Inventory Accelerator in General Equilibrium*

Pengfei Wang
Hong Kong University of Science & Technology

Yi Wen
Federal Reserve Bank of St. Louis
and
Tsinghua University

First Version: December, 2008. This Version: March 6, 2011

Abstract

We develop a general-equilibrium model of inventories with explicit microfoundations by embedding the production-cost–smoothing motive (e.g., Eichenbaum, 1989) into a DSGE model with imperfect competition. We show that monopolistic firms facing idiosyncratic cost shocks have incentives to bunch production and smooth sales by carrying inventories. The model is broadly consistent with key stylized facts of aggregate inventory fluctuations, such as the procyclical inventory investment and the countercyclical inventory-to-sales ratio. In addition, the model yields novel predictions for the role of inventories in macroeconomic stability: Inventories may greatly amplify and propagate the business cycle, provided that markups or the variance of idiosyncratic cost shocks are sufficiently large. That is, a strong incentive to accumulate inventories under the cost-smoothing motive at the firm level may give rise to hump-shaped aggregate output dynamics and significantly higher volatility of GDP. Such predictions are in contrast to the implications of the recent general-equilibrium inventory literature, which shows that inventory investment induced by more conventional mechanisms (e.g., the stockout-avoidance motive and the (S,s) rule) does not increase the variance of aggregate output.

Keywords: Inventories, Inventory-Accelerator, Business Cycle, General-Equilibrium Inventory Theory, Hump-Shaped Output Dynamics, Stock-Adjustment Model.

JEL codes: E13, E20, E32.

*We thank three anonymous referees and the co-editor for very helpful comments, and Judy Ahlers for editorial assistance. Correspondence: Yi Wen, Research Division, Federal Reserve Bank of St. Louis, St. Louis, MO 63144. Phone: 314-444-8559. Fax: 314-444-8731. Email: yi.wen@stls.frb.org.

1 Introduction

"Relative to its importance in business fluctuations, inventory investment must be the most under-researched aspect of macroeconomic activity" (Blinder, 1981, p.444). Blinder's assessment probably remains true today despite his and others' best efforts in developing a well-established and empirically validated theory of inventories over the past quarter century.[1] The general-equilibrium inventory literature with rigorous microfoundations is sparse,[2] although the empirical evidence continues to remind us of Blinder's (1990) famous claim that "business cycles are, to a surprisingly large degree, inventory cycles."[3] In particular, the "Great Moderation" of the U.S. economy since the 1980s appears to be paralleled by a significant reduction in inventory volatility and inventory-to-sales ratio (especially for durable goods), which has led some (e.g., Kahn, McConnell, and Perez-Quiros, 2002) to argue that reduced inventories due to improved inventory-management technologies explain the reduced volatility in gross domestic product (GDP).[4]

Overwhelming empirical evidence indicates that the variance of production is larger than that of sales and inventory investment is procyclical in a variety of sectors and subsectors and in the entire economy. Because of this, conventional wisdom views inventories as a destabilizing force to the economy and key to understanding the business cycle.[5]

However, the conventional wisdom is based on a partial-equilibrium logic: *Given* sales, procyclical inventory investment implies a higher variance of production; hence, output is more variable than it would be if inventories did not exist or were not procyclical. Such an argument ignores the possible general-equilibrium effects of inventories on sales through prices. Indeed, a recent general-equilibrium inventory literature challenges Blinder's view that inventories are key to understanding economic fluctuations. Khan and Thomas (KT) (2007a) and Wen (2008) develop dynamic stochastic general-equilibrium (DSGE) models in which inventories are rigorously introduced through firms' optimization behavior via either the (S,s) policy or the stockout-avoidance motive and show that procyclical inventory investment does not increase the volatility of output.[6] This is so be-

[1] An incomplete list of important early works includes Blanchard (1983), Blinder (1981, 1986a, 1986b), Blinder and Maccini (1991), Eichenbaum (1989), Kahn (1987), Ramey (1991), West (1986), and many others.

