%}$) to make most of the adjustment in the medium-term,

then mark-ups ($\mu_t^{\%}$) would rise and so wages ($W_t^{\%}$) would drop. However, an expected drop in wages will decrease invention today, since inventor returns are increasing in the expected future wage. Thus, there is no scenario in which a negative labour supply shock does not cause a fall in invention.

Following both a positive and a negative labour supply shock we see jumps in productivity followed by persistent returns, where the jump goes in the same direction as the initial labour supply shock. These are caused by the fact that an increase in invention decreases the proportion of industries that are producing non-protected products, reweighting aggregate productivity towards productive protected industries. Twenty years later this reweighting effect begins to go in the opposite direction since the unusually large industry cohort is now non-protected. This is particularly noticeable in the aggregate mark-up and the aggregate productivity of non-protected industries. We do not consider these twenty year lagged spikes in mark-ups to be particularly plausible, so in section 5 we show that a small modification to our model can produce much smoother processes.
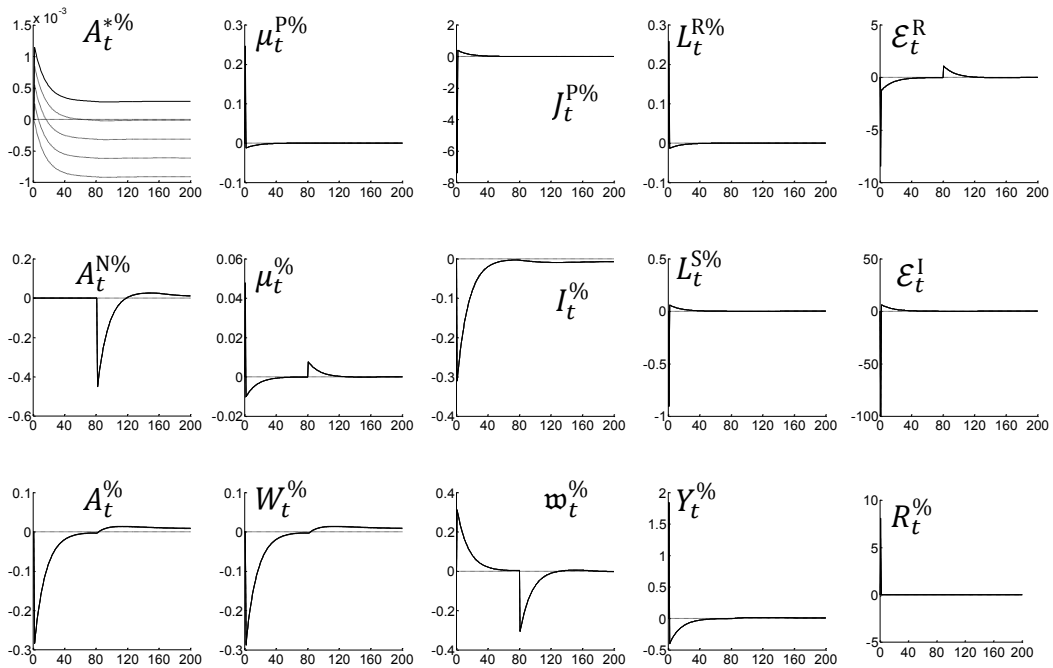
Figure 4: Response to a 2%,…,10% positive demand shock (i.e. a positive shock to Θ_t)

The contrast between positive and negative demand shocks (Figure 4 and Figure 5 respectively) is even more stark, with the positive demand shock having an almost identical effect at each magnitude considered, since hitting the bound almost perfectly undoes the effects of the shock. The mechanism behind their effects is identical to the labour supply shock though, with both having their initial effects through the free entry condition. This means that positive demand shocks are actually recessionary in the medium term.
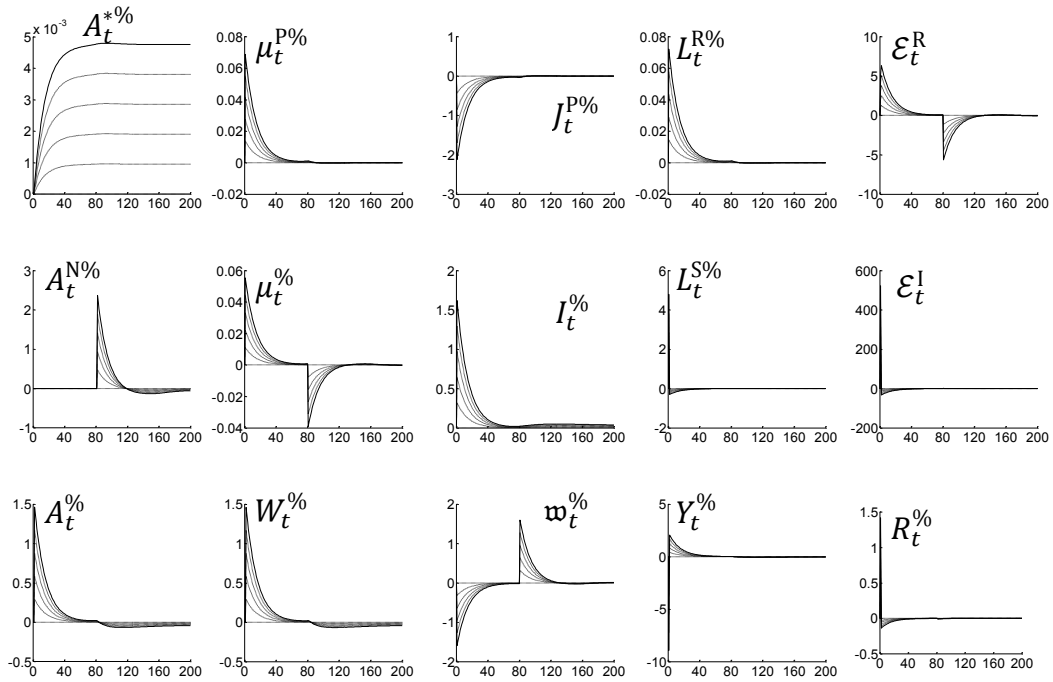
Figure 5: Response to a 2%,…,10% negative demand shock (i.e. a negative shock to Θ_t)

