



Endogenous Entry, Product Variety, and Business Cycles

Author(s): Florin O. Bilbiie, Fabio Ghironi, and Marc J. Melitz

Reviewed work(s):

Source: *Journal of Political Economy*, Vol. 120, No. 2 (April 2012), pp. 304-345

Published by: [The University of Chicago Press](http://www.press.uchicago.edu)

Stable URL: <http://www.jstor.org/stable/10.1086/665825>

Accessed: 07/06/2012 09:24

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Political Economy*.

Endogenous Entry, Product Variety, and Business Cycles

Florin O. Bilbiie

Paris School of Economics, Université Paris 1 Panthéon-Sorbonne, and Centre for Economic Policy Research

Fabio Ghironi

Boston College, Federal Reserve Bank of Boston, and National Bureau of Economic Research

Marc J. Melitz

Harvard University, Centre for Economic Policy Research, and National Bureau of Economic Research

This paper builds a framework for the analysis of macroeconomic fluctuations that incorporates the endogenous determination of the number of producers and products over the business cycle. Economic

Previously circulated under the title “Business Cycles and Firm Dynamics” and first presented in the summer of 2004. For helpful comments, we thank Robert Shimer, two anonymous referees, Susanto Basu, Christian Broda, Sanjay Chugh, Andrea Colciago, Diego Comin, Fiorella De Fiore, Massimo Giovannini, Jean-Olivier Hairault, Robert Hall, Kólver Hernández, Boyan Jovanovic, Nobuhiro Kiyotaki, Oleksiy Kryvtsov, Philippe Martin, Kris Mitchener, José-Víctor Ríos-Rull, Nicholas Sim, Viktors Stebunovs, Michael Woodford, and participants in many conferences and seminars. We are grateful to Brent Bundick, Massimo Giovannini, Nicholas Sim, Viktors Stebunovs, and Pinar Uysal for excellent research assistance. Remaining errors are our responsibility. Bilbiie thanks the NBER, the Centre for Economic Performance at London School of Economics, and the European Central Bank for hospitality in the fall of 2006, 2005, and summer of 2004, respectively; Nuffield College at Oxford for financial support during the 2004–7 period; and Banque de France for financial support through the Chaire Banque de France at the Paris School of Economics. Ghironi and Melitz thank the National Science Foundation for financial support through a grant to the NBER. The views expressed in this paper are those of the authors and do not necessarily reflect those of Banque de France, the Federal Reserve Bank of Boston, or Federal Reserve policy.

[*Journal of Political Economy*, 2012, vol. 120, no. 2]

© 2012 by The University of Chicago. All rights reserved. 0022-3808/2012/12002-0004\$10.00

expansions induce higher entry rates by prospective entrants subject to sunk investment costs. The sluggish response of the number of producers generates a new and potentially important endogenous propagation mechanism for business cycle models. The return to investment determines household saving decisions, producer entry, and the allocation of labor across sectors. Our framework replicates several features of business cycles and predicts procyclical profits even for preference specifications that imply countercyclical markups.

I. Introduction

This paper studies the role of endogenous producer entry and creation of new products in propagating business cycle fluctuations. Toward that goal, we develop a dynamic, stochastic, general equilibrium model with monopolistic competition, consumer love for variety, and sunk entry costs. We seek to understand the contributions of the intensive and extensive margins—changes in production of existing goods and in the range of available goods—to the response of the economy to changes in aggregate productivity.

Empirically, new products are introduced not only by new firms but also by existing firms (most often at their existing production facilities). We therefore take a broad view of producer entry (and exit) as also incorporating product creation (and destruction) by existing firms, although our model does not address the determinants of product variety within firms. Even though new firms account for a small share of overall production (for US manufacturing, new firms account for 2–3 percent of both overall production and employment), the contribution of new products (including those produced at existing firms) is substantially larger—important enough to be a major source of aggregate output changes. Furthermore, as is the case with firm entry, new product creation is also very strongly procyclical.¹

The important contribution of product creation and destruction to aggregate output is convincingly documented in a recent paper by Bernard, Redding, and Schott (2010), who are the first to measure product creation and destruction within firms across a large portion of the US economy (all US manufacturing firms). For each firm, they record production levels (dollar values) across five-digit US Standard Industrial Classification (SIC) categories, which still represent a very coarse defi-

¹ The working paper version of this study (Bilbiie, Ghironi, and Melitz 2007) contains evidence on the procyclicality of net firm entry (measured as new incorporations minus failures) and profits for the period 1947–98. Our conclusions there are in line with the pioneering work of Dunne, Roberts, and Samuelson (1988). Here, we focus on product creation rather than firm entry.

nition of products.² Bernard et al. first document that 94 percent of product additions by US manufacturing firms occur within their pre-existing production facilities (as opposed to new plants or via mergers and acquisitions). They further show that 68 percent of firms change their product mix within a 5-year census period (representing 93 percent of firms weighted by output). Of these firms, 66 percent both add and drop products (representing 87 percent of firms weighted by output). Thus, product creation over time is not just a secular trend at the firm level (whereby firms steadily increase the range of products they produce over time). Most important, Bernard et al. show that product creation and destruction account for important shares of overall production: over a 5-year period—a horizon usually associated with the length of business cycles—the value of new products (produced at existing firms) is 33.6 percent of overall output during that period (−30.4 percent of output for the lost value from product destruction at existing firms). These numbers are almost twice (1.8 times) as large as those accounted for by changes at the intensive margin: production increases and decreases for the same product at existing firms. The overall contribution of the extensive margin (product creation and destruction) would be even higher if a finer level of product disaggregation (beyond the five-digit level) were available.³

Put together, product creation (by both existing firms and new firms) accounts for 46.6 percent of output in a 5-year period, and the lost value from product destruction (by existing and exiting firms) accounts for 44 percent of output. This represents a minimal annual contribution of 9.3 percent (for product creation) and 8.8 percent (for product destruction). The actual annual contributions are likely larger, not only because the coarse definition of a product potentially misses much product creation and destruction within the five-digit SIC category but also because additions to and subtractions from output across years within the same 5-year interval (for a given firm-product combination) are not recorded. Relatedly, Den Haan and Sedlacek (2010) estimate the contribution of the extensive margin (measured along the employment dimension) to total value added. They calculate the contribution of “cyclical workers” (workers who during the period under scrutiny ex-

² As an example, the five-digit SIC codes within the four-digit SIC category 3949 (sporting and athletics goods) are 39491 (fishing tackle and equipment), 39492 (golf equipment), 39493 (playground equipment), 39494 (gymnasium and exercise equipment), and 39495 (other sporting and athletic goods). For all of US manufacturing, there are 1,848 five-digit products.

³ In the example of five-digit SIC 39494 (gymnasium and exercise equipment) from the previous footnote, any production of a new equipment product, whether a treadmill, an elliptical machine, a stationary bike, or any weight machine, would be recorded as production of the same product and hence be counted toward the intensive margin of production.

perienced a nonemployment spell) over a 3-year interval for Germany and the United States and find that this amounts to roughly half of total value added.

The substantial contribution of product creation and destruction is also confirmed by Broda and Weinstein (2010), who measure products at the finest possible level of disaggregation: the product bar code. Their data cover all the purchases of products with bar codes by a representative sample of US consumers. An important feature of the evidence in Bernard et al. (2010) is confirmed by Broda and Weinstein's highly disaggregated data: 92 percent of product creation occurs within existing firms. Broda and Weinstein find that 9 percent of the consumers' purchases in a year are devoted to new goods not previously available.⁴ Similarly to Bernard et al., Broda and Weinstein find that the market share of new products is four times larger than the market share of new firms (measured in terms of either output or employment), precisely because most product creation occurs within the firm (the same conclusion arises for product destruction vs. firm exit). Furthermore, Broda and Weinstein report that this product creation is strongly procyclical at a quarterly business cycle frequency. The evidence on the strong procyclicality of product creation is also confirmed by Axarloglou (2003) for US manufacturing at a monthly frequency.

In our model, we assume symmetric, homothetic preferences over a continuum of goods. This nests several tractable specifications (including constant elasticity of substitution [CES]) as special cases. To keep the setup simple, we do not model multiproduct firms. In our model presentation below and in the discussion of results, there is a one-to-one identification between a producer, a product, and a firm. This is consistent with much of the macroeconomic literature with monopolistic competition, which similarly uses "firm" to refer to the producer of an individual good. However, the relevant profit-maximizing unit in our setup is best interpreted as a production line, which could be nested within a multiproduct firm. The boundary of the firm across products is then not determined. Strategic interactions (within and across firms) do not arise because of our assumption of a continuum of goods as long as each multiproduct firm produces a countable set of goods of measure 0.⁵ In this interpretation of our model, producer entry and

⁴ This 9 percent figure is low relative to its 9.3 percent counterpart from Bernard et al. (2010), given the substantial difference in product disaggregation across the two studies (the extent of product creation increases monotonically with the level of product disaggregation). We surmise that this is due to the product sampling of Broda and Weinstein's (2010) data: including only final goods with bar codes. Food items, which have the lowest levels of product creation rates, tend to be overrepresented in those samples.

⁵ This differentiates our approach from that of Jaimovich and Floetotto (2008), who assume a discrete set of producers within each sector. In that case, the boundaries of firms crucially determine the strategic interaction between individual competitors.

exit capture the product-switching dynamics within firms documented by Bernard et al. (2010).

In our baseline setup, each individual producer/firm produces output using only labor. However, the number of firms that produce in each period can be interpreted as the capital stock of the economy, and the decision of households to finance entry of new firms is akin to the decision to accumulate physical capital in the standard real business cycle (RBC) model. Product creation (or, more broadly, entry) takes place subject to sunk product development costs, which are paid by investors in the expectation of future profits. Free entry equates the value of a product (the present discounted value of profits) to the sunk cost; subsequent to entry, the per-period profits fluctuate endogenously. This distinguishes our framework from earlier studies that modeled entry in a frictionless way: there, entry drives profits to zero in every period. (We discuss the relation between our work and these studies later on.) Our framework is hence closer to that of variety-based endogenous growth models (see, e.g., Romer 1990; Aghion and Howitt 1991; Grossman and Helpman 1991). Indeed, just as the RBC model is a discrete-time, stochastic, general equilibrium version of the exogenous growth model that abstracts from growth to focus on business cycles, our model can be viewed as a discrete-time, stochastic, general equilibrium version of variety-based, endogenous growth models that abstracts from endogenous growth (we discuss in greater detail further on why we have chosen to abstract from growth).

From a conceptual standpoint, linking innovation-based growth and business cycle theory is not new: the history of this idea goes back at least to Schumpeter (1934). Aghion and Howitt (1991) review some attempts at unifying growth and business cycles. Shleifer's (1986) theory of implementation cycles is one example of the conceptual link between (endogenous) business cycles and innovation-based growth theory: cycles occur because firms, expecting higher profits in booms due to a demand externality, innovate simultaneously in the expectation of a boom; the boom therefore becomes self-fulfilling. However, to the best of our knowledge this is the first study that blends elements of variety-based endogenous growth theory and RBC methodology (including the focus on exogenous aggregate productivity as the only source of uncertainty). Moreover, our framework also uses a general structure of preferences for variety that implies that markups fall when market size increases (which can be viewed as a dynamic extension of Krugman's [1979] insights about the effects of market size on firm size and markups).

The investment in new productive units is financed by households through the accumulation of shares in the portfolio of firms. The stock market price of this investment fluctuates endogenously in response to

shocks and is at the core of our propagation mechanism. Together with the shares' payoff (monopolistic profits), it determines the return to investment/entry, which in turn determines household saving decisions, producer entry, and the allocation of labor across sectors in the economy. This contrasts with the standard, one-sector RBC models, where the price of physical capital is constant without capital adjustment costs and the return to investment is simply equal to the marginal product of physical capital. This approach to investment and the price of capital provides an alternative to adjustment costs in order to obtain a time-varying price of capital. It also introduces a direct link between investment and (the expectation of) economic profits. In our model, labor is allocated to production of existing goods and creation of new ones, and the total number of products acts as capital in the production of the consumption basket. This structure is close to two-sector versions of the RBC model in which labor is allocated to production of the consumption good and to investment that augments the capital stock, and capital is also used to produce the consumption good. We discuss this relationship in further detail below.

In terms of matching key second moments of the US business cycle, our baseline model performs at least as well as a traditional RBC model (it does better at matching the volatility of output and hours). Importantly, our model can additionally account for stylized facts pertaining to entry, profits, and markups. With translog preferences (for which the elasticity of substitution is increasing in the number of goods produced), our model is further able to simultaneously generate countercyclical markups and procyclical profits; it also reproduces the time profile of the markup's correlation with the business cycle. These are well-known challenges for models of countercyclical markups based on sticky prices (see Rotemberg and Woodford [1999] for a discussion). To the best of our knowledge, our framework is the first to address and explain these issues simultaneously.⁶ Moreover, we develop an extension of our framework that also incorporates investment in physical capital. This significantly improves the performance of the model (relative to both our baseline without physical capital and the standard RBC model) in reproducing the volatilities of output, hours worked, and total investment.

The paper is structured as follows. Section II presents the baseline model. Section III computes impulse responses and second moments for a numerical example and illustrates the properties of the model for

⁶ Perfect-competition models, such as the standard RBC, address none of these facts. Imperfect-competition versions (with or without sticky prices) generate fluctuations in profits (and, for sticky prices, in markups) but no entry. Frictionless-entry models discussed later generate fluctuations in entry (and, in some versions—such as Cook [2001], Comin and Gertler [2006], or Jaimovich and Floetotto [2008]—also markups) but with zero profits.

transmission of economic fluctuations. Section IV outlines the extension of our model to include investment in physical capital and illustrates its second-moment properties. Section V discusses the relation between our work and other contributions to the literature on entry and business cycles. Section VI presents conclusions.