Positive and negative invention productivity shocks also have very similar effects to demand and labour supply shocks, with increases in invention productivity (decreases in its cost) resulting in surges in invention. These are shown in Figure 6 and Figure 7.
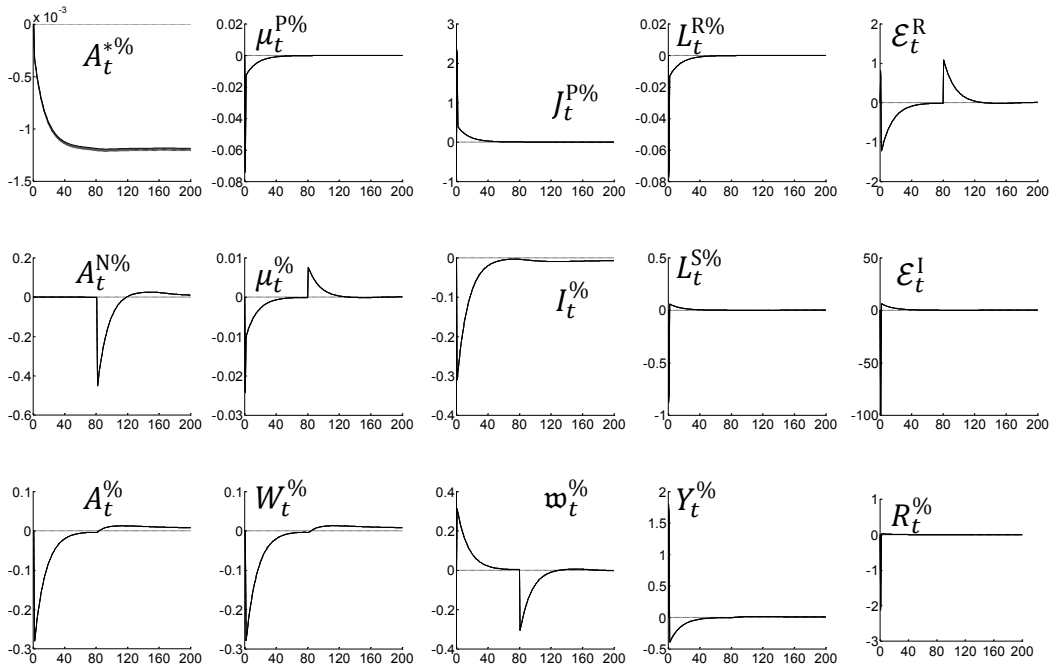
**Figure 6: Response to a 2%,…,10% negative invention productivity shock
(i.e. a positive shock to L_t^{I})**
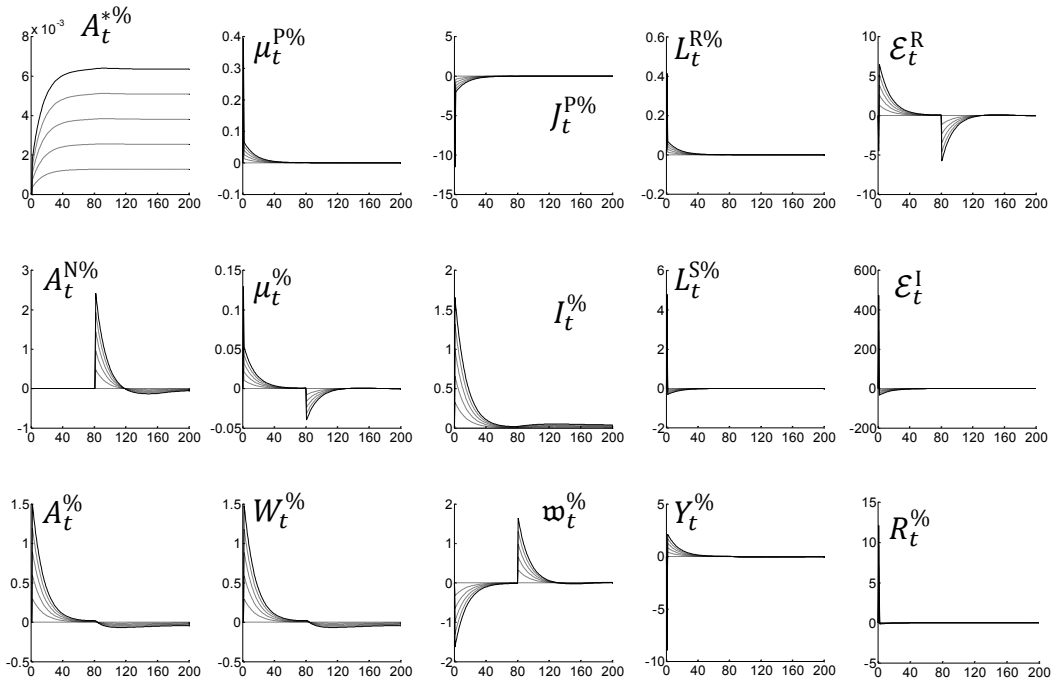
**Figure 7: Response to a 2%,…,10% positive invention productivity shock
(i.e. a negative shock to L_t^I)**
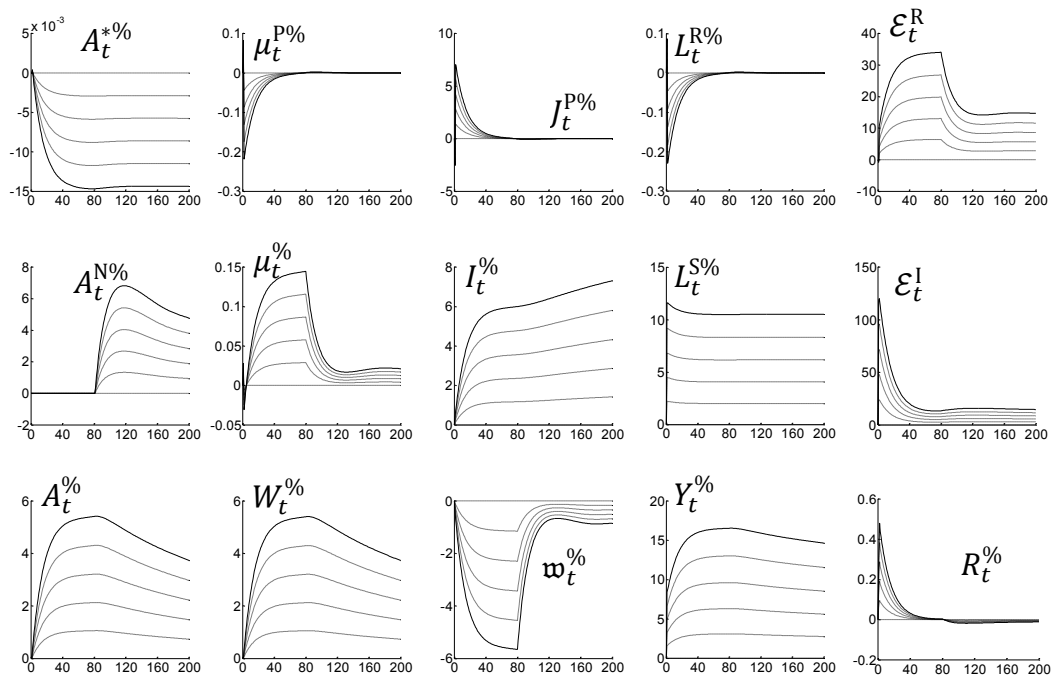
Figure 8: Response to a 2%,…,10% positive population shock (i.e. a positive shock to $G_{N,t}$)

We turn finally to population shocks, which also work through the invention channel. After a population shock arrives, aggregate demand is permanently higher. If the measure of products ($I_t^\%$) did not adjust, this would mean more firms in each industry ($J_t^{P\%}$), and hence that inventor returns would be higher. As a result, there must be an increase in invention. However, this will not happen instantaneously because greater invention results in higher wages ($W_t^\%$) both because of increased

demand, and because of the reweighting effect. These higher wages then push up the cost of invention. This is shown in Figure 8.

Interestingly, the "idea" shock was the only shock to cause a significant movement in the frontier growth rate, and even there only the direct effect is quantitatively important. This is despite the fact that we calibrated to maximize the responsiveness of frontier growth rates to business cycle shocks. In empirical studies, the identifying assumption is often made that, in the long-run, only technology shocks affect productivity. We see then that this assumption may be approximately correct, providing long enough horizons are taken. However, in our model, invention, population and preference shocks still have some effect on productivity at fifty-year horizons. This means that were we to attempt to apply the methodology of, for example, Beaudry and Lucke (2009) to simulated data from our model, using a similar number of sample periods to them (around fifty years), it seems highly improbable that the long run restrictions would succeed in recovering the original shocks. This intuition is supported by Gospodinov's (2010) proof of the inconsistency of the IRF estimator in an SVAR identified by long-run restrictions in which one of the processes is "local to unity" (i.e. has a near unit root). This lends support to the claim that preference shocks could be a major contributor to the shock labelled as a news shock by Beaudry and Lucke (2009). In a sense, preference shocks act as endogenous news shocks in our model.

4. Alternative calibration

The calibration of the previous section had set the rate at which research and invention get harder over time (ζ) to approximately zero, meaning that productivity had no significant effect on the equilibrium number of products. However, if we attempt to calibrate our model to match higher aggregate mark-ups, we are pushed towards higher values of ζ and hence lower invention growth rates. This is because ζ gives us an additional instrument with which to control invention expenditures, which otherwise would be too high since the higher mark-ups mean higher steady state rent levels. Recall though that $I_t = \hat{I}_t N_t A_t^{*-\zeta} \Psi$: therefore, at least in the long run, an increase in frontier productivity must reduce invention, introducing a new channel into our model.

In particular, we calibrate mark-ups to match Del Negro et al.'s (2007a) finding of an aggregate mark-up of 35%. Our other calibration targets are identical, with the exception that we no longer attempt to maximize the difference in mark-ups between protected and non-protected industries, in favour of getting invention rates as high as possible. This leads us to set:

Variable	ζ	λ	η	γ	$\wp$	L^I	g_N	β	ν
Value to 3 s.f.[33]	0.454	5.65	0.0611	2.89	0.131	2124	0.00311	0.99	2

---

[33] In steady state, for stochastic variables.

In Figure 9 we plot the impulse response to a positive "idea" shock under this alternative calibration. As before, the frontier productivity level ($A_t^{*\%}$) immediately jumps up, causing a similar initial jump in aggregate productivity ($A_t^{\%}$). This also means that in the very long run aggregate productivity must be higher. Again as before, the stream of products going out of patent protection transmits this jump in frontier productivity in to a slow initial rise in the aggregate productivity of non-protected industries ($A_t^{N\%}$). However, the jump in frontier productivity increases the difficulty of invention so much that the stock of products ($I_t^{\%}$) falls monotonically after the arrival of the shock. This leads to a higher proportion of products being non-protected ($\mathfrak{w}_t^{\%}$), and hence to the fall in aggregate productivity which begins immediately after the shock arrives. Additionally, after twenty years this fall in the invention rate means that a smaller proportion of non-protected industries will have relatively new technologies, and so the aggregate productivity of non-protected industries ($A_t^{N\%}$) will also start to fall.

Figure 9: Response to a 2%,…,10% positive "idea" shock (i.e. a positive shock to Z_t) under our alternative calibration

Most importantly though, the drop in invention causes a drop in labour supply ($L_t^{S\%}$) as a direct consequence of the drop in invention labour demand coupled with the relatively weak initial positive response of output. This gives a possible explanation for Gali's (1999) finding that positive technology shocks cause drops in hours worked.

5.   Smooth invention extension

Inventing a new product typically involves considering many possible ideas and absorbing a considerable amount of background material. Much of this will turn out to have been unnecessary,

but pursuing these dead-ends is unavoidable in the initial creative process. These dead-ends are unlikely to be a complete waste though, since the human capital acquired while pursuing them will tend to make future invention easier. Furthermore, the migration of staff will often make it difficult for the initial inventor to hold on to the value of this human capital, so the benefits may be shared by all future inventors.

Additionally, many product inventions lead directly to future inventions: for example, new drugs are often invented by slightly tweaking the chemical structure of another previously discovered drug. Again, the original inventor may not be able to internalise the value of these follow-on inventions if patents are sufficiently narrow and publication is mandatory.

This suggests modelling recent invention as making current invention easier for everyone. We expect similar results may be obtained without resorting to an externality, but internalisation of the value of these follow-on inventions introduces some technical complications.[34]

We make two changes to the stationary model of section 3. Firstly, we assume that there is a stock of "invention wisdom", s_t, that accumulates according to $s_t = (1 - \delta)\frac{L_t^I}{N_t} + \delta s_{t-1}$, where $\mathcal{L}_t^I$ is the total amount of labour used in invention at t and where δ controls the depreciation rate of this invention wisdom. Secondly, we assume that the cost of inventing a new product is given by $L_t^I \exp[-\alpha(s_{t-1} - \mathbb{E}s_{t-1})] A_t^{*\zeta}$, with $\alpha > 0$, so the more invention wisdom has been accumulated, the cheaper invention is. By subtracting the unconditional mean of s. we ensure that the steady-state of our original model is also a steady-state of this extended one under the same parameterization.

We set $\delta = 0.9$ which means the half-life of "invention wisdom" is around one and a half years, and we set $\alpha = 3$ which results in movements of invention costs of under one per-cent in response to reasonable shocks. The rest of our parameterization is identical to that given in section 3.2. Figure 10 plots the impulse response to a negative labour supply shock in our extended model with these parameter values.