II. The Model

A. Household Preferences and the Intra-temporal Consumption Choice

The economy is populated by a unit mass of atomistic, identical households. All contracts and prices are written in nominal terms. Prices are flexible. Thus, we solve only for the real variables in the model. However, as the composition of the consumption basket changes over time because of firm entry (affecting the definition of the consumption-based price index), we introduce money as a convenient unit of account for contracts. Money plays no other role in the economy. For this reason, we do not model the demand for cash currency and resort to a cashless economy as in Woodford (2003).

The representative household supplies L_t hours of work each period t in a competitive labor market for the nominal wage rate W_t and maximizes expected intertemporal utility $E_t[\sum_{s=t}^{\infty} \beta^{s-t} U(C_s, L_s)]$, where C is consumption and $\beta \in (0, 1)$ the subjective discount factor. The period utility function takes the form $U(C_t, L_t) = \ln C_t - \chi(L_t)^{1+1/\varphi}/(1+1/\varphi)$, $\chi > 0$, where $\varphi \geq 0$ is the Frisch elasticity of labor supply to wages and the intertemporal elasticity of substitution in labor supply. Our choice of functional form for the utility function is guided by results in King, Plosser, and Rebelo (1988): Given separable preferences, log utility from consumption ensures that income and substitution effects of real wage variation on effort cancel out in the steady state; this is necessary to have constant steady-state effort and balanced growth if there is productivity growth.

At time t , the household consumes the basket of goods C_t , defined over a continuum of goods Ω . At any given time t , only a subset of goods $\Omega_t \subset \Omega$ is available. Let $p_t(\omega)$ denote the nominal price of a good $\omega \in \Omega_t$. Our model can be solved for any parameterization of symmetric homothetic preferences. For any such preferences, there exists a well-defined consumption index C_t and an associated welfare-based price index P_t . The demand for an individual variety, $c_t(\omega)$, is then obtained as $c_t(\omega)d\omega = C_t \partial P_t / \partial p_t(\omega)$, where we use the conventional notation for quantities with a continuum of goods as flow values (see the Appendix for more details).

We anticipate symmetric equilibrium across products. Given the demand level per variety, the symmetric price elasticity of demand ζ is in

general a function of the number N_i of goods (where N_i is the mass of Ω_i): $\zeta(N_i) \equiv [\partial c_i(\omega)/\partial p_i(\omega)][p_i(\omega)/c_i(\omega)]$, for any symmetric variety ω . The benefit of additional product variety is described by the relative price $\rho_i(\omega) = \rho(N_i) \equiv p_i(\omega)/P_i$, for any symmetric variety ω , or, in elasticity form, $\epsilon(N_i) \equiv \rho'(N_i)N_i/\rho(N_i)$. Together, $\zeta(N_i)$ and $\rho(N_i)$ completely characterize the effects of consumption preferences in our model; explicit expressions for these objects can be obtained upon specifying functional forms for preferences, as will become clear in the discussion below.

B. Firms

There is a continuum of monopolistically competitive firms, each producing a different variety $\omega \in \Omega$. Production requires only one factor, labor (this assumption is relaxed in Sec. IV, where we introduce physical capital). Aggregate labor productivity is indexed by Z_t , which represents the effectiveness of one unit of labor. The term Z_t is exogenous and follows an AR(1) process (in logarithms). Output supplied by firm ω is $y_i(\omega) = Z_t l_i(\omega)$, where $l_i(\omega)$ is the firm's labor demand for productive purposes. The unit cost of production, in units of the consumption basket C_t , is w_t/Z_t , where $w_t \equiv W_t/P_t$ is the real wage.

Prior to entry, firms face an exogenous sunk entry cost of f_E effective labor units (as in Judd [1985], Romer [1990], and Grossman and Helpman [1991], among others), equal to $w_t f_E/Z_t$ units of the consumption basket. This specification ensures that exogenous productivity shocks are truly aggregate in our model, as they affect symmetrically both production of existing goods and creation of new products.⁷ Given our modeling assumption relating each firm to a product line, we think of the entry cost as the development and setup cost associated with a particular variety.

There are no fixed production costs. Hence, all firms that enter the economy produce in every period until they are hit with a "death" shock, which occurs with probability $\delta \in (0, 1)$ in every period. The assumption of exogenous exit is adopted here only in the interest of tractability. Recent evidence suggests that this assumption is a reasonable starting point for analysis. At the product level, Broda and Weinstein (2010) report that product destruction is much less cyclical than product creation. A similar pattern also holds at the plant level: using US Census (annual) data, Lee and Mukoyama (2007) find that, while plant entry is highly procyclical (the entry rate is 8.1 percent in booms and 3.4 percent in recessions), annual exit rates are similar across booms and recessions (5.8 and 5.1 percent, respectively). They also find that plants

⁷ Thus, the "production function" for new goods is $N_{E,t} = Z_t I_{E,t} / f_E$.

TABLE 1
TWO FRAMEWORKS

CES	Translog
$\mu(N_i) = \mu = \frac{\theta}{\theta-1}$	$\mu(N_i) = \mu_i = 1 + \frac{1}{\frac{\sigma N_i}{N}}$
$\rho(N_i) = N_i^{\mu-1} = N_i^{1/(\theta-1)}$	$\rho(N_i) = \exp\left(-\frac{1}{2} \frac{N-N_i}{\sigma N_i}\right)$, $\bar{N} \equiv \text{Mass}(\Omega)$
$\epsilon(N_i) = \mu - 1$	$\epsilon(N_i) = \frac{1}{2\sigma N_i} = \frac{1}{2}[\mu(N_i) - 1]$

exiting in recessions are very similar to those exiting in booms (in terms of employment or productivity).

In units of consumption, variety ω 's price will be set to $\rho_i(\omega) \equiv p_i(\omega)/P_i = \mu_i w_i/Z_i$, where μ_i is the price markup over marginal cost (anticipating symmetric equilibrium). Given our demand specification with endogenous price elasticity of residual demand, this markup is a function of the number of producers: $\mu_i = \mu(N_i) \equiv \zeta(N_i)/[\zeta(N_i) + 1]$. The profits generated from the sales of each variety (expressed in units of consumption) are $d_i(\omega) = d_i = [1 - \mu(N_i)^{-1}]C_i/N_i$ and are returned to households as dividends.

1. Preference Specifications and Markups

In our quantitative exercises, we consider two alternative specifications that are nested within our general analysis of symmetric homothetic preferences. The first specification features CES between goods as in Dixit and Stiglitz (1977). For these CES preferences, the consumption aggregator is $C_i = [\int_{\omega \in \Omega} c_i(\omega)^{\theta-1} d\omega]^{\theta/(\theta-1)}$, where $\theta > 1$ is the symmetric elasticity of substitution across goods. The consumption-based price index is then $P_i = [\int_{\omega \in \Omega} p_i(\omega)^{1-\theta} d\omega]^{1/(1-\theta)}$ and the household's demand for each individual good ω is $c_i(\omega) = [p_i(\omega)/P_i]^{-\theta} C_i$. It follows that the markup and the benefit of variety are independent of the number of goods ($\epsilon(N_i) = \epsilon$, $\mu(N_i) = \mu$) and related by $\epsilon = \mu - 1 = 1/(\theta - 1)$. The second specification uses the translog expenditure function proposed by Feenstra (2003), which introduces demand-side pricing complementarities. For this preference specification, the symmetric price elasticity of demand is $-(1 + \sigma N_i)$, $\sigma > 0$: as N_i increases, goods become closer substitutes, and the elasticity of substitution $1 + \sigma N_i$ increases. If goods are closer substitutes, then the markup $\mu(N_i)$ and the benefit of additional varieties in elasticity form ($\epsilon(N_i)$) must decrease. This property occurs whenever the price elasticity of residual demand decreases with quantity consumed along the residual demand curve. The change in $\epsilon(N_i)$ is only half the change in net markup generated by an increase in the number of producers. Table 1 contains the expressions for markup, relative price,

and the benefit of variety (the elasticity of ρ to the number of firms) for each preference specification.

2. Firm Entry and Exit

In every period, there is a mass N_t of firms producing in the economy and an unbounded mass of prospective entrants. These entrants are forward looking and correctly anticipate their expected future profits $d_t(\omega)$ in every period $s \geq t + 1$ as well as the probability δ (in every period) of incurring the exogenous exit-inducing shock. Entrants at time t start producing only at time $t + 1$, which introduces a one-period time-to-build lag in the model. The exogenous exit shock occurs at the very end of the time period (after production and entry). A proportion δ of new entrants will therefore never produce. Prospective entrants in period t compute their expected postentry value ($v_t(\omega)$) given by the present discounted value of their expected stream of profits $\{d_s(\omega)\}_{s=t+1}^\infty$:

$$v_t(\omega) = E_t \sum_{s=t+1}^\infty Q_{t,s} d_s(\omega), \tag{1}$$

where $Q_{t,s}$ is the stochastic discount factor that is determined in equilibrium by the optimal investment behavior of households. This also represents the value of incumbent firms after production has occurred (since both new entrants and incumbents then face the same probability $1 - \delta$ of survival and production in the subsequent period). Entry occurs until firm value is equalized with the entry cost, leading to the free-entry condition $v_t(\omega) = w_t f_E / Z_t$. This condition holds as long as the mass $N_{E,t}$ of entrants is positive. We assume that macroeconomic shocks are small enough for this condition to hold in every period. Finally, the timing of entry and production we have assumed implies that the number of producing firms during period t is given by $N_t = (1 - \delta)(N_{t-1} + N_{E,t-1})$. The number of producing firms represents the stock of capital of the economy. It is an endogenous state variable that behaves much like physical capital in the benchmark RBC model but in contrast to the latter has an endogenously fluctuating price given by (1).

3. Symmetric Firm Equilibrium

All firms face the same marginal cost. Hence, equilibrium prices, quantities, and firm values are identical across firms: $p_t(\omega) = p_t$, $\rho_t(\omega) = \rho_t$, $l_t(\omega) = l_t$, $y_t(\omega) = y_t$, $d_t(\omega) = d_t$, and $v_t(\omega) = v_t$. In turn, equality of prices across firms implies that the consumption-based price index P_t and the firm-level price p_t are such that $p_t/P_t \equiv \rho_t = \rho(N_t)$. An increase in the

number of firms implies necessarily that the relative price of each individual good increases, $\rho'(N_t) > 0$. When there are more firms, households derive more welfare from spending a given nominal amount; that is, *ceteris paribus*, the price index decreases. It follows that the relative price of each individual good must rise. The aggregate consumption output of the economy is $N_t \rho_t y_t = C_t$, which we can rewrite as $C_t = Z_t \rho(N_t)(L_t - \int_{E,t} N_{E,t}/Z_t)$. An increase in the number of entrants $N_{E,t}$ absorbs productive resources and acts like an overhead labor cost in production of consumption. Importantly, in the symmetric firm equilibrium, the option value of waiting to enter is zero, despite the presence of sunk costs and exit risk. This happens because all uncertainty in our model is aggregate, and the death shock is symmetric across firms and time invariant.⁸

C. Household Budget Constraint and Optimal Behavior

Households hold shares in a mutual fund of firms. Let x_t be the share in the mutual fund held by the representative household entering period t . The mutual fund pays a total profit in each period (in units of currency) equal to the total profit of all firms that produce in that period, $P_t N_t d_t$. During period t , the representative household buys x_{t+1} shares in a mutual fund of $N_t + N_{E,t}$ firms (those already operating at time t and the new entrants). The mutual fund covers all firms in the economy, even though only $1 - \delta$ of these firms will produce and pay dividends at time $t + 1$. The date t price (in units of currency) of a claim to the future profit stream of the mutual fund of $N_t + N_{E,t}$ firms is equal to the nominal price of claims to future firm profits, $P_t v_t$.

The household enters period t with mutual fund share holdings x_t . It receives dividend income on mutual fund share holdings, the value of selling its initial share position, and labor income. The household allocates these resources between purchases of shares to be carried into next period and consumption. The period budget constraint (in units of consumption) is

$$v_t(N_t + N_{E,t})x_{t+1} + C_t = (d_t + v_t)N_t x_t + w_t L_t. \quad (2)$$

The household maximizes its expected intertemporal utility subject to (2).

The Euler equations for share holdings is

$$v_t = \beta(1 - \delta)E_t \left[\frac{C_t}{C_{t+1}} (v_{t+1} + d_{t+1}) \right].$$

⁸ See the Appendix for the proof. This contrasts with, among others, Caballero and Hammour (1994) and Campbell (1998). See also Jovanovic (2006) for a more recent contribution in that vein.

As expected, forward iteration of the equation for share holdings and the absence of speculative bubbles yield the asset price solution in equation (1), with the stochastic discount factor $Q_{t,s} = [\beta(1 - \delta)]^s C_t / C_{t+s}$.

Finally, the allocation of labor effort obeys the standard intratemporal first-order condition:

$$\chi(L_t)^{1/\varphi} = \frac{w_t}{C_t}. \tag{3}$$

D. Aggregate Accounting, Labor Market Dynamics, and the Relation with RBC Theory

Different from the benchmark, one-sector, RBC model of Kydland and Prescott (1982) and many other studies, our model economy is a two-sector economy in which one sector employs part of the labor supply to produce consumption and the other sector employs the rest of the labor supply to produce new firms. Labor market equilibrium requires that these two components of labor demand sum to aggregate labor supply: $L_t^C + L_t^E = L_t$, where $L_t^C = N_t l_t$ is the total amount of labor used in production of consumption, and $L_t^E = N_{E,t} f_E / Z_t$ is labor used to create new firms.