In our main model, a negative labour supply shock resulted in a large negative initial spike in both productivity ($A_t^\%$) and mark-ups ($\mu_t^\%$), which was followed twenty years later by a large positive spike in mark-ups. All of these spikes were driven by the large initial drop in the measure of products. Thanks to the accumulation of "invention wisdom", the fall in invention is a lot smoother here, enabling us to replace the sudden spikes with smooth, medium-frequency oscillations.

---

[34] We investigated several specifications in which inventors can hold onto the returns from follow-on inventions; however, they all resulted in aggregate indeterminacy. This may be explained intuitively. If there is a lot of invention today then the stream of cheap follow-on inventions will reduce invention labour demand in the future, causing labour substitution into production and higher consumption. This in turn increases the value of those follow-on inventions, thus encouraging inventors to invent now, to capture the higher value of the follow-on inventions. While aggregate indeterminacy is not implausible, introducing sunspots here would be an unnecessarily large departure.
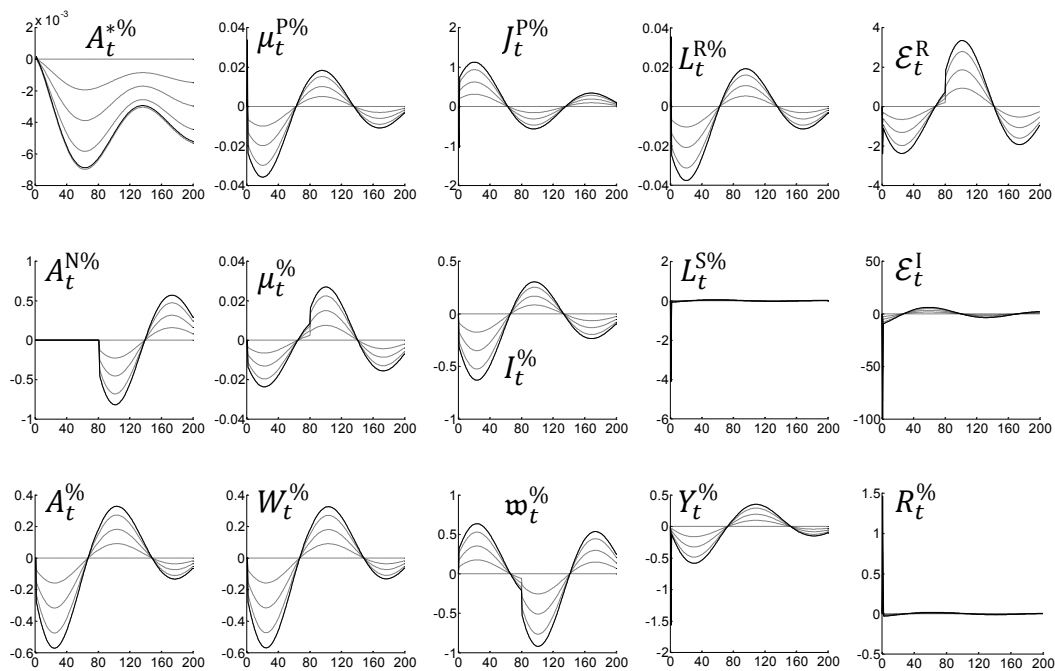
Figure 10: Response to a 2%,…,10% negative labour supply shock (i.e. a positive shock to Φ_t) in our extended model

6.   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"[35]. 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. We showed that old technologies become enshrined in industries producing non patent-protected products, which leads to gradual increases in aggregate productivity in response to process innovation shocks. We also showed that almost all shocks lead to changes in the rate of product invention that have significant consequences for aggregate productivity and mark-ups, due to fluctuations in the proportion of industries that are producing patent-protected products. Our model's propagation mechanisms thus lend persistence to all shocks, not just productivity ones.

All of this means that exogenous models of productivity may be missing important effects, even over the short to medium term. No matter how persistent an exogenous process productivity is assumed to be, unless ad hoc assumptions are made about its correlations with other shocks, productivity will never move following a preference shock. Additionally, whereas introducing news

---

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

about future productivity into standard models requires arbitrary assumptions about the timing and accuracy of that news, our model generates news about future productivity movements endogenously from preference shocks. This means we may reinterpret the results of Beaudry and Lucke (2009) as providing evidence for the major role of preference shocks in business cycle fluctuations. Furthermore, we showed that under some calibrations, our model generates a negative response of hours to productivity shocks, something that is impossible with exogenous productivity unless frictions are added. Finally, we demonstrated that a small change to our core model results in smooth medium-frequency cycles in output and productivity.

Our model suggests that a switch to indefinite patent protection might 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 leaving its growth rate unchanged and only slightly increasing mark-ups. A less radical policy change would 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. Additionally, we wish to assess the extent to which the non-separable preferences of Jaimovich and Rebelo (2009) can help our model to generate a positive co-movement in output and hours in response to a wider range of shocks, given that our model generates endogenous news shocks. We are also keen to integrate some of the standard frictions from new-Keynesian models, to investigate the interaction between sticky-price driven counter-cyclicality in mark-ups and the research channel.

Our model has many strong testable implications and while some of these may not be robust to the inclusion of frictions or non-separable preferences, others certainly should be, not least the differences between protected and non-protected industries. We hope to test these implications in future empirical work, along with quantifying the importance of the channels presented here.

7.   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, April. http://ideas.repec.org/p/cpm/dynare/001.html.

Aghion, Philippe, and Peter Howitt. 1992. A Model of Growth Through Creative Destruction. *Econometrica* 60, no. 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, no. 2: 300-323.

Beaudry, Paul, and Bernd Lucke. 2009. Letting Different Views about Business Cycles Compete. *National Bureau of Economic Research Working Paper Series* No. 14950 (May). http://www.nber.org/papers/w14950.

Bilbiie, Florin O., Fabio Ghironi, and Marc J. Melitz. 2007. Monetary Policy and Business Cycles with Endogenous Entry and Product Variety. *National Bureau of Economic Research Working Paper Series* No. 13199 (June). http://www.nber.org/papers/w13199.

Boulhol, Hervé. 2008. The Upward Bias of Markups Estimated from the Price-Based Methodology. *Annales d'Économie et de Statistique*, no. 89 (March): 131-156.

Broda, Christian, and David E Weinstein. 2010. Product Creation and Destruction: Evidence and Price Implications. *American Economic Review* 100, no. 3 (June): 691-723.

Comin, Diego. 2009. On the integration of growth and business cycles. *Empirica* 36, no. 2 (May 1): 165-176.

Comin, Diego, and Mark Gertler. 2004. Procyclical Speed of Technological Diffusion. Unpublished paper, cited in Comin and Gertler (2006).