Aggregating the budget constraint (2) across households and imposing the equilibrium condition $x_{t+1} = x_t = 1$ for all t yields the aggregate accounting identity for GDP $Y_t \equiv C_t + N_{E,t} v_t = w_t L_t + N_t d_t$. Total consumption, C_t , plus investment (in new products or firms), $N_{E,t} v_t$, must be equal to total income (labor income $w_t L_t$ plus dividend income $N_t d_t$). Thus, v_t is the relative price of the investment “good” in terms of consumption. In a one-sector RBC model, only the interest rate dictates the allocation of resources between consumption and investment. In our model, this allocation is reflected in the allocation of labor across the two sectors (producing consumption goods and new goods). The key distinction is that the relative price of investment v_t fluctuates and dictates the allocation of labor across sectors, in conjunction with the return on shares, $r_{t+1}^E \equiv (v_{t+1} + d_{t+1})/v_t$. This is reminiscent of a two-sector RBC model in which the relative price of investment is also endogenous and affects the allocation of resources to consumption versus investment (see, e.g., Long and Plosser 1983; Boldrin, Christiano, and Fisher 2001).

Despite this similarity, there are important features that differentiate our framework from a two-sector RBC structure: First, we model explicitly the microeconomic incentives for product creation from consumer love for variety and profit incentives for innovators. Second, we have a different notion of investment, directed entirely toward the extensive margin (the creation of new goods), whereas all investment takes

TABLE 2
MODEL SUMMARY

Variable	Definition
Pricing	$\rho_t = \mu_t \frac{w_t}{Z}$
Markup	$\mu_t = \mu(N_t)$
Variety effect	$\rho_t = \rho(N_t)$
Profits	$d_t = \left(1 - \frac{1}{\mu_t}\right) \frac{C_t}{N_t}$
Free entry	$v_t = w_t \frac{f_t}{Z}$
Number of firms	$N_t = (1 - \delta)(N_{t-1} + N_{E,t-1})$
Intratemporal optimality	$\chi(L_t)^{1/\varphi} = \frac{w_t}{C_t}$
Euler equation (shares)	$v_t = \beta(1 - \delta)E_t \left[\frac{C_t}{C_{t+1}} (v_{t+1} + d_{t+1}) \right]$
Aggregate accounting	$C_t + N_{E,t} v_t = w_t L_t + N_t d_t$

place at the intensive margin (machines used to produce more of the same good) in the RBC model (one-sector or two-sector). Both forms of investment take place in reality, and the version of our model introduced in Section IV addresses this. Third, our model can address facts about entry, profits, and markups. A two-sector RBC model that is otherwise isomorphic to ours would need the ad hoc assumption of a labor share in consumption output that is an appropriate function of capital to generate a procyclical labor share in GDP (as our model does under translog preferences). Fourth, since aggregate production of consumption in our model features a form of increasing returns due to variety, one needs to introduce increasing returns in the consumption sector of the RBC model to make it isomorphic to ours. But since internal increasing returns are inconsistent with perfect competition, one needs to adopt the ad hoc assumption of a labor externality in the consumption sector to avoid internal increasing returns at the firm level (or otherwise assume that firms price at average cost).⁹ For these reasons, and its traditional role as benchmark, we keep the one-sector RBC model as a reference point for performance comparison below.

E. Model Summary

Table 2 summarizes the main equilibrium conditions of the model (the labor market equilibrium condition is redundant once the variety effect equation is included). The equations in the table constitute a system of nine equations in nine endogenous variables: ρ_t , μ_t , d_t , w_t , L_t , $N_{E,t}$, N_t ,

⁹ Evidence in Harrison (2003) does not support the assumptions needed to make the models isomorphic. In particular, Harrison finds that returns to scale are slightly increasing in the investment sector, but they are decreasing or constant in the consumption sector.

v_t , and C_t . Of these endogenous variables, one is predetermined as of time t : the total number of firms, N_t . Additionally, the model features one exogenous variable: aggregate productivity, Z_t .

F. Steady State

We assume that productivity is constant in the steady state and denote steady-state levels of variables by dropping the time subscript: $Z_t = Z$. We conjecture that all endogenous variables are constant in the steady state and show that this is indeed the case. We define the steady-state interest rate as a function of the rate of time preference, $1 + r \equiv \beta^{-1}$. We exploit this below to treat r as a parameter in the solution. The full steady-state solution is presented in the Appendix. Here, we present the most important long-run properties of our model.

The gross return on shares is $1 + d/v = (1 + r)/(1 - \delta)$, which captures a premium for expected firm destruction. The number of new entrants makes up for the exogenous destruction of existing firms: $N_E = \delta N/(1 - \delta)$. Calculating the shares of profit income and investment in consumption output and GDP allows us to draw another transparent comparison between our model and the standard RBC setup. The steady-state profit equation gives the share of profit income in consumption output: $dN/C = (\mu - 1)/\mu$. Using this result in conjunction with those obtained above, we have the share of investment in consumption output, denoted by γ : $vN_E/C = \gamma \equiv (\mu - 1)\delta/[\mu(r + \delta)]$. This expression is similar to its RBC counterpart. There, the share of investment in output is given by $s_K\delta/(r + \delta)$, where δ is the depreciation rate of capital and s_K is the share of capital income in total income. In our framework, $(\mu - 1)/\mu$ can be regarded as governing the share of “capital” since it dictates the degree of monopoly power and hence the share of profits that firms generate from producing consumption output (dN/C). Note that $Y = C + vN_E$. Then the shares of investment and profit income in GDP are $vN_E/Y = \gamma/(1 + \gamma)$ and $dN/Y = [(r + \delta)\gamma]/[\delta(1 + \gamma)]$, respectively. It follows that the share of consumption in GDP is $C/Y = 1/(1 + \gamma)$. The share of labor income in total income is $wL/Y = 1 - [(r + \delta)\gamma]/[\delta(1 + \gamma)]$. Importantly, all these ratios are constant. If we allowed for long-run growth (either via an exogenous trend in Z_t or endogenously by assuming entry cost f_E/N_t as in Grossman and Helpman [1991]), these long-run ratios would still be constant with CES preferences, consistent with the Kaldorian growth facts. In fact, regardless of preference specification within the homothetic class, our model’s long-run properties with growth would be consistent with two stylized facts originally found by Kaldor (1957): a constant share of profits in total capital, $dN/vN = (r + d)/(1 - d)$, and, relatedly, a high correlation between the profit share in GDP and the investment share in GDP. These

facts are absent from both the standard RBC model and the frictionless entry models discussed in Section V.

We abstract from growth for two reasons (beyond the fact that it is the subject of its own extensive literature). In variety-based models, endogenous growth occurs whenever costs of product creation decrease with the number of existing products; in other words, the production function for new goods exhibits constant returns to scale in an accumulating factor, namely, the number of goods. The growth rate in such models (such as in the standard AK model) is a function of the level of productivity: any shock to productivity would immediately put the economy on the new balanced growth path with no transition dynamics. We focus instead on short-run fluctuations in which the extensive margin does play a significant role in propagating shocks. Second, the growth rate is also a function of the elasticity of substitution between goods, which is not constant (in general) in our model. Reconciling an endogenous time-varying markup with stylized growth facts (that imply constant markups and profit shares in the long run) is a challenge to growth theory that is worth future investigation but is beyond the scope of this paper.¹⁰

G. Dynamics

We solve for the dynamics in response to exogenous shocks by log-linearizing the model around the steady state. However, the model summary in table 2 already allows us to draw some conclusions on the properties of shock responses for some key endogenous variables. It is immediate to verify that firm value is such that $v_t = w_t f_E / Z_t = f_E \rho(N_t) / \mu(N_t)$. Since the number of producing firms is predetermined and does not react to exogenous shocks on impact, firm value is predetermined with respect to productivity shocks. An increase in productivity results in a proportional increase in the real wage on impact through its effect on labor demand. Since the entry cost is paid in effective labor units, this does not affect firm value. An implication of the wage schedule $w_t = Z_t \rho(N_t) / \mu(N_t)$ is that also marginal cost, w_t / Z_t , is predetermined with respect to the shock.

We can reduce the system in table 2 to a system of two equations in two variables, N_t and C_t (see the Appendix). With sans serif fonts used to denote percentage deviations from steady-state levels, log linearization around the steady state under assumptions of log normality and homoskedasticity yields

¹⁰ Balanced growth would be restored under translog preferences by making the ad hoc assumption that the parameter σ decreases at the same rate at which N_t increases in the long run.

$$\begin{aligned} N_{t+1} = & \left[1 - \delta + \frac{r + \delta}{\mu - 1} \epsilon + \left(\frac{r + \delta}{\mu - 1} + \delta \right) \varphi (\epsilon - \eta) \right] N_t \\ & - \left[\varphi \left(\frac{r + \delta}{\mu - 1} + \delta \right) + \frac{r + \delta}{\mu - 1} \right] C_t + (1 + \varphi) \left(\frac{r + \delta}{\mu - 1} + \delta \right) Z_t, \end{aligned} \tag{4}$$

$$\begin{aligned} C_t = & \frac{1 - \delta}{1 + r} E_t C_{t+1} - \left[\frac{1 - \delta}{1 + r} (\epsilon - \eta) - \frac{r + \delta}{1 + r} \left(1 - \frac{\eta}{\mu - 1} \right) \right] N_{t+1} \\ & + (\epsilon - \eta) N_t, \end{aligned} \tag{5}$$

where $\eta \equiv \mu'(N)N/\mu(N) \leq 0$ is the elasticity of the markup function with respect to N , which takes the value of zero under CES and $-(1 + \sigma N)^{-1}$ under translog preferences. Equation (4) states that the number of firms producing at $t + 1$ increases if consumption at time t is lower (households save more in the form of new firms) or if productivity is higher. Equation (5) states that consumption at time t is higher the higher expected future consumption and the larger the number of firms producing at time t . The effect of N_{t+1} depends on parameter values. For realistic parameter values, we have $\epsilon - \eta > (r + \delta)/(1 - \delta)$: an increase in the number of firms producing at $t + 1$ is associated with lower consumption at t . (Higher productivity at time t lowers contemporaneous consumption through this channel, as households save to finance faster entry in a more attractive economy. However, we shall see below that the general equilibrium effect of higher productivity will be that consumption rises.)

In the Appendix, we show that the system (4)–(5) has a unique, nonexplosive solution for any possible parameterization. To solve the system, we assume $Z_t = \phi_Z Z_{t-1} + \varepsilon_{Z,t}$, where $\varepsilon_{Z,t}$ is an independent and identically distributed, Normal innovation with zero mean and variance $\sigma_{\varepsilon_Z}^2$.

III. Business Cycles: Propagation and Second Moments

In this section we explore the properties of our model by means of a numerical example. We compute impulse responses to a productivity shock. The responses substantiate the results and intuitions in the previous section. Then, we compute second moments of our artificial economy and compare them to second moments in the data and those produced by a standard RBC model.¹¹

¹¹ Numerical results are obtained using the Matlab Toolkit described in Uhlig (1999).

A. Empirically Relevant Variables and Calibration

An issue of special importance when comparing our model to properties of the data concerns the treatment of variety effects. As argued in Ghironi and Melitz (2005), when discussing model properties in relation to empirical evidence, it is important to recognize that empirically relevant variables—as opposed to welfare-consistent concepts—net out the effect of changes in the range of available products. The reason is that construction of consumer price index (CPI) data by statistical agencies does not adjust for availability of new products as in the welfare-consistent price index. Furthermore, adjustment for variety, when it happens, certainly does not happen either at the frequency represented by periods in our model or using one of the specific functional forms for preferences that our model assumes. It follows that CPI data are closer to p_t than P_t . For this reason, when investigating the properties of the model in relation to the data, one should focus on real variables deflated by a data-consistent price index. For any variable X_t in units of the consumption basket (other than the return to investment), we define its data-consistent counterpart as $X_{R,t} \equiv P_t X_t / p_t = X_t / \rho_t = X_t / \rho(N_t)$. We define the data-consistent return to investment using data-consistent share prices and dividends as $r_{R,t+1}^E \equiv (v_{R,t+1} + d_{R,t+1}) / v_{R,t}$.

In our baseline calibration, we interpret periods as quarters and set $\beta = 0.99$ to match a 4 percent annualized average interest rate. We set the size of the exogenous firm exit shock $\delta = 0.025$. This implies a 10 percent annual production destruction rate (both as a share of products and as market share) and is consistent with the Bernard et al. (2010) finding of an 8.8 percent minimum production destruction rate (measured as a market share).¹² Under CES preferences, we use the value of θ from Bernard et al. (2003) and set $\theta = 3.8$, which was calibrated to fit US plant and macro trade data. In our model, this choice implies a share of investment in GDP ($v_R N_E / Y_R = v N_E / Y$) around 16 percent.¹³ We calibrate the parameter σ under translog preferences to ensure equality of steady-state markup and number of firms across preference specifications as described in the Appendix. This implies $\sigma = 0.35323$.

¹² This calibration also implies a 10 percent annual job destruction rate, which is consistent with the empirical evidence.

¹³ It may be argued that the value of θ results in a steady-state markup that is too high relative to the evidence. However, it is important to observe that, in models without any fixed cost, $\theta / (\theta - 1)$ is a measure of both markup over marginal cost and average cost. In our model with entry costs, free entry ensures that firms earn zero profits net of the entry cost. This means that firms price at average cost (inclusive of the entry cost). Thus, although $\theta = 3.8$ implies a fairly high markup over marginal cost, our parameterization delivers reasonable results with respect to pricing and average costs. The main qualitative features of the impulse responses below are not affected if we set $\theta = 6$, resulting in a 20 percent markup of price over marginal cost as in Rotemberg and Woodford (1992) and several other studies.