———. 2006. Medium-Term Business Cycles. *American Economic Review* 96, no. 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, June. RePEc. http://ideas.repec.org/p/nbr/nberwo/15029.html.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters. 2005. *On the fit and forecasting performance of New-Keynesian models*. Working Paper Series. European Central Bank, June. http://ideas.repec.org/p/ecb/ecbwps/20050491.html.

———. 2007a. On the Fit of New Keynesian Models. *Journal of Business & Economic Statistics* 25: 123-143.

———. 2007b. Rejoinder. *Journal of Business and Economic Statistics* 25, no. 2: 159.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. Monopolistic Competition and Optimum Product Diversity. *The American Economic Review* 67, no. 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. http://ideas.repec.org/p/tor/tecipa/tecipa-348.html.

Ellis, Colin. 2006. *Elasticities, Markups and Technical Progress: Evidence from a State-Space Approach*. SSRN. http://ssrn.com/paper=932580.

Ethier, Wilfred J. 1982. National and International Returns to Scale in the Modern Theory of International Trade. *The American Economic Review* 72, no. 3 (June): 389-405.

Gali, Jordi. 1999. Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? *American Economic Review* 89, no. 1 (March): 249-271.

Gospodinov, Nikolay. 2010. Inference in Nearly Nonstationary SVAR Models With Long-Run Identifying Restrictions. *Journal of Business and Economic Statistics* 28, no. 1 (January): 1-12.

Groth, Christian, Karl-Josef Koch, and Thomas Steger. 2009. When economic growth is less than exponential. *Economic Theory* (July 4). http://dx.doi.org/10.1007/s00199-009-0480-y.

Hansen, N., A.S.P. Niederberger, L. Guzzella, and P. Koumoutsakos. 2009. A Method for Handling Uncertainty in Evolutionary Optimization With an Application to Feedback Control of Combustion. *Evolutionary Computation, IEEE Transactions on* 13, no. 1: 180-197.

Jaimovich, Nir. 2007. Firm dynamics and markup variations: Implications for sunspot equilibria and endogenous economic fluctuations. *Journal of Economic Theory* 137, no. 1 (November): 300-325.

Jaimovich, Nir, and Sergio Rebelo. 2009. Can News about the Future Drive the Business Cycle? *American Economic Review* 99, no. 4 (September): 1097-1118.

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). http://dx.doi.org/10.1016/j.jedc.2010.04.003.

Mansfield, Edwin. 1993. The diffusion of industrial robots in Japan and the United States. *Research Policy* 22, no. 2 (April): 105.

Nekarda, Christopher, and Valerie Ramey. 2010. The Cyclical Behavior of the Price-Cost Markup. http://econ.ucsd.edu/~vramey/research/markupcyc.pdf.

Peretto, Pietro F. 1999. Cost reduction, entry, and the interdependence of market structure and economic growth. *Journal of Monetary Economics* 43, no. 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, no. 11 (November): 1885-1913.

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

Roeger, Werner. 1995. Can Imperfect Competition Explain the Difference between Primal and Dual Productivity Measures? Estimates for U.S. Manufacturing. *The Journal of Political Economy* 103, no. 2 (April): 316-330.

Rubinstein, Ariel. 1982. Perfect Equilibrium in a Bargaining Model. *Econometrica* 50, no. 1 (January): 97-109.

Schmitt-Grohé, Stephanie, and Martín Uribe. 2008. What's News in Business Cycles. *National Bureau of Economic Research Working Paper Series* No. 14215 (August). http://www.nber.org/papers/w14215.

The Associated Press. 2008. Court Rules Business Concept Cannot Be Patented. *The New York Times*, October 31, sec. Business. http://www.nytimes.com/2008/10/31/business/31patent.html?_r=1.

Wälde, Klaus. 2005. Endogenous Growth Cycles. *International Economic Review* 46, no. 3: 867-894.

World Trade Organisation, The. 1994. Agreement on Trade-Related Aspects of Intellectual Property Rights. http://www.wto.org/english/docs_e/legal_e/27-trips.pdf.

8. Appendices

8.1. The free-entry and first order conditions

Let $m_t^{\mathrm{R}}(i,j)W_t$ be the Lagrange multiplier on research's positivity constraint and $m_t^{\mathrm{A}}(i,j)W_t$ be the Lagrange multiplier on appropriation's positivity constraint. Then in a symmetric equilibrium the two first order conditions and the free entry condition (respectively) mean:

$$\beta \frac{1}{I_t J_t(i)} \frac{\mu_t(i)}{1+\mu_t(i)} \left(\frac{1+\mu_t}{1+\mu_t(i)}\right)^{\frac{1}{\lambda}} \mathbb{E}_t \Xi_{t+1} Y_{t+1} \left(\frac{A_{t+1}(i)}{A_{t+1}}\right)^{\frac{1}{\lambda}} \frac{1}{\mu_t(i)} \frac{Z_{t+1} A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi}{1+\gamma Z_{t+1} A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi L_t^{\mathrm{R}}(i)}$$
$$= W_t \left(1 - m_t^{\mathrm{R}}(i)\right)$$

$$\beta \frac{1}{I_t J_t(i)} \frac{\mu_t(i)}{1+\mu_t(i)} \left(\frac{1+\mu_t}{1+\mu_t(i)}\right)^{\frac{1}{\lambda}} \mathbb{E}_t \Xi_{t+1} Y_{t+1} \left(\frac{A_{t+1}(i)}{A_{t+1}}\right)^{\frac{1}{\lambda}} \frac{1}{\mu_t(i)} \frac{1+(\gamma-\zeta^{\mathrm{R}}) Z_{t+1} A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi L_t^{\mathrm{R}}(i)}{1+\gamma Z_{t+1} A_t^{**}(i)^{-\zeta^{\mathrm{R}}}\Psi L_t^{\mathrm{R}}(i)}$$
$$\cdot \frac{1}{\tau} \frac{A_t(i)^{-\zeta^{\mathrm{A}}}\Upsilon(A_t^{*\tau} - A_t(i)^\tau)}{A_t^{**}(i)^\tau \left(1 + A_t(i)^{-\zeta^{\mathrm{A}}}\Upsilon L_t^{\mathrm{A}}(i)\right)^2} = W_t \left(1 - m_t^{\mathrm{A}}(i)\right)$$

$$\beta \frac{1}{I_t J_t(i)} \frac{\mu_t(i)}{1+\mu_t(i)} \left(\frac{1+\mu_t}{1+\mu_t(i)}\right)^{\frac{1}{\lambda}} \mathbb{E}_t \Xi_{t+1} Y_{t+1} \left(\frac{A_{t+1}(i,j)}{A_{t+1}(i)}\right)^{\frac{1}{\eta\lambda}} \left(\frac{A_{t+1}(i)}{A_{t+1}}\right)^{\frac{1}{\lambda}}$$
$$= \left[L_t^{\mathrm{R}}(i,j) + L_t^{\mathrm{A}}(i,j) + L_t^{\Re}(i) + L_t^{\mathrm{F}}\right] W_t$$

(where we have dropped j indices on variables which are the same across the industry).