We set steady-state productivity to $Z = 1$. The entry cost f_E does not affect any impulse response under CES preferences and under translog preferences with our calibration procedure. Therefore, we set $f_E = 1$ without loss of generality (basically, changing f_E amounts to changing the unit of measure for output and number of firms). We set the weight of the disutility of labor in the period utility function, χ , so that the steady-state level of labor effort is equal to one—and steady-state levels of all variables are the same—regardless of φ . This requires $\chi = 0.924271$. This choice is a mere normalization with no effect on the quantitative results. We set the elasticity of labor supply φ to four for consistency with King and Rebelo's (1999) calibration of the benchmark RBC model, to which we will compare our results.¹⁴

We use the same productivity process as King and Rebelo (1999), with persistence $\phi_z = 0.979$ and standard deviation of innovations $\sigma_{\varepsilon_z} = 0.0072$ to facilitate comparison of results with the baseline RBC setup. In King and Rebelo's benchmark RBC model with Cobb-Douglas production, the exogenous productivity process coincides with the Solow residual by construction, and persistence and the standard deviation of innovations are obtained by fitting an AR(1) process to Solow residual data. In our model, the aggregate GDP production function is not Cobb-Douglas, and hence the Solow residual does not coincide with exogenous productivity. In fact, it is not clear how one should define the Solow residual in our model to account for capital accumulation through the stock of firms N_t .¹⁵ Moreover, the Solow residual (however defined) is just another endogenous variable in our model. We could try to match its moments to the estimates in King and Rebelo's study, but we would face the same difficulty as for other endogenous variables—that our model, like the RBC model, does not generate enough endogenous persistence. We therefore opt for the same parameter values for the exogenous productivity process as used by King and Rebelo. In so doing, we place the test of the model's ability to outperform the RBC model (based on the standard benchmark against a set of macroeconomic aggregates) on the transmission mechanism rather than on the implications of different parameter choices for the exogenous driving force. This makes the comparison between models much more transparent.

¹⁴ The period utility function is defined over leisure $(1 - L_t)$ in King and Rebelo (1999), where the endowment of time in each period is normalized to one. The elasticity of labor supply is then the risk aversion to variations in leisure (set to one in their benchmark calibration) multiplied by $(1 - L)/L$, where L is steady-state effort, calibrated to $1/5$. This yields $\varphi = 4$ in our specification.

¹⁵ This issue is still relevant for our model extension with physical capital in Sec. IV.

B. Impulse Responses

Figure 1 shows the responses of key endogenous variables to a 1 percent positive innovation to Z_t under CES preferences. The number of years after the shock is on the horizontal axis. The responses for all real variables are shown using both the welfare-relevant price index P_t (represented as circles) and the data-consistent CPI price index p_t (represented as crosses). Both measures are important. The data-consistent series provides the link back to the empirical evidence. On the other hand, the dynamics are driven by optimizing behavior with respect to their welfare-relevant counterparts.

Consider first the effects of the shock on impact. Note that the relative price $\rho_t \equiv p_t/P_t$ depends only on the number of products N_t and is thus predetermined at time t . The impact responses for both the data- and welfare-consistent measures are thus identical. The productivity improvement spurs profit expectations generated by the increased demand for all individual goods y_t . Without any entry, this would translate into a higher (ex ante) value for each variety. However, the free-entry mechanism induces an immediate response of entry that drives the (ex post) equilibrium value of a variety back down to the level of the entry cost; recall that this is equal to the marginal cost of producing an extra unit of an existing good. Since marginal cost (and hence the entry cost) moves in lockstep with the—constant on impact—individual relative price (ρ_t), it follows that on impact there is no reaction in marginal cost; therefore, entry occurs up to the point at which the (ex post) equilibrium firm value does not react to the shock on impact.

The remaining question is then what is the optimal relative allocation of the productivity increase between the two sectors: consumption C_t and investment (entry) $N_{E,t}$. To understand why consumption increases less than proportionally with productivity, it is important to consider the investment decision of households. The price of a share (value of a firm) together with its payoff (dividends obtained from monopolistic firms) determine the return on a share: the return to entry (product creation). On impact, the rate of return to investing (r_{t+1}^E , evaluated from the ex ante perspective of investment decisions) is high, both because the present share price is low relative to the future and because next period's share payoffs (firm profits) are expected to be high. Intertemporal substitution logic implies that the household should postpone consumption into the future. Since the only means to transfer resources intertemporally is the introduction of new varieties, investment (measured either as the number of entrants $N_{E,t}$ or in consumption units $I_t^E \equiv v_t N_{E,t}$) increases on impact; this is the mirror ("demand") image of the new firms' decision to enter discussed earlier. This allocation of resources, driven by intertemporal substitution, is also reflected

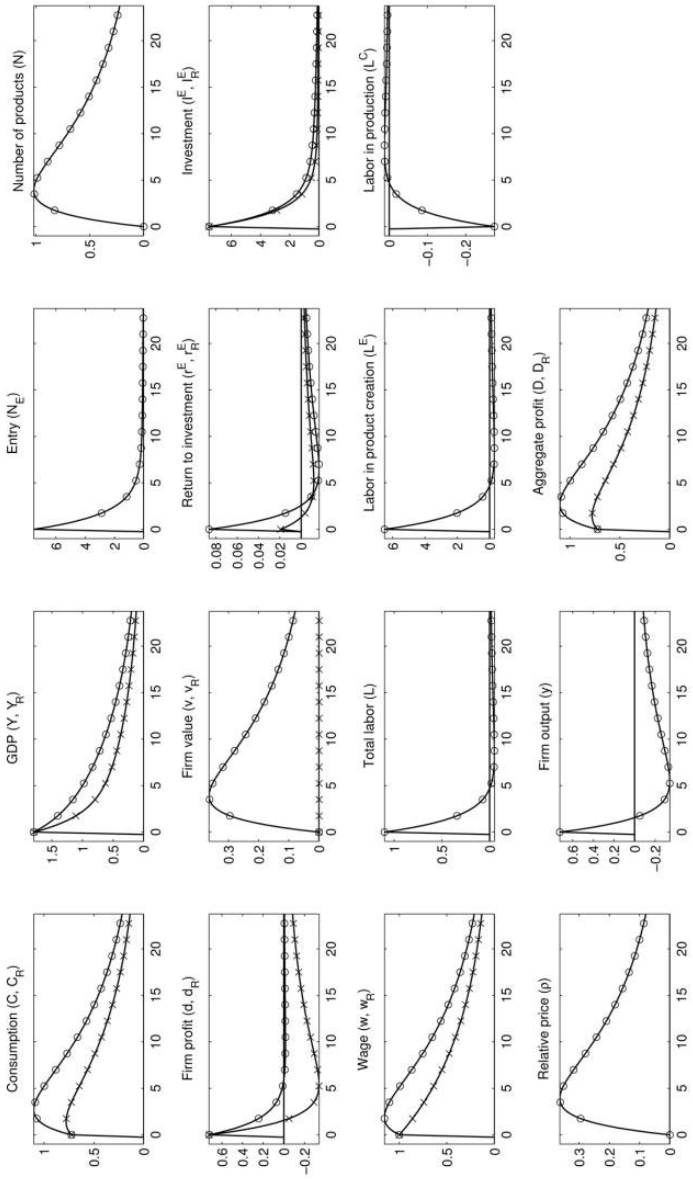


FIG. 1.—Impulse responses to a productivity increase, CES preferences (crosses denote data-consistent variables)

in the allocation of labor across the two sectors: on impact, labor is reallocated into product creation (L_t^E) from the production of existing goods (L_t^C).¹⁶ Finally, the real wage increases on impact in line with the increase in productivity, and faced with this higher wage, the household optimally decides to work more hours in order to attain a higher consumption level. GDP (Y_t) increases because both consumption and investment increase.

Over time, increased entry translates into a gradual increase in the number of products N_t and reduces individual good demand (output of each good falls below the steady state for most of the transition). More product variety also generates a love of variety welfare effect that is reflected in the increase in the relative price ρ_t . This increase is also reflected one-for-one in the welfare-consistent measure of the value of a variety (since the opportunity cost of investment in terms of forgone consumption is now higher with more varieties). Profits per variety fall with the reduction in demand per variety. Together with the higher opportunity cost of investment from higher product variety, this generates a fall in the return to investment/entry below its steady-state value and a reversal in the allocation of labor: labor is reallocated back from product creation to production. The hump-shaped pattern of aggregate consumption is consistent with the dynamics of the return to investment. After a certain amount of time, the number of products peaks and then progressively declines back to its old steady-state level. This also unwinds the welfare effects driven by the additional product variety (ρ_t decreases). The decrease in product variety is also reflected in a reversal of the decrease in individual good demand and profits per variety, which then increase back up to their steady-state levels. Importantly, however, aggregate profit $D_t \equiv N_t d_t$ and its data-consistent counterpart $D_{R,t}$ remain above the steady state throughout the transition. The response of data-consistent consumption ($C_{R,t}$) is still hump shaped but relatively more muted than its welfare-consistent counterpart as it does not factor in the additional benefits from product variety. The data-consistent firm value is constant because with CES preferences the markup is constant, namely, $v_{R,t} = f_E/\mu = f_E(\theta - 1)/\theta$. Finally, the data-consistent real wage $w_{R,t}$ declines monotonically toward the steady state, tracking the behavior of productivity.¹⁷

¹⁶ The negative correlation between labor inputs in the two sectors of our economy is inconsistent with evidence concerning sectoral comovement. This feature, however, is shared by all multisector models in which labor is perfectly mobile (see Christiano and Fitzgerald [1998] for an early review of the evidence and implications for a two-sector RBC model). One natural way to induce comovement would be to introduce costs of reallocating labor across sectors as in Boldrin et al. (2001).

¹⁷ The welfare-consistent real wage w_t increases by more than productivity in all periods after impact because a higher number of firms puts upward pressure on labor demand. With logarithmic utility from consumption, labor supply depends on $w_t/C_t = w_{R,t}/C_{R,t}$. In

Figure 2 repeats the experiment of figure 1 for the case of translog preferences. The qualitative behavior of several variables is similar to the CES case, but key differences emerge. With translog preferences, varieties become closer substitutes as the increased product variety induces a crowding-out effect in product space. These demand-side changes, in turn, lead to lower markups. Relative to CES, the profit incentive for product creation is thus weaker and is reflected in a muted response of entry. However, the hump-shape response for overall product variety is still very similar to the CES case, and this induces the countercyclical response of the markup, μ_t , which declines over time before settling on the path back to the steady state.¹⁸ The muted response of the relative price under translog preferences implies that individual firm output does not drop below the steady state during the transition (as it did in the CES case): it is relatively more profitable to keep producing old goods, since investing in new ones erodes profit margins and yields a smaller welfare gain to consumers. This is also evident in the dynamics of labor across sectors: the reallocation of labor from product creation back into the production of existing goods takes place faster than in the CES case.

Importantly, although markups are countercyclical, aggregate profits (both welfare- and data-consistent) remain strongly procyclical. It is notoriously difficult to generate both countercyclical markups and procyclical aggregate profits in models with a constant number of producers/products (for instance, based on sticky prices). These models imply that profits become countercyclical, in stark contrast with the data (see Rotemberg and Woodford 1999). Our model naturally breaks this link between the responses of markups and aggregate profits via the endogenous fluctuation in the number of products. Procyclical product entry pushes up aggregate profits relative to the change in the product-level markup. We return to this issue when computing the second moments of our artificial economy below.¹⁹

Finally, we note that these responses differ from the efficient ones generated by solving the social planner's problem for our economy.

other words, variety has no effect on labor supply. This would no longer be the case with a different utility function.

¹⁸ The fluctuations of the markup over time also generate differences relative to the CES responses for data-consistent measures. For example, the data-consistent value of a variety $v_{R,t} = f_E/\mu(N_t)$ increases with the markup, since the latter implies a higher opportunity cost of forgone production.

¹⁹ A discussion of the responses to a permanent increase in productivity can be found in Bilbiie et al. (2007), along with a discussion of the consequences of different values for the elasticity of labor supply. The most salient feature of the responses to a permanent shock is that, with CES preferences, GDP expansion takes place entirely at the intensive margin in the short run, whereas it is entirely driven by the extensive margin (with firm-level output back at the initial steady state) in the long run. With translog preferences, extensive and intensive margin adjustments coexist in the long run.

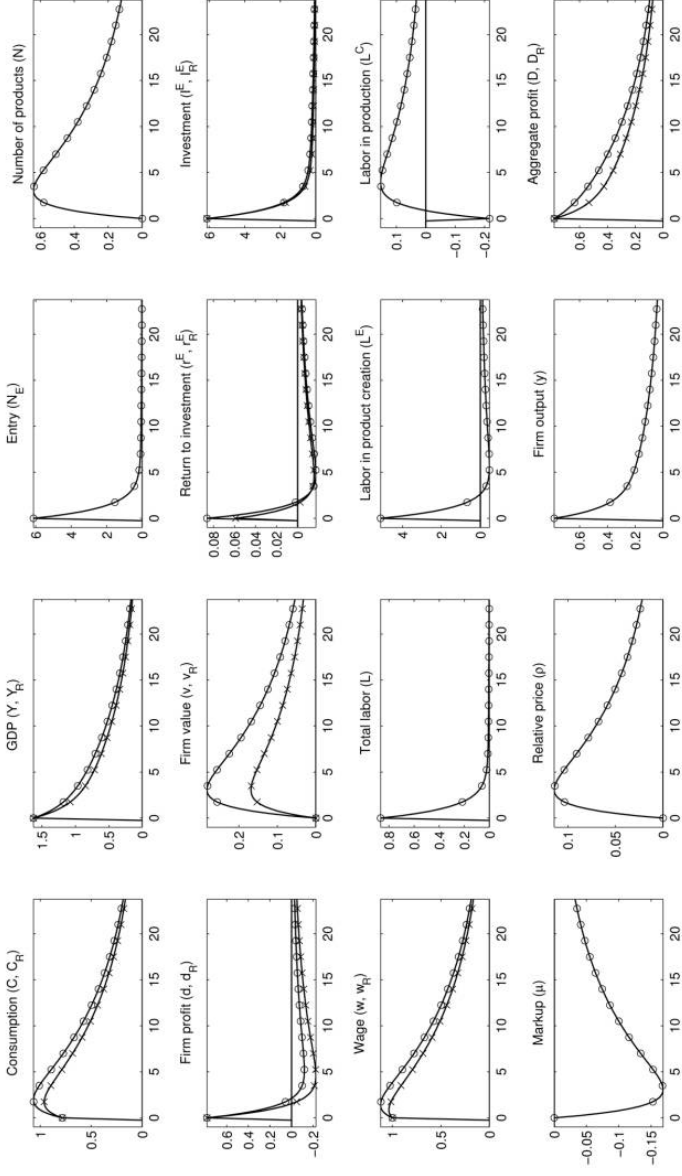


FIG. 2.—Impulse responses to a productivity increase, translog preferences (crosses denote data-consistent variables)

TABLE 3
MOMENTS FOR DATA, CES MODEL, AND BASELINE RBC

VARIABLE X	σ_X			σ_X/σ_{YR}			FIRST-ORDER AUTOCORRELA- TION			corr(X, Y_R)		
	Data	CES	RBC	Data	CES	RBC	Data	CES	RBC	Data	CES	RBC
Y_R	1.81	1.63	1.39	1.00			.84	.69	.72	1.00		
C_R	1.35	.71	.61	.74	.44	.44	.80	.75	.79	.88	.95	.94
Investment, $v_R N_E$	5.30	6.82	4.09	2.93	4.18	2.95	.87	.67	.71	.80	.99	.99
L	1.79	1.01	.67	.99	.62	.48	.88	.67	.71	.88	.98	.97

SOURCE.—King and Rebelo (1999) for data and RBC moments.

There is the standard markup distortion of the differentiated goods relative to leisure (this is also a feature of models without endogenous entry). Moreover, an intertemporal distortion occurs when the markups on goods are not synchronized over time. Finally, endogenous entry generates another distortion whenever the incentives for entry are not aligned with the welfare benefit of product variety. The CES preferences introduced by Dixit and Stiglitz (1977) represent a knife-edge case that eliminates those last two distortions. On the other hand, the translog case compounds all three distortions. See Bilbiie, Ghironi, and Melitz (2008*b*) for a detailed discussion of these distortions and the associated planner remedies.

C. Second Moments

To further evaluate the properties of our baseline model, we compute the implied second moments of our artificial economy for some key macroeconomic variables and compare them to those of the data and those produced by the benchmark RBC model. While discussing the behavior of welfare-consistent variables was important to understand the impulse responses above, here we focus only on empirically relevant variables, as we compare the implications of the model to the data. Table 3 presents the results for our CES model.²⁰ In each panel, the first number is the empirical moment implied by the US data reported in King and Rebelo (1999), the second number is the moment implied by our model, and the third number is the moment generated by King and Rebelo’s baseline RBC model. We compute model-implied second moments for Hodrick-Prescott-filtered variables for consistency with data

²⁰ The moments in table 3 change only slightly under translog preferences, without affecting the main conclusions. Details are available on request.

and standard RBC practice, and we measure investment in our model with the real value of household investment in new firms ($v_R N_E$).

Remarkably, the performance of the simplest model with entry subject to sunk costs and constant markups is similar to that of the baseline RBC model in reproducing some key features of US business cycles. Our model fares better insofar as reproducing the volatilities of output and hours. The ratio between model and data standard deviations of output is 0.90, compared to 0.77 for the standard RBC model; and the standard deviation of hours is 50 percent larger than that implied by the RBC model. On the other hand, investment is too volatile, and our baseline framework faces the same well-known difficulties of the standard RBC model: consumption is too smooth relative to output, there is not enough endogenous persistence (as indicated by the first-order autocorrelations), and all variables are too procyclical relative to the data.

Additionally, however, our model can jointly reproduce important facts about product creation and the dynamics of profits and markups: procyclical entry (as reviewed in the introduction), procyclical profits, and, in the version with translog preferences, countercyclical markups. To substantiate this point, figure 3 plots model-generated cross-correlations of entry, aggregate real profits, and GDP for CES preferences and translog preferences. In both cases, entry and profits are strongly procyclical, and the contemporaneous correlation of profits and entry is positive.²¹ Figure 4 shows the model-generated correlation of the markup with GDP at various lags and leads under translog preferences, comparing it to that documented by Rotemberg and Woodford (1999). Our model almost perfectly reproduces the contemporaneous countercyclicity of the markup; furthermore, the time profile of its correlation with the business cycle also matches well with the empirical evidence.²² There is a straightforward intuition for this result, which follows from the slow movement of the number of firms in our model: when productivity increases, GDP increases on impact and then declines

²¹ In Bilbiie et al. (2007), we show that the tent-shaped patterns in fig. 3 are not too distant from reproducing the evidence for net firm entry as measured by the difference between new incorporations and failures.

²² Of the various labor share–based empirical measures of the markup considered by Rotemberg and Woodford, the one that is most closely related to the markup in our model is the version with overhead labor, whose cyclicity is reported in col. 2 of their table 2, p. 1066, and reproduced in fig. 4. In our model, the inverse of the markup is equal to the share of production labor (labor net of workers in the “investment” sector who develop new products) in total consumption: $1/\mu_t = [w_t(L_t - L_{E,t})]/C_t$. This also implies that the share of aggregate profits in consumption is the remaining share $1 - (1/\mu_t)$. Countercyclical markups therefore entail a countercyclical profit share and a procyclical labor share, as documented by Rotemberg and Woodford (1999). Those authors also measure shares in consumption rather than GDP. Since the share of consumption in GDP is relatively acyclical, this difference in the use of denominators will not be consequential.

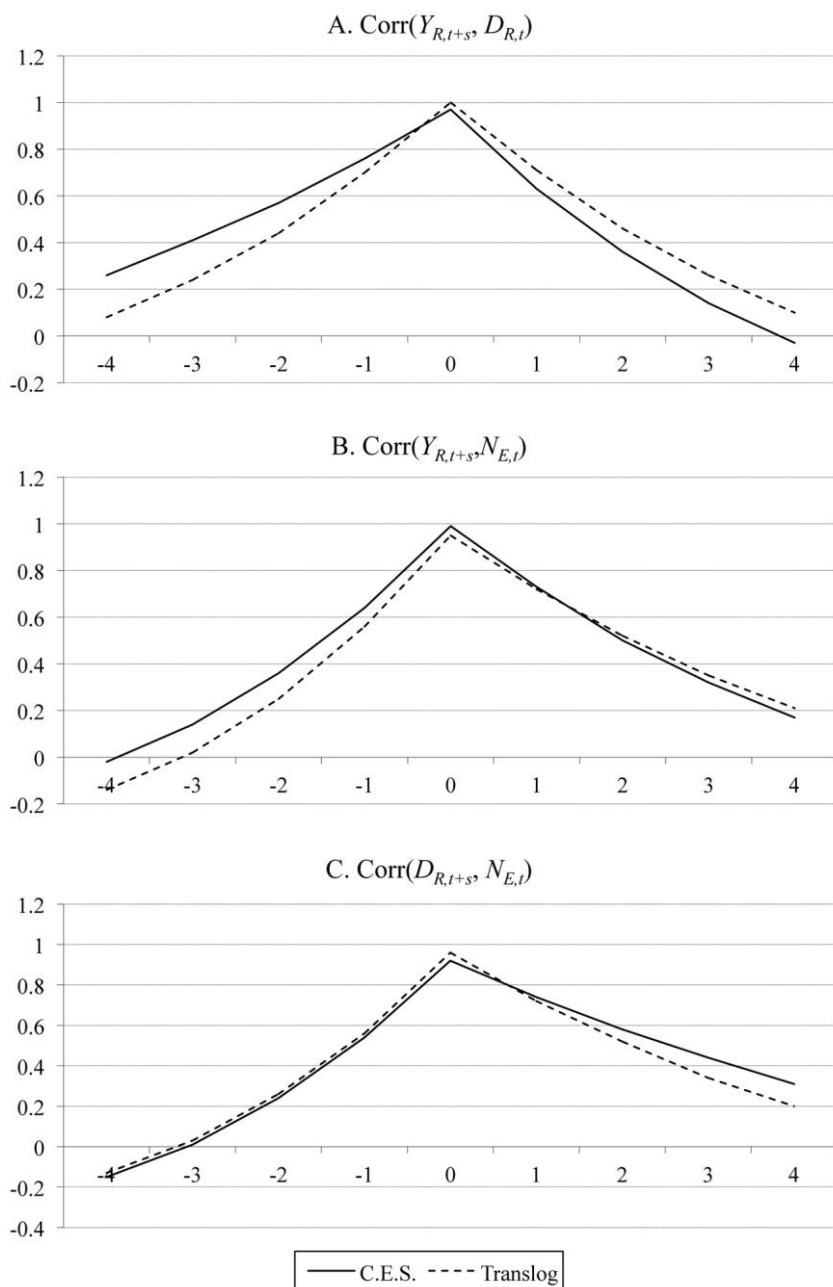


FIG. 3.—Model-based correlations: GDP, real profits, and entry

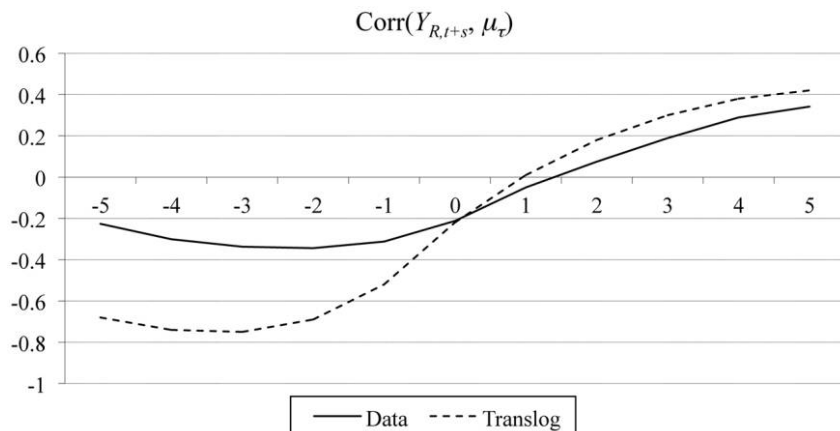


FIG. 4.—The cyclical correlation of the markup. Source for data: Rotemberg and Woodford (1999, 1066, table 2, col. 2).

toward the steady state, whereas the number of firms builds up gradually before returning to the steady state. Since the markup is a decreasing function of the number of firms, it also falls gradually in response to a technology shock. As a consequence, the markup is more negatively correlated with lags of GDP and positively correlated with its leads.

We view the performance of our model as a relative success. First, the model, although based on a propagation mechanism different from one in which traditional physical capital is absent, has second-moment properties that are comparable to the RBC model concerning macroeconomic variables of which that model speaks; indeed, our model fares better insofar as generating output and hours volatility is concerned. Second, our model can explain (at least qualitatively) stylized facts about which the benchmark RBC model is silent. Third, to the best of our knowledge, our model is the first that can account for all these additional facts simultaneously: earlier models that address entry (such as those we discuss in Sec. V) fail to account for the cyclical correlation of profits (since they assume entry subject to a period-by-period zero-profit condition), and models that generate procyclical profits (due to monopolistic competition) abstract from changes in product space. Finally, we view the ability to generate procyclical profits with a countercyclical markup and to reproduce the time pattern of the markup's correlation with the cycle in the simplest version of our model as major improvements relative to other (e.g., sticky-price-based) theories of cyclical markup variation.

TABLE 4
MOMENTS FOR DATA, CES MODEL WITH CAPITAL, AND BASELINE RBC

VARIABLE X	FIRST-ORDER AUTOCORRELA- TION											
	σ_x			σ_x/σ_{Y_R}						$\text{corr}(X, Y_R)$		
	Data	CES	RBC	Data	CES	RBC	Data	CES	RBC	Data	CES	RBC
Y_R	1.81	1.75	1.39	1.00			.84	.70	.72	1.00		
C_R	1.35	.62	.61	.74	.35	.44	.80	.63	.79	.88	.98	.94
Investment, TI_R	5.30	4.39	4.09	2.93	2.51	2.95	.87	.72	.71	.80	1.00	.99
L	1.79	1.62	.67	.99	.93	.48	.88	.71	.71	.88	.99	.97

SOURCE.—King and Rebelo (1999) for data and RBC moments.

IV. The Role of Physical Capital

We now extend our model and incorporate physical capital as well as the capital embodied in the stock of available product lines. We explore this for two reasons. First, our benchmark model studies an extreme case in which all investment goes toward the creation of new production lines and their associated products. While this is useful to emphasize the new transmission mechanism provided by producer entry, it is certainly unrealistic: part of observed investment is accounted for by the need to augment the capital stock used in production of existing goods. Second, the introduction of physical capital may improve the model's performance in explaining observed macroeconomic fluctuations. Since inclusion of capital in the model does not represent a major modeling innovation, we relegate the presentation of the augmented setup to the Appendix and limit ourselves to mentioning the main assumptions here.