That the solution for research when $Z_{t+1} \equiv 1$ is given by equation (2.2) is a trivial consequence of the complementary slackness condition and the fact that $\frac{1}{\mu_t(i)} < \gamma$. Deriving (2.3) is less trivial though.

Begin by defining $\mathcal{k}_t(i) := \frac{1+(\gamma-\zeta^R)\mathfrak{L}_t^R(i)}{1+\gamma\mathfrak{L}_t^R(i)}$, and note that since we are assuming $\gamma > \zeta^R \geq 0$, we have that $0 < \mathcal{k}_t(i) \leq 1$.

Also define:

$$n_t(i) := \frac{\mathcal{k}_t(i)}{\tau\mu_t(i)} A_t^*(i)^{-\zeta^A} \Upsilon \left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right] \left[L_t^R(i) + L_t^\mathfrak{R}(i) + L_t^F\right] \geq 0,$$

which is not a function of $L_t^A(i)$, given $L_t^R(i)$. We can then combine the appropriation first order condition with the free entry condition to obtain:

$$\frac{1}{\left(1+\mathfrak{L}_t^A(i)\right)^2} \left(\frac{A_t^*(i)}{A_t^{**}(i)}\right)^\tau \left[\frac{\mathcal{k}_t(i)}{\tau\mu_t(i)}\left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right]\mathfrak{L}_t^A(i) + n_t(i)\right] = 1 - m_t^A(i).$$

Since the left hand side is weakly positive, from the dual feasibility condition we know $m_t^A(i) \in [0,1]$. Now when $L_t^A(i) = 0$, this becomes:

$$n_t(i) = 1 - m_t^A(i),$$

since in this case $A_t^*(i) = A_t^{**}(i)$. Therefore when $L_t^A(i) = 0$, $n_t(i) \leq 1$.

We now prove the converse. Suppose then for a contradiction that $L_t^A(i) > 0$, but $n_t(i) \leq 1$. By complementary slackness we must have $m_t^A(i) = 0$, hence:

$$1 \geq n_t(i) = \left(1+\mathfrak{L}_t^A(i)\right)^2 \left(\frac{A_t^{**}(i)}{A_t^*(i)}\right)^\tau - \frac{\mathcal{k}_t(i)}{\tau\mu_t(i)}\left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right]\mathfrak{L}_t^A(i)$$

$$\geq \left(1+\mathfrak{L}_t^A(i)\right)^2 \left(\frac{A_t^{**}(i)}{A_t^*(i)}\right)^\tau - \left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right]\mathfrak{L}_t^A(i)$$

$$= \left(1+\mathfrak{L}_t^A(i)\right)\left[\left(1+\mathfrak{L}_t^A(i)\right) + \mathfrak{L}_t^A(i)\left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right]\right] - \left[\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau - 1\right]\mathfrak{L}_t^A(i),$$

where we have used the facts that $\mathcal{k}_t(i) \leq 1$ and $\frac{1}{\mu_t(i)} < \tau$ to derive the second inequality. Expanding the brackets then gives that:

$$1 \geq 1 + 2\mathfrak{L}_t^A(i) + \left(\frac{A_t^*}{A_t^*(i)}\right)^\tau \mathfrak{L}_t^A(i)^2,$$

i.e. that $0 \geq 2 + \left(\frac{A_t^*}{A_t^*(i)}\right)^\tau \mathfrak{L}_t^A(i)$ which is a contradiction as $\left(\frac{A_t^*}{A_t^*(i)}\right)^\tau \mathfrak{L}_t^A(i) \geq 0$.

We have proven then that providing $\frac{1}{\mu_t(i)} < \tau$, $L_t^A(i) = 0$ if and only if $n_t(i) \leq 1$. It just remains for us to solve for $L_t^A(i)$ when it is strictly positive. From the above, we have that, in this case:

$$\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau [n_t(i) - 1] = 2\left[1 - \frac{1}{2}\left[1 + \frac{\hbar_t(i)}{\tau\mu_t(i)}\right]\left[1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^\tau\right]\right]\mathfrak{L}_t^A(i) + \mathfrak{L}_t^A(i)^2.$$

Hence:

$$\mathfrak{L}_t^A(i) = -\left[1 - \frac{1}{2}\left[1 + \frac{\hbar_t(i)}{\tau\mu_t(i)}\right]\left[1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^\tau\right]\right]$$

$$+ \sqrt{\left[1 - \frac{1}{2}\left[1 + \frac{\hbar_t(i)}{\tau\mu_t(i)}\right]\left[1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^\tau\right]\right]^2 + \left(\frac{A_t^*(i)}{A_t^*}\right)^\tau [n_t(i) - 1]},$$

since the lower solution is guaranteed to be negative as $n_t(i) > 1$ when $L_t^A(i) > 0$.

8.2. The steady state for non-patent-protected industries

In an industry i which is not patent-protected and in which appropriation, but no research, is performed, from (2.1) and (2.3):

$$-\mathfrak{K}_t(i) + \sqrt{\mathfrak{K}_t(i)^2 + \mathfrak{U}_t(i)} = \mathfrak{L}_t^A(i) = \left[1 - \frac{\left(\frac{A_{t+1}^*(i)}{A_t^*(i)}\right)^\tau - 1}{1 - \left(\frac{A_t^*(i)}{A_t^*}\right)^\tau}\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau\right]^{-1} - 1.$$

If we treat $\mathfrak{p}_1 := \tau\mu_t(i) - 1 \approx 0$, $\mathfrak{p}_2 := A_t^*(i)^{-\zeta^A}\Upsilon_t L_t^F \approx 0$ and $\mathfrak{p}_3 := \left(\frac{A_{t+1}^*(i)}{A_t^*(i)}\right)^\tau - 1 \approx 0$ as fixed, this leaves us with a cubic in $\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau$, for which only one solution will be feasible (i.e. strictly less than 1). Taking a second order Taylor approximation of this solution in $\mathfrak{p}_1$, $\mathfrak{p}_2$ and $\mathfrak{p}_3$, reveals (after some messy computation), that:

$$\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau \approx \mathfrak{p}_2\big(1 - (\mathfrak{p}_1 + \mathfrak{p}_2)\big) = A_t^*(i)^{-\zeta^A}\Upsilon_t L_t^F\big(2 - \tau\mu_t(i) - A_t^*(i)^{-\zeta^A}\Upsilon_t L_t^F\big)$$

(The effect of $\mathfrak{p}_3$ on $\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau$ is third order and hence it does not appear in this expression.)