We assume that households accumulate the stock of capital (K_t) and rent it to firms producing at time t in a competitive capital market. Investment in the physical capital stock (I_t) requires the use of the same composite of all available varieties as the consumption basket. Physical capital obeys a standard law of motion with rate of depreciation $\delta^K \in (0, 1)$. For simplicity, we follow Grossman and Helpman (1991) and assume that the creation of new firms does not require physical capital. Producing firms then use capital and labor to produce goods according to the Cobb-Douglas production function $y_t(\omega) = Z_t l_t(\omega)^\alpha k_t(\omega)^{1-\alpha}$, with $0 < \alpha < 1$.

As with the baseline model, we use the model with physical capital to compute second moments of the simulated economy. Table 4 reports results for key macro aggregates for CES preferences compared again to data and moments of the baseline RBC model.²³ All parameters take

²³ To save space, we do not report impulse responses for the model with capital. These, as well as second moments for the translog case (which are not significantly different from those in table 4 for the relevant variables), are available on request.

the same values as in Section III; in addition, the labor share parameter is set to $\alpha = 0.67$ and physical capital depreciation to $\delta^K = 0.025$, values that are standard in the RBC literature (e.g., King and Rebelo 1999). For comparison with investment data, we now measure investment with the real value of total investment in physical capital and new firm creation, $TI_{R,t} \equiv v_{R,t}N_{E,t} + I_{R,t}$, where $I_{R,t}$ is real investment in physical capital accumulation.

Inclusion of physical capital alters some of the key second-moment properties of the model relative to table 3. In particular, the model with capital reproduces almost the entire data variability of output and hours worked, thus clearly outperforming both our baseline and the RBC model (the ratio between model and data standard deviations of output is 0.97, and the relative standard deviation of hours is twice as large as that implied by the RBC model). The volatility of investment is also much closer to its data counterpart (whereas in our baseline model without physical capital investment was too volatile). On a more negative note, the model still generates too smooth consumption, fails to reproduce persistence, and overstates correlations; all these shortcomings are shared with the baseline RBC model and many of its extensions. Finally, the correlations pertaining to entry, profits, and markup are not significantly affected with respect to the baseline model without physical capital (results available on request). In summary, we show that the incorporation of physical capital significantly affects some of the business cycle properties of the model, in particular those pertaining to volatility of output, hours, and investment, bringing them closer to the data.

V. Discussion: Entry in Business Cycle Models

We argued that the introduction of endogenous producer entry and product variety is a promising avenue for business cycle research, for the ability of the mechanism to explain several features of evidence and improve on the basic RBC setup. To be fair, ours is not the first paper that introduces producer entry in a business cycle framework. But our model differs from earlier ones along important dimensions. In this section, we discuss the relation between our model and earlier models with producer entry, as well as some recent studies in the same vein.

Chatterjee and Cooper (1993) and Devereux, Head, and Lapham (1996*a*, 1996*b*) documented the procyclical nature of entry and developed general equilibrium models with monopolistic competition to study the effect of entry and exit on the dynamics of the business cycle. However, entry is frictionless in their models: There is no sunk entry cost, and firms enter instantaneously in each period until all profit opportunities are exploited. A fixed period-by-period cost then serves

to bound the number of operating firms. Free entry implies zero profits in all periods, and the number of producing firms in each period is not a state variable. Thus, these models cannot jointly address the procyclicality of profits and entry. In contrast, entry in our model is subject to a sunk entry cost and a time-to-build lag, and the free-entry condition equates the expected present discounted value of profits to the sunk cost.²⁴ Thus, profits are allowed to vary and the number of firms is a state variable in our model, consistent with evidence and the widespread view that the number of producing firms is fixed in the short run.²⁵ Finally, our model exhibits a steady state in which (i) the share of profits in capital is constant and (ii) the share of investment is positively correlated with the share of profits. These are among the Kaldorian growth facts outlined in Cooley and Prescott (1995), which neither the standard RBC model nor the frictionless entry model can account for (the former because it is based on perfect competition, the latter because the share of profits is zero).

Entry subject to sunk costs, with the implications that we stressed above, also distinguishes our model from more recent contributions such as those by Comin and Gertler (2006) and Jaimovich and Floetotto (2008), who also assume a period-by-period, zero-profit condition.²⁶ Our model further differs from Comin and Gertler's along three dimensions: (i) we focus on a standard definition of the business cycle, whereas they focus on the innovative notion of "medium-term" cycles; (ii) our model generates countercyclical markups due to demand-side pricing complementarities, whereas Comin and Gertler, like Galí (1995), postulate a function for markups that is decreasing in the number of firms; and (iii) our model features exogenous, RBC-type productivity shocks, whereas Comin and Gertler consider endogenous technology and use

²⁴ The pattern of product creation and destruction documented by Bernard et al. (2010) and Broda and Weinstein (2010) is also most consistent with sunk product development costs subject to uncertainty—as featured in our model.

²⁵ In fact, our model features a fixed number of producing firms within each period and a fully flexible number of firms in the long run. Ambler and Cardia (1998) and Cook (2001) take a first step in our direction. A period-by-period zero-profit condition holds only in expectation in their models, allowing for ex post profit variation in response to unexpected shocks, and the number of firms in each period is predetermined relative to shocks in that period. Benassy (1996) analyzes the persistence properties of a variant of the model developed by Devereux et al. (1996*a*, 1996*b*).

²⁶ Sunk entry costs are a feature of Hopenhayn and Rogerson's (1993) model, which is designed to analyze the employment consequences of firm entry and exit, and thus directly addresses the evidence in Davis, Haltiwanger, and Schuh (1996). However, Hopenhayn and Rogerson assume perfect competition in goods markets (as in Hopenhayn's [1992] seminal model) and abstract from aggregate dynamics by focusing on stationary equilibria in which prices, employment, output, and the number of firms are all constant. Lewis (2006) builds on the framework of this paper and estimates vector autoregression responses (including those of profits and entry as measured by net business formation) to macroeconomic shocks, finding support for the sunk-cost-driven dynamics predicted by our model.

wage markup shocks as the source of business cycles. The source of cyclical movements in markups further differentiates our work from Jaimovich and Floetotto's (and that of Cook [2001]), where countercyclical markups occur because of supply-side considerations, that is, increased competition leading to lower markups. We prefer a demand-based, preference-based explanation for countercyclical markups since data suggest that most of the entering and exiting firms are small, and much of the change in the product space is due to product switching within existing firms rather than entry of entirely new firms, pointing to a limited role for supply-driven competitive pressures in explaining markup dynamics over the business cycle.²⁷

A lively literature has emerged in the last few years that focuses on the role of producer entry and exit in the business cycle, in some cases building on our framework. Samaniego (2008) explores the issue in the heterogeneous establishment model with perfect competition of Hopenhayn and Rogerson (1993). He argues that entry and exit play little role in aggregate fluctuations. Lee and Mukoyama (2007) also build on Hopenhayn and Rogerson' work, but they conclude that the determinants of entry and exit are important for their model to match the data, and they point to the sensitivity of Samaniego's results to his assumptions on entry costs. Our entry setup differs by virtue of the assumption of imperfect competition. As we showed, this allows entry to explain features of the business cycle (such as markup dynamics) that pose a challenge to other models. Moreover, in contrast to Samaniego, we take a broader view of entry and exit as product creation and destruction that take place over the length of a cycle rather than purely entry and exit of establishments.²⁸ Other strands of the literature have focused on the consequences of alternative production and labor market structures and modes of competition for macroeconomic dynamics. For instance, Wang and Wen (2007) argue that producer entry as in our model and a Leontief production structure can reconcile flexible-price business cycle modeling with the evidence on the responses to technology shocks in Galí (1999) and Basu, Fernald, and Kimball (2006). Shao and Silos (2008) extend our model to incorporate search and matching in the labor market. They argue that firm entry introduces an endogenously time-varying value of vacancy creation, which contributes to the volatility of unemployment and generates an empirically plausible relationship between vacancies and unemployment. Colciago and Etro (2008) extend our model to consider Cournot and Bertrand competition as sources of markup variation, and they find that the extended model performs

²⁷ Dos Santos Ferreira and Dufourt (2006) motivate markup fluctuations in their model with the influence of "animal spirits" that affect firm entry and exit decisions.

²⁸ Sim (2007) develops a version of our model with heterogeneous productivity and endogenous producer exit.

better than the RBC setup at matching impulse responses and business cycle moments for US data.²⁹

VI. Conclusions

We developed a model of business cycle transmission with an endogenous number of producers and products subject to sunk entry costs, a time-to-build lag, and exogenous risk of firm destruction. The assumption of a general structure of homothetic preferences allows the model to nest the familiar CES specification with constant markups and a translog setup with time-varying markups as special cases. The model shows that variation in the number of producers and products over the horizon generally associated to the length of a business cycle can be an important propagation mechanism for fluctuations, consistent with the evidence documented by Bernard et al. (2010). Our setup explains stylized facts such as the procyclical behavior of entry and profits. Assuming translog preferences, it results in countercyclical markups with procyclical profits, resolving a puzzle for models that motivate cyclical markup variation with nominal rigidity; moreover, our model generates a time profile of the markup's correlation with the business cycle that is in line with the data. Finally, when it comes to the second-moment properties of variables that are the focus of traditional RBC models, our setup does at least as well as the latter (for a benchmark productivity process), and when we include physical capital, the model can simultaneously reproduce most of the variance of GDP, hours worked, and total investment found in the data.

There are several directions for future research. We took on the implications of a sticky-price version of our model for business cycle dynamics and the conduct of monetary policy in Bilbiie, Ghironi, and Melitz (2008*a*). The analysis of optimal monetary policy in that article is limited to a first-best environment in which the policy maker has access to lump-sum fiscal instruments. Bilbiie, Fujiwara, and Ghironi (2009) study Ramsey-optimal monetary policy in a more realistic, second-best world.³⁰ Chugh and Ghironi (2009) focus on optimal fiscal policy in a second-best environment.

However, the most important direction that our model points to for future research is empirical. The evidence reviewed in the introduction should be regarded as only a preliminary step in the direction of investigating empirically how much product creation and variety matter

²⁹ The dynamics of producer entry and exit have also received recent attention in a large number of open-economy studies. See, e.g., Ghironi and Melitz (2005) and Corsetti, Martin, and Pesenti (2007).

³⁰ See also Bergin and Corsetti (2008) for an analysis of monetary policy in a model with producer entry.

for business cycles. Indeed, our model should be viewed as providing the motivation for a deeper investigation of the empirical features of product dynamics, in the sense that a model that relies on product creation has some relative virtues in terms of explaining macroeconomic stylized facts. Ideally, data on product creation and destruction for a large set and a fine disaggregation of products at business cycle frequency would be needed for appropriate tests of our theory. Moreover, data on the product development costs at the same (or comparable) level of disaggregation would be important to gauge the relevance of sunk costs in determining product introduction over the cycle.³¹ To the best of our knowledge, these data are hitherto unavailable. Construction and investigation of these data are fundamental tasks for the future.

Appendix

A. Homothetic Consumption Preferences

Consider an arbitrary set of homothetic preferences over a continuum of goods Ω . Let $p(\omega)$ and $c(\omega)$ denote the prices and consumption level (quantity) of an individual good $\omega \in \Omega$. These preferences are uniquely represented by a price index function $P \equiv h(\mathbf{p})$, $\mathbf{p} \equiv [p(\omega)]_{\omega \in \Omega}$, such that the optimal expenditure function is given by PC , where C is the consumption index (the utility level attained for a monotonic transformation of the utility function that is homogeneous of degree 1). Any function $h(\mathbf{p})$ that is nonnegative, nondecreasing, homogeneous of degree 1, and concave uniquely represents a set of homothetic preferences. With the conventional notation used for quantities with a continuum of goods as flow values, the derived Marshallian demand for any variety ω is then given by $c(\omega)d\omega = C\partial P/\partial p(\omega)$.

B. No Option Value of Waiting to Enter

Let the option value of waiting to enter for firm ω be $\Lambda_t(\omega) \geq 0$. In all periods t , $\Lambda_t(\omega) = \max[v_t(\omega) - w_t f_E/Z_t, \beta \Lambda_{t+1}(\omega)]$, where the first term is the payoff of undertaking the investment and the second term is the discounted payoff of waiting. If firms are identical (there is no idiosyncratic uncertainty) and exit is exogenous (uncertainty related to firm death is common across firms), this becomes $\Lambda_t = \max[v_t - w_t f_E/Z_t, \beta \Lambda_{t+1}]$. Because of free entry, the first term is always zero, so the option value obeys $\Lambda_t = \beta \Lambda_{t+1}$. This is a contraction mapping because of discounting, and by forward iteration, under the assumption $\lim_{T \rightarrow \infty} \beta^T \Lambda_{t+T} = 0$ (i.e., there is a zero value of waiting when reaching the terminal period), the only stable solution for the option value is $\Lambda_t = 0$.

³¹ This will also make it possible to measure the extent of extensive margin investment in the economy, i.e., the part of investment in National Income and Product Accounts (NIPA) that goes toward enlarging the set of available goods.