From this approximate solution for $\left(\frac{A_t^*(i)}{A_t^*}\right)^\tau$ then, we have that the relative productivity of a non-protected industry is decreasing in its mark-up. Furthermore, from dropping to a first order

approximation, we have that $A_t^*(i)^{1+\frac{\zeta^A}{\tau}} \approx A_t^*(\Upsilon_t L_t^F)^{\frac{1}{\tau}}$, so asymptotically non-protected industries are growing at $\left[1 + \frac{\zeta^A}{\tau}\right]^{-1}$ times the growth rate of the frontier.

8.3. The inventor-firm bargaining process

We model the entire process of setting and paying rents as follows:

1) Firms enter, paying the fixed cost.

2) Firms who have entered conduct appropriation, then research.

3) The "idea shock" for next period's production, Z_{t+1}, is realised and firms and patent holders learn its level.

4) Finally, firms arrive at the patent-holder to conduct bargaining, with these arrivals taking place sequentially but in a random order. (For example, all firms phone the patent-holder sometime in the week before production is to begin.) In this bargaining we suppose that the patent-holder has greater bargaining power, since they have a longer outlook[36] and since they lose nothing if bargaining collapses[37]. We also suppose that neither patent-holders nor firms are able to observe or verify either how many (other) firms paid the fixed cost or what research and appropriation levels they chose. This is plausible because until production begins it is relatively easy to keep such things hidden (for example, by purchasing the licence under a spin-off company), and because it is hard to ascertain ahead of production exactly what product a firm will be producing. We assume bargaining takes an alternating offer form, (Rubinstein 1982) but that it happens arbitrarily quickly (i.e. in the no discounting limit).

5) Firms pay the agreed rents if bargaining was successful. Since this cost is expended before production, we continue to suppose firms have to borrow in the period before production in order to cover it. Firms will treat it as a fixed cost, sunk upon entry, since our unobservability assumptions mean bargaining's outcome will not be a function of research and appropriation levels.

6) The next period starts, other aggregate shocks are realised and production takes place.

7) The patent-holder brings court cases against any firms who produced but decided not to pay the rent. For simplicity, we assume the court always orders the violating firm to pay damages to the patent-holder, which are given as follows:

---

[36] Consider what happens as the time gap between offers increases. When this gap is large enough only one offer would be made per-period, meaning the patent-holder would make a take-it-or-leave-it offer giving (almost) nothing to the firm, which the firm would then accept.

[37] The firm owner may, for example, face restrictions from starting businesses in future if as a result of the bargaining collapse they are unable to repay their creditors.

a) When the courts believe rents were not reasonable (i.e. $L_t^{\mathfrak{R}}(i) > L_t^{\mathfrak{R}*}(i)$, where $L_t^{\mathfrak{R}*}(i)W_t$ is the level courts determine to be "reasonable royalties"), they set damages greater than $L_t^{\mathfrak{R}*}(i)W_t$, as *"the infringer would have nothing to lose, and everything to gain if he could count on paying only the normal, routine royalty non-infringers might have paid"*[38]. We assume excess damages over $L_t^{\mathfrak{R}*}(i)W_t$ are less than the patent-holder's legal costs however.

b) When the courts consider the charged rent to have been reasonable (i.e. $L_t^{\mathfrak{R}}(i) \leq L_t^{\mathfrak{R}*}(i)$) the courts award punitive damages of more than $\max\left\{ L_t^{\mathfrak{R}*}(i)W_t, \left(\frac{1}{1-\wp}\right) L_t^{\mathfrak{R}}(i)W_t, \right\}$, where $\wp$ is the bargaining power of the firm, in the sense of the generalized Nash bargaining solution.[39]

Under this specification:

$$L_t^{\mathfrak{R}}(i) = \min\left\{ L_t^{\mathfrak{R}*}(i), (1-\wp)\left[L_t^{R}(i) + L_t^{A}(i) + L_t^{\mathfrak{R}}(i) + L^{F} \right] \right\}$$

since entry is fixed when bargaining takes place, since patent-holders know that bargaining to a rent level any higher than $L_t^{\mathfrak{R}*}(i)W_t$ will just result in them having to pay legal costs,[40] and since $\left[L_t^{R}(i) + L_t^{A}(i) + L_t^{\mathfrak{R}}(i) + L^{F} \right]W_t$ is equal to the production period profits of each firm in industry i, by the free entry condition.[41] Therefore, in equilibrium:

$$L_t^{\mathfrak{R}}(i) = \min\left\{ L_t^{\mathfrak{R}*}(i), L_t^{\mathfrak{R}\dagger}(i) \right\}, \tag{8.1}$$

where $L_t^{\mathfrak{R}\dagger}(i)$ is a solution to equations (2.2), (2.3) along with equation (2.4), (i.e. $L_t^{\mathfrak{R}}(i) = \frac{1-\wp}{\wp}\left[L_t^{R}(i) + L_t^{A}(i) + L^{F} \right]$) if one exists, or $+\infty$ otherwise. Because damages are always greater than $L_t^{\mathfrak{R}*}(i)W_t$, these rents will be sufficiently low to ensure firms are always prepared to licence the patent at the bargained price in equilibrium.

Now suppose we are out of equilibrium and fewer firms than expected have entered. Since neither the patent-holder nor firms can observe how many firms have entered, and since firms arrive at the patent-holder sequentially, both sides will continue to believe that the equilibrium number of firms has entered and so rents will not adjust. On the other hand, suppose that (out of equilibrium) too

---

[38] Panduit Corp. v. Stahlin Brothers Fibre Works, Inc., 575 F.2d 1152, 1158 (6th Circuit 1978), cited in Pincus (1991).

[39] The level $\left(\frac{1}{1-\wp}\right) L_t^{\mathfrak{R}}(i)W_t$ is chosen to ensure that, with equilibrium rents, firms prefer not to produce at all rather than to produce without paying rents.

[40] The disagreement point is zero since it is guaranteed that $L_t^{\mathfrak{R}}(i) \leq L_t^{\mathfrak{R}*}(i)$ and so punitive damages would be awarded were the firm to produce without paying rents, which, by construction, leaves them worse off than not producing.

[41] A similar expression can also be derived if we assume instead that courts guarantee infringers a fraction $\wp$ of production profits, or if we assume courts always award punitive damages but firms are able to hide a fraction $\wp$ of their production profits.

many firms enter. When the first unexpected firm arrives at the patent-holder to negotiate, the patent-holder will indeed realise that too many firms have entered. However, since the firm they are bargaining with has no way of knowing this,[42] the patent-holder can bargain for the same rents as in equilibrium. Therefore, even out of equilibrium:

$$L_t^{\Re}(i) = \min\{L_t^{\Re*}(i), L_t^{\Re\dagger}(i)\}$$

where we stress $L_t^{\Re\dagger}(i)$ is not a function of the decisions any firm happened to take. This ensures that any solution of equations (2.2), (2.3) and (8.1) for research, appropriation and rents will also be an equilibrium, even allowing for the additional condition that the derivative of firm profits with respect to the number of firms must be negative at an optimum.

We now just have to pin down "reasonable royalties", $L_t^{\Re*}(i)W_t$. Certainly it must be the case that $L_t^{\Re*}(i) \le L_t^{\bar{\Re}}(i)$, where $L_t^{\bar{\Re}}(i)$ is the level of rents at which $J_t(i) = 1$, since rents so high that no one is prepared to pay them must fall foul of the courts' desire to ensure licensees can make a profit. [43] However, since when $J_t(i) = 1$ the sole entering firm (almost) may as well be the patent-holder themselves, where possible the courts will set $L_t^{\Re*}(i)$ sufficiently low to ensure that $J_t(i) > 1$ in equilibrium, again following the idea that licensees ought to be able to make a profit. When there is a $J_t(i) > 1$ solution to equations (2.2), (2.3) and (2.4) already (i.e. $L_t^{\Re\dagger}(i) < \infty$), the courts will just set $L_t^{\Re*}(i)$ at the rent level that would obtain in that solution, thus preventing the possibility of $J_t(i) = 1$ being an equilibrium. It may be shown that for sufficiently large t such a solution is guaranteed to exist, so in this case $L_t^{\Re*}(i) = L_t^{\Re\dagger}(i) = L_t^{\Re}(i)$.[44]

8.4.   Dealing with the zero lower bound on net product creation

The following algorithm generates correct impulse responses, although it cannot be used for path simulations. (We believe this algorithm to be novel and so it may perhaps also prove useful for imposing the zero lower bound on interest rates in medium-scale new-Keynesian models.)

---

[42] Either they are a firm that thinks the equilibrium number of firms has entered, or they are a firm that thinks more than the equilibrium number of firms has entered, but that does not know whether the patent-holder has yet realised this.

[43] "…*the very definition of a reasonable royalty assumes that, after payment, the infringer will be left with a profit.*" Georgia-Pacific Corp. v. U.S. Plywood-Champion Papers Corp., 446 F.2d 295, 299 & n.1 (2d Cir.), *cert. denied,* 404 U.S. 870 (1971), cited in Pincus (1991).

[44] There may still be multiple solutions for rents (as (2.2), (2.3) and (2.4) might have multiple solutions), but of these only the one with minimal entry is really plausible, since this is both weakly Pareto dominant (firms always make zero profits and it may be shown that the patent-holder prefers minimal entry) and less risky for entering firms (if entering firms are unsure if the patent-holder will play the high rent or the low rent equilibrium, they are always better off assuming the high rent one since if that assumption is wrong they make strict profits, whereas had they assumed low rents but rents were in fact high they would make a strict loss).

1) Generate the full set of positive and negative IRFs ignoring the bound. Make a note of those IRFs in which the bound was hit. (In our model the bound is only hit after a positive shock to Φ_t.)

2) For each shock ϵ^* for which the bound was hit:

   a. Choose a large number $\mathcal{T}$ after which you believe the constraint will no longer bind. (We set $\mathcal{T} = 40$ which was certainly sufficient.)

   b. Edit the equation determining the constrained variable to include a "shadow price" term equal to $\sum_{s=0}^{\mathcal{T}-1} \epsilon_{s,t-s}^{SP}$ (taking exponents if appropriate), where $\epsilon_{s,t}^{SP} \sim \text{NIID}(0,1)$ if $t = 0$ and is 0 otherwise. Note that each of the shocks in the sum thus becomes known in period 0, but does not hit the equation until period s. (For us, since we work with the log of the measure of industries, the inventor entry equation becomes: $L_t^I \widehat{W}_t = \frac{1-\wp}{\wp} \mathfrak{I}_{t,0} \exp\left[\sum_{s=0}^{\mathcal{T}-1} \epsilon_{s,t-s}^{SP}\right]$.)

   c. Generate impulse responses for this extended model for shocks coming from each of the $\epsilon_{s,\cdot}^{SP}$ for the variable of interest.

   d. Let v be the relative impulse response of the zero bounded variable to a ϵ^* shock (as a column vector) and let M be the matrix formed from horizontally concatenating the relative impulse responses to each $\epsilon_{s,\cdot}^{SP}$ (as column vectors). Let M^* be the square sub-matrix of M with an identical top row. Let m be the steady state value of the zero bounded variable.

   e. Solve the following quadratic optimization problem:

   $$\alpha^* = \underset{\substack{\alpha \geq 0, \\ v + M\alpha \geq 0}}{\text{argmin}}[\alpha'(m + v + M^*\alpha)] = \underset{\substack{\alpha \geq 0, \\ v + M\alpha \geq 0}}{\text{argmin}}\left[\alpha'(m + v) + \frac{1}{2}\alpha'(M^* + M^{*\prime})\alpha\right]$$

   where α should be thought of as determining the linear combination $(\sum_{s=0}^{\mathcal{T}-1} \alpha_{s+1}\epsilon_{s,t-s}^{SP})$ of the $\epsilon_{s,\cdot}^{SP}$ shocks to which the imposition of the bound is equivalent. (In our model it turns out that only the first element of α^* is positive.)

   f. If the minimand is strictly positive at the optimum, α^* cannot be used. In this case increase $\mathcal{T}$ and go back to b (or try a different minimization algorithm). Otherwise, continue.

3) The desired impulse response is then given by $v + M\alpha^*$.

To see why this works, recall that by construction $\alpha^{*\prime}(m + v + M^*\alpha^*) = 0$. Therefore, it must be the case that for each s, either $\alpha_{s+1}^* = 0$ so the shock $\epsilon_{s,\cdot}^{SP}$ is getting zero weighting, or that the zero lower bound binds in period s. Therefore, these "shadow price shocks" are only being introduced when the positivity constraint does indeed bind. As the full future history of "shadow price shocks" is known as soon as the ϵ^* shock hits, these shocks are consistent with rational expectations.