C. Model Solution

We can reduce the system in table 2 to a system of two equations in two variables, N_t and C_t . To see this, write firm value as a function of the endogenous state N_t and the exogenous state f_E by combining free entry, the pricing equation, and the markup and variety effect equations:

$$v_t = f_E \frac{\rho(N_t)}{\mu(N_t)}. \quad (\text{A1})$$

The number of new entrants as a function of consumption and number of firms is $N_{E,t} = Z_t L_t / f_E - C_t / [f_E \rho(N_t)]$. Substituting this, equations (3) and (A1), and the expression for profits in the law of motion for N_t (scrolled one period forward) and the Euler equation for shares yields

$$N_{t+1} = (1 - \delta) \left[N_t + \frac{Z_t}{f_E} \left[\frac{1 - \rho(N_t)}{C_t \mu(N_t)} \frac{Z_t}{\chi} \right] - \frac{C_t}{f_E \rho(N_t)} \right], \quad (\text{A2})$$

$$f_E \frac{\rho(N_t)}{\mu(N_t)} = \beta(1 - \delta) E_t \left[\frac{C_t}{C_{t+1}} \left[f_E \frac{\rho(N_{t+1})}{\mu(N_{t+1})} + \left[1 - \frac{1}{\mu(N_{t+1})} \right] \frac{C_{t+1}}{N_{t+1}} \right] \right]. \quad (\text{A3})$$

Equations (A2) and (A3) allow us to solve for the steady-state number of firms and consumption (and therefore all other variables) by solving the equations

$$N = [\chi(r + \delta)]^{-\varphi} \left[(1 - \delta) \frac{Z}{f_E} \right]^{1+\varphi} \left[\frac{\mu(N) - 1}{N\mu(N)} \right]^\varphi \left[\delta + \frac{r + \delta}{\mu(N) - 1} \right], \quad (\text{A4})$$

$$C = \frac{(r + \delta)\rho(N)}{(1 - \delta)[\mu(N) - 1]} N f_E. \quad (\text{A5})$$

In the CES case, the markup is always equal to a constant, $\mu(N) = \theta/(\theta - 1)$, and the variety effect is governed by $\rho(N) = N^{1/(\theta-1)}$. The solution is

$$N^{\text{CES}} = \frac{1 - \delta}{\chi\theta(r + \delta)} \left[\frac{\chi\theta(r + \delta)}{\theta(r + \delta) - r} \right]^{1/(1+\varphi)} \frac{Z}{f_E}, \quad (\text{A6})$$

$$C^{\text{CES}} = \frac{(r + \delta)(\theta - 1)}{1 - \delta} f_E (N^{\text{CES}})^{\theta/(\theta-1)}. \quad (\text{A7})$$

In the translog case, the steady-state markup function is $\mu(N) = 1 + 1/(\sigma N)$. The number of firms solves the equation

$$N = \left[(1 - \delta) \frac{Z}{f_E} \right]^{1+\varphi} \left[\frac{1}{\chi(r + \delta)} \right]^\varphi \frac{[N(1 + \sigma N)]^{-\varphi}}{\delta + \sigma N(r + \delta)} \equiv H(N), \quad (\text{A8})$$

which shows that N^{Trans} is a fixed point of the function $H(N)$. Since $H(N)$ is continuous and $\lim_{N \rightarrow 0} H(N) = \infty$ and $\lim_{N \rightarrow \infty} H(N) = 0$, $H(N)$ has a unique fixed point if and only if $H'(N) \leq 0$. Straightforward differentiation of $H(N)$ shows that this is indeed the case, and hence there exists a unique N^{Trans} that solves the nonlinear equation (A8). In the special case of inelastic labor ($\varphi = 0$), a closed-form solution can be obtained as

$$N_{\varphi=0}^{\text{Trans}} = \frac{-\delta + \sqrt{\delta^2 + 4\sigma(Z/f_E)(r+\delta)(1-\delta)}}{2\sigma(r+\delta)}. \quad (\text{A9})$$

Steady-state labor effort under both preference scenarios is

$$L = \left\{ \frac{1}{\chi} \left[1 - \frac{r}{\theta(r+\delta)} \right] \right\}^{\varphi/(1+\varphi)}. \quad (\text{A10})$$

Note that hours are indeed constant relative to variation in long-run productivity.

In the quantitative exercises below, we use a specific calibration scheme, which ensures that the steady-state number of firms and markup under translog preferences are the same as under CES. (We make this assumption since we observe only one set of data and, hence, only one value for N and μ .) We can achieve this for translog preferences by an appropriate choice of the parameter σ (denoted by σ^* below). The choice of σ that ensures equalization of steady states across CES and translog preferences can be explained intuitively for the case $\varphi = 0$ with reference to figure A1. In the CES case, the relevant $H(N)$ function is a constant, and the equilibrium is given by

$$H^{\text{CES}} \equiv \frac{1-\delta}{\theta(r+\delta) - r f_E} \frac{Z}{f_E} = N,$$

represented by the dotted horizontal line. The intersection of this with the 45-degree line determines the number of firms in the steady state. Choosing the value of σ that equates the steady-state number of firms across CES and translog cases (denoted σ^*) amounts to choosing the $H(N)$ function for the translog

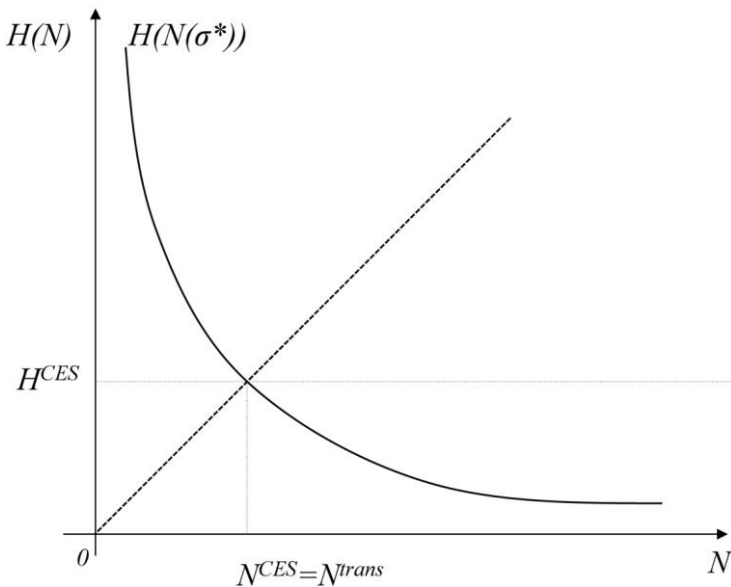


FIG. A1.—The steady-state number of firms: CES versus translog

case whose fixed point is precisely the same (i.e., that crosses the 45-degree line at the same point); this is given by the solid curve in the figure.

Algebraically, this can be achieved as follows in the general case $\varphi \geq 0$. For any preference specification, the steady-state number of firms solves equation (A4), which can be rewritten as

$$N = [\chi(r + \delta)]^{-[\varphi/(1+\varphi)]} (1 - \delta) \frac{Z}{f_E} \left\{ \left[\frac{\mu(N) - 1}{\mu(N)} \right]^{\varphi/(1+\varphi)} \left[\delta + \frac{r + \delta}{\mu(N) - 1} \right]^{1/(1+\varphi)} \right\}. \quad (A11)$$

Since the terms up to Z/f_E in the right-hand side of this equation are independent of N , equalization of N for translog and CES preferences reduces to ensuring that the last fraction is invariant to the preference specifications. That is, we need to find the value of σ that ensures that $N^{\text{Trans}} = N^{\text{CES}}$, which holds as long as

$$\frac{(1 + \sigma N^{\text{CES}})^{-[\varphi/(1+\varphi)]}}{[\delta + (r + \delta)\sigma N^{\text{CES}}]^{1/(1+\varphi)}} = \frac{\theta^{-[\varphi/(1+\varphi)]}}{[\delta + (r + \delta)(\theta - 1)]^{1/(1+\varphi)}},$$

where we used the expression for N^{CES} in (A6). It is easily verified that $\sigma^* = (\theta - 1)/N^{\text{CES}}$ is a solution and is unique (exploiting monotonicity of the markup function). Substituting the expression for N^{CES} , we can write the value of σ^* as a function of structural parameters:

$$\sigma^* = \frac{\theta - 1}{1 - \delta} \frac{[\chi\theta(r + \delta)]^{\varphi/(1+\varphi)} f_E}{[\theta(r + \delta) - r]^{-1/(1+\varphi)} Z}.$$

D. Local Equilibrium Determinacy and Nonexplosiveness

To analyze local determinacy and nonexplosiveness of the rational expectations equilibrium, we can focus on the perfect-foresight version of the system (4)–(5) and restrict attention to endogenous variables. Rearranging yields

$$\begin{bmatrix} C_{t+1} \\ N_{t+1} \end{bmatrix} = M \begin{bmatrix} C_t \\ N_t \end{bmatrix},$$

$$M \equiv \begin{bmatrix} \frac{1+r}{1-\delta} - \Theta \frac{r+\delta}{\mu-1} & \Theta\Phi - \frac{1+r}{1-\delta}(\epsilon - \eta) \\ -\frac{r+\delta}{\mu-1} & \Phi \end{bmatrix},$$

where

$$\Theta \equiv \epsilon - \eta - \frac{r+\delta}{1-\delta} \left(1 - \frac{\eta}{\mu-1} \right)$$

and

$$\Phi \equiv 1 - \delta + \frac{r+\delta}{\mu-1} \epsilon.$$

Existence of a unique, nonexplosive, rational expectations equilibrium requires that one eigenvalue of M be inside and one outside the unit circle. The characteristic polynomial of M takes the form $J(\lambda) = \lambda^2 - [\text{trace}(M)]\lambda + \det(M)$,

where the trace is

$$\text{trace}(M) = 1 - \delta + \frac{1+r}{1-\delta} + \eta \frac{r+\delta}{\mu-1} \left[1 - \frac{r+\delta}{(1-\delta)(\mu-1)} \right] + \frac{r+\delta}{1-\delta} \frac{r+\delta}{\mu-1},$$

and the determinant

$$\det(M) = 1 + r + \frac{r+\delta}{\mu-1} \frac{1+r}{1-\delta} \eta.$$

The condition for existence of a unique, nonexplosive rational expectations equilibrium is $J(-1)J(1) < 0$, where

$$J(1) = -\frac{r+\delta}{1-\delta} \left(\delta + \frac{r+\delta}{\mu-1} \right) + \eta \frac{(r+\delta)^2}{1-\delta} \frac{\mu}{(\mu-1)^2} < 0$$

if and only if

$$\eta < \frac{\mu-1}{\mu} \frac{r+\delta\mu}{r+\delta}.$$

Since $\eta \leq 0$ and the right-hand side of the latter inequality is always positive, this condition is always satisfied. Moreover, $J(-1) = 4 + 2r - J(1) > 0$ whenever $J(1) < 0$, so there exists a unique, stable, rational expectations equilibrium for any possible parameterization. The elasticity of the number of firms producing in period $t+1$ to its past level is the stable root of $J(\lambda) = 0$, that is,

$$\frac{\text{trace}(M) - \sqrt{[\text{trace}(M)]^2 - 4 \det(M)}}{2}.$$

E. The Model with Physical Capital

On the household side, we now have the capital accumulation equation (I_t is investment)

$$K_{t+1} = (1 - \delta^K)K_t + I_t, \quad (\text{A12})$$

where $\delta^K \in (0, 1)$ is the rate of depreciation, which acts as an additional dynamic constraint. The budget constraint becomes

$$B_{t+1} + v_t N_{H,t} x_{t+1} + C_t + I_t + T_t = (1 + r_t)B_t + (d_t + v_t)N_{H,t} x_t + w_t L_t + r_t^K K_t,$$

where r_t^K is the rental rate of capital. Euler equations for bonds and share holdings and the labor supply equation are unchanged. The Euler equation for capital accumulation requires

$$1 = \beta E_t \left[\left(\frac{C_{t+1}}{C_t} \right)^{-1} (r_{t+1}^K + 1 - \delta^K) \right]. \quad (\text{A13})$$

On the firm side, the production function is now Cobb-Douglas in labor and capital: $y_t(\omega) = Z_t l_t(\omega)^\alpha k_t(\omega)^{1-\alpha}$. When $\alpha = 1$, this model reduces to our previous model without physical capital. When we impose symmetry of the equilibrium, cost minimization taking factor prices w_t and r_t^K as given implies

$$r_t^K = (1 - \alpha) \frac{y_t}{k_t} \lambda_{\nu}, \quad w_t = \alpha \frac{y_t}{l_t} \lambda_{\nu}, \quad (\text{A14})$$

TABLE A1
MODEL WITH PHYSICAL CAPITAL: SUMMARY

Variable	Definition
Pricing	$\rho_t = \mu(N_t)\lambda_t$
Profits	$d_t = \left[1 - \frac{1}{\mu(N_t)}\right] \frac{Y_t^C}{N_t}$
Capital accumulation	$K_{t+1} = (1 - \delta^K)K_t + I_t$
Euler equation (capital)	$1 = \beta E_t \left[\left(\frac{C_{t+1}}{C_t}\right)^{-1} (r_{t+1}^K + 1 - \delta^K) \right]$
Aggregate accounting	$Y_t^C + v_t N_{E,t} = w_t L_t + N_t d_t + r_t^K K_t$
Total manufacturing output	$Y_t^C = C_t + I_t$
Real wage	$w_t = \frac{\alpha Y_t^C}{\mu_t L_t^C}$
Rental rate	$r_t^K = \frac{1-\alpha}{\mu_t} \frac{Y_t^C}{K_t}$
Labor in manufacturing	$Y_t^C = \rho_t Z_t (L_t^C)^\alpha K_t^{1-\alpha}$
Labor in entry	$L_t^E = N_{E,t} f_E / Z_t$

where λ_t is marginal cost. The profit function becomes $d_t = \rho_t y_t - w_t l_t - r_t^K k_t$, where optimal pricing yields $\rho_t = \mu_t \lambda_t$. Finally, market clearing for physical capital requires

$$K_{t+1} = N_{t+1} k_{t+1}, \tag{A15}$$

since capital entering $t + 1$ is rent to firms that are producing at time $t + 1$. Importantly, at the end of the period (when the capital market clears) there is a “reshuffling” of capital among firms such that there is no scrap value for the capital of disappearing firms. The other equations remain unchanged.

We have thus introduced five new variables, K_t , k_t , I_t , r_t^K , and λ_t , and five new equations (all the equations displayed above except for the budget constraint). We can write the equations as in the version without capital, using only aggregate variables. Take factor prices, multiply the numerator and denominator by N_t , and substitute out marginal cost from the pricing equation:

$$r_t^K = (1 - \alpha) \frac{\rho_t N_t y_t}{\mu_t N_t k_t} = \frac{1 - \alpha}{\mu_t} \frac{Y_t^C}{K_t},$$

$$w_t = \frac{\alpha}{\mu_t} \frac{Y_t^C}{L_t^C},$$

where $L_t^C \equiv N_t l_t$. Finally, note that labor market clearing and the profits equation are unchanged, and the resource constraint becomes

$$C_t + I_t + N_{E,t} v_t = w_t L_t + N_t d_t + r_t^K K_t.$$

The complete model can then be summarized by adding the equations in table A1 to the equations in table 2 that remain unchanged (markup, variety effect, free entry, number of firms, intratemporal optimality, and Euler equations for bonds and shares). An additional variable of interest is then total investment, $TI_t \equiv I_t + v_t N_{E,t}$, which aggregates investment in physical capital for production of consumption goods and in new firms.

The steady state.—In the steady state, the Euler equation for shares, combined with expressions for firm value, pricing and profits, yields

$$\alpha \frac{f_E}{Z} = \frac{1 - \delta}{r + \delta} [\mu(N) - 1] \frac{L^C}{N}.$$

From labor market clearing (or the aggregate accounting identity), combined with factor prices, the free-entry condition, profit and pricing equations, and the steady-state number of entrants, labor used to produce goods is

$$L^C = L - \frac{f_E}{Z} \frac{\delta}{1 - \delta} N.$$

Combining these two results, we have

$$N = (1 - \delta) Z L \left[\alpha \frac{r + \delta}{\mu(N) - 1} + \delta \right] f_E.$$

This equation yields a value for N that depends on structural parameters.³² Under translog preferences, precisely the same calibration scheme as that described in section D for the baseline model ensures that the steady-state markup and number of firms N^{Trans} are the same as under CES preferences: $\sigma^* N^{\text{CES}} = \theta - 1$. Finally, from the rental rate expression, the steady-state stock of capital can be determined once the steady-state number of firms N is known:

$$K = \left[Z \frac{1 - \alpha}{r + \delta^K} \frac{\rho(N)}{\mu(N)} \right]^{1/\alpha} \left(L - N \frac{f_E}{Z} \frac{\delta}{1 - \delta} \right).$$

All other variables can be easily determined once N and K are known.

The steady-state shares dN/Y^C and vN_E/Y^C are the same as in the model without physical capital. From the factor price expressions, the shares of physical capital and manufacturing labor income into manufacturing output Y^C are, respectively, $r^K K/Y^C = (1 - \alpha)/\mu$ and $wL^C/Y^C = \alpha/\mu$. It follows that the share of total labor income into manufacturing output is

$$\frac{wL}{Y^C} = \frac{1}{\mu} \left[\alpha + \frac{\delta}{r + \delta} (\mu - 1) \right].$$

The share of total investment is made up of two components: investment in new products/firms vN_E/Y_C and investment in new physical capital I/Y_C . The latter can be found from the expression for the rental rate, using $I/K = \delta^K$ and $r^K = r + \delta^K$, as $I/Y^C = \delta^K(1 - \alpha)/[\mu(r + \delta^K)]$. Note that the share of investment in physical capital is smaller than its RBC counterpart, $(1 - \alpha)\delta^K/(r + \delta^K)$. But the share of total investment in total GDP can be higher since it includes investment in new firms, namely (using that the share of manufacturing output into total output is $Y_C/Y = [1 + vN_E/Y_C]^{-1}$),

$$\frac{TI}{Y} = \left(\frac{\delta}{r + \delta} \frac{\mu - 1}{\mu} + \frac{\delta^K}{r + \delta^K} \frac{1 - \alpha}{\mu} \right) \frac{1}{1 + vN_E/Y_C}.$$

In principle, it is possible to use this expression to calibrate the shares of labor α and capital $1 - \alpha$ as follows. The term TI/Y can be found from NIPA data, as usual in RBC exercises. Then we can use micro data on firm (job) destruction and markups to find the share of new goods' investment in GDP and get $1 -$

³² Note that when $\alpha = 1$, we obtain the same value of N as in the model without capital.

α from the equation above (using also a standard value for physical capital depreciation).

References

- Aghion, P., and P. W. Howitt. 1991. *Endogenous Growth Theory*. Cambridge, MA: MIT Press.
- Ambler, S., and E. Cardia. 1998. "The Cyclical Behaviour of Wages and Profits under Imperfect Competition." *Canadian J. Econ.* 31:148–64.
- Axaroglou, K. 2003. "The Cyclicity of New Product Introductions." *J. Bus.* 76: 29–48.
- Basu, S., J. G. Fernald, and M. S. Kimball. 2006. "Are Technology Improvements Contractionary?" *A.E.R.* 96:1418–48.
- Benassy, J.-P. 1996. "Monopolistic Competition, Increasing Returns to Specialization and Output Persistence." *Econ. Letters* 52:187–91.
- Bergin, P. R., and G. Corsetti. 2008. "The Extensive Margin and Monetary Policy." *J. Monetary Econ.* 55:1222–37.
- Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum. 2003. "Plants and Productivity in International Trade." *A.E.R.* 93:1268–90.
- Bernard, A. B., S. J. Redding, and P. K. Schott. 2010. "Multi-product Firms and Product Switching." *A.E.R.* 100:70–97.
- Bilbiie, F. O., I. Fujiwara, and F. Ghironi. 2009. "Optimal Monetary Policy with Endogenous Product Variety." Manuscript, Paris School Econ., Bank of Japan, and Boston Coll.
- Bilbiie, F. O., F. Ghironi, and M. J. Melitz. 2007. "Endogenous Entry, Product Variety and Business Cycles." Working Paper no. 13646, NBER, Cambridge, MA.
- . 2008a. "Monetary Policy and Business Cycles with Endogenous Entry and Product Variety." In *NBER Macroeconomics Annual 2007*, edited by D. Acemoglu, K. S. Rogoff, and M. Woodford. Chicago: Univ. Chicago Press.
- . 2008b. "Monopoly Power and Endogenous Product Variety: Distortions and Remedies." Working Paper no. 14383, NBER, Cambridge, MA.
- Boldrin, M., L. J. Christiano, and J. Fisher. 2001. "Habit Persistence, Asset Returns, and the Business Cycle." *A.E.R.* 91:149–66.
- Broda, C., and D. E. Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *A.E.R.* 100:691–723.
- Caballero, R. J., and M. L. Hammour. 1994. "The Cleansing Effect of Recessions." *A.E.R.* 84:1350–68.
- Campbell, J. R. 1998. "Entry, Exit, Embodied Technology, and Business Cycles." *Rev. Econ. Dynamics* 1:371–408.
- Chatterjee, S., and R. Cooper. 1993. "Entry and Exit, Product Variety and the Business Cycle." Working Paper no. 4562, NBER, Cambridge, MA.
- Christiano, L. J., and T. J. Fitzgerald. 1998. "The Business Cycle: It's Still a Puzzle." *Fed. Reserve Bank Chicago Econ. Perspectives* 22:56–83.
- Chugh, S. K., and F. Ghironi. 2009. "Optimal Fiscal Policy with Endogenous Product Variety." Manuscript, Univ. Maryland and Boston Coll.
- Colciago, A., and F. Etro. 2008. "Endogenous Market Structures and Business Cycles." Manuscript, Univ. Milan Bicocca.
- Comin, D., and M. Gertler. 2006. "Medium Term Business Cycles." *A.E.R.* 96: 523–51.
- Cook, D. 2001. "Time to Enter and Business Cycles." *J. Econ. Dynamics and Control* 25 (8): 1241–61.

- Cooley, T. F., and E. C. Prescott. 1995. "Economic Growth and Business Cycles." In *Frontiers of Business Cycle Research*, edited by T. F. Cooley, 1–38. Princeton, NJ: Princeton Univ. Press.
- Corsetti, G., P. Martin, and P. Pesenti. 2007. "Productivity Spillovers, Terms of Trade, and the 'Home Market Effect.'" *J. Internat. Econ.* 73:99–127.
- Davis, S., J. C. Haltiwanger, and S. Schuh. 1996. *Job Creation and Destruction*. Cambridge, MA: MIT Press.
- Den Haan, W. J., and P. Sedlacek. 2010. "Inefficient Employment Decisions, Entry Costs, and the Cost of Fluctuations." Manuscript, Univ. Amsterdam.
- Devereux, M. B., A. C. Head, and B. J. Lapham. 1996a. "Aggregate Fluctuations with Increasing Returns to Specialization and Scale." *J. Econ. Dynamics and Control* 20:627–56.
- . 1996b. "Monopolistic Competition, Increasing Returns, and the Effects of Government Spending." *J. Money, Credit and Banking* 28:233–54.
- Dixit, A. K., and J. E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *A.E.R.* 67:297–308.
- Dos Santos Ferreira, R., and F. Dufourt. 2006. "Free Entry and Business Cycles under the Influence of Animal Spirits." *J. Monetary Econ.* 53:311–28.
- Dunne, T., M. J. Roberts, and L. Samuelson. 1988. "Patterns of Firm Entry and Exit in US Manufacturing Industries." *RAND J. Econ.* 19:495–515.
- Feenstra, R. C. 2003. "A Homothetic Utility Function for Monopolistic Competition Models, without Constant Price Elasticity." *Econ. Letters* 78:79–86.
- Galí, J. 1995. "Product Diversity, Endogenous Markups, and Development Traps." *J. Monetary Econ.* 36:39–64.
- . 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *A.E.R.* 89:249–71.
- Ghironi, F., and M. J. Melitz. 2005. "International Trade and Macroeconomic Dynamics with Heterogeneous Firms." *Q.J.E.* 120:865–915.
- Grossman, G. M., and E. Helpman. 1991. *Innovation and Growth in the Global Economy*. Cambridge, MA: MIT Press.
- Harrison, H. 2003. "Returns to Scale and Externalities in the Consumption and Investment Sectors." *Rev. Econ. Dynamics* 6 (4): 963–76.
- Hopenhayn, H. A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60:1127–50.
- Hopenhayn, H. A., and R. Rogerson. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *J.P.E.* 101:915–38.
- Jaimovich, N., and M. Floetotto. 2008. "Firm Dynamics, Markup Variations, and the Business Cycle." *J. Monetary Econ.* 55:1238–52.
- Jovanovic, B. 2006. "Investment Options and the Business Cycle." Manuscript, New York Univ.
- Judd, K. L. 1985. "On the Performance of Patents." *Econometrica* 53:567–85.
- Kaldor, N. 1957. "A Model of Economic Growth." *Econ. J.* 67:591–624.
- King, R. G., C. I. Plosser, and S. T. Rebelo. 1988. "Production, Growth, and Business Cycles: I. The Basic Neoclassical Model." *J. Monetary Econ.* 21:195–232.
- King, R. G., and S. T. Rebelo. 1999. "Resuscitating Real Business Cycles." In *Handbook of Macroeconomics*, vol. 1B, edited by J. B. Taylor and M. Woodford, 927–1007. Amsterdam: Elsevier.
- Krugman, Paul R. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *J. Internat. Econ.* 9 (4): 469–79.
- Kydland, F. E., and E. C. Prescott. 1982. "Time to Build and Aggregate Fluctuations." *Econometrica* 50:1345–70.

- Lee, Y., and T. Mukoyama. 2007. "Entry, Exit, and Plant-Level Dynamics over the Business Cycle." Working Paper no. 07-18, Fed. Reserve Bank Cleveland.
- Lewis, V. 2006. "Macroeconomic Fluctuations and Firm Entry: Theory and Evidence." Manuscript, Catholic Univ. Leuven.
- Long, J. B., Jr., and C. I. Plosser. 1983. "Real Business Cycles." *J.P.E.* 91 (1): 39–69.
- Romer, P. 1990. "Endogenous Technological Change." *J.P.E.* 98:71–102.
- Rotemberg, J. J., and M. Woodford. 1992. "Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity." *J.P.E.* 100:1153–1207.
- . 1999. "The Cyclical Behavior of Prices and Costs." In *Handbook of Macroeconomics*, vol. 1B, edited by J. B. Taylor and M. Woodford, 1051–1135. Amsterdam: Elsevier.
- Samaniego, R. M. 2008. "Entry, Exit and Business Cycles in a General Equilibrium Model." *Rev. Econ. Dynamics* 11:529–41.
- Schumpeter, J. A. 1934. *The Theory of Economic Development*. Cambridge, MA: Harvard Univ. Press.
- Shao, E., and P. Silos. 2008. "Firm Entry and Labor Market Dynamics." Manuscript, San Francisco State Univ. and Fed. Reserve Bank Atlanta.
- Shleifer, A. 1986. "Implementation Cycles." *J.P.E.* 94 (6): 1163–90.
- Sim, J. W. 2007. "Uncertainty, Firm Dynamics and Business Cycles." Manuscript, Board Governors, Fed. Reserve System, Washington, DC.
- Uhlig, H. 1999. "A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily." In *Computational Methods for the Study of Dynamic Economies*, edited by R. Marimon and A. Scott, 30–61. Oxford: Oxford Univ. Press.
- Wang, P., and Y. Wen. 2007. "A Defense of RBC: Understanding the Puzzling Effects of Technology Shocks." Manuscript, Hong Kong Univ. Sci. and Technology and Fed. Reserve Bank St. Louis.
- Woodford, M. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton Univ. Press.