[2] Exceptions include Fisher and Hornstein (2000), Khan and Thomas (2007a, 2007b), Kryvtsov and Midrigany (2008), and Wen (2008).

[3] Inventory investment accounts for less than 1% of GDP, but its movement accounts for more than 60% of the variations in GDP (see, e.g., Blinder, 1981 and 1986a; and Blinder and Maccini, 1991). Using updated data, Romer (2001, p170, Table 4.2) shows that declines in inventory investment still account for more than 40% of the drop in GDP for post-war U.S. recessions.

[4] This view is controversial. See Kahn (2008), Irvinea and Schuh (2005a, 2005b), Iacoviello, Schiantarelli, and Schuh (2007), and Ramey and Vine (2006), among others for more recent empirical studies on this issue.

[5] See the previously cited literature in footnote 1 for reference.

[6] In particular, Wen (2008) shows that inventories *reduce* the variance of GDP.

cause in general equilibrium inventories may stabilize sales as much as (or even more than) they destabilize production. For example, in the model of KT (2007a), a rise in the incentive for ordering intermediate goods and accumulating input inventories diverts productive resources (such as labor) away from the final-goods sector to intermediate-goods production, thereby dampening the rise in final sales. In Wen's (2008) model, the asset price of inventories is procyclical because of an endogenously determined probability of inventory stockout; hence, it attenuates the increase of sales in booms and mitigates the decrease of demand in recessions. Consequently, these papers show that eliminating inventories from the economy does not necessarily decrease the variance of GDP, contradicting the conventional wisdom that business cycles are driven largely by inventory fluctuations.

Although the general-equilibrium analyses of KT (2007a) and Wen (2008) are provocative, there exist other theoretical possibilities linking inventories to output volatility and alternative motives inducing firms to hold inventories. For example, firms may use inventories to smooth sales when production costs are uncertain. That is, profit-maximizing firms (facing cost shocks) may opt to "bunch" production by producing more than sales and carry the excess supply as inventories when costs are low, and use inventories to meet demand when costs are high. This is the production-cost–smoothing motive emphasized by Eichenbaum (1989). Empirical studies by Eichenbaum (1989) and others show that cost shocks are indeed important in explaining inventory fluctuations at the industry level.[7]

This paper provides a microfounded general-equilibrium model of inventories based on the cost-smoothing motive. A fundamental challenge for building a general-equilibrium inventory model with the production-cost–smoothing motive is that the rate of return to inventory investment is negative and dominated by that of capital accumulation. Kydland and Prescott (1982) have bypassed this difficulty by assuming inventory as a factor of production.[8] This paper confronts this problem directly by introducing three key frictions into an otherwise standard real-business-cycle (RBC) model: (i) imperfect competition, (ii) heterogeneous productivity across firms, and (iii) borrowing constraints. The first two frictions imply that a firm's profit function is strictly concave in sales and marginal costs are idiosyncratic; consequently, inventories can emerge as an optimal device to smooth sales and production costs.[9] The third friction avoids infinite inventory holdings for low-cost firms under constant returns to scale technology. With these frictions, there exists a

[7]Costs shocks as a distinct source of uncertainty driving inventory behavior are also emphasized by Blanchard (1983), Eichenbaum (1984, 1989), Durlauf and Maccini (1995), Ramey (1989), and West (1986), among others. Also see the references in Eichenbaum (1989). However, this literature does not provide an explicit microtheory to explain the existence of inventories. For example, in this literature the existence of an optimal target inventory level is assumed, rather than derived from firms' cost-minimization problems.

[8]In fact, proving the existence of inventories in a frictionless general-equilibrium model is as difficult as proving the existence of money.

[9]Inventories do not exist in a representative-agent model even if the profit function is concave in sales.

well-defined distribution of inventory stocks across firms in equilibrium.[10]

In contrast to the partial-equilibrium inventory literature, the optimal inventory target of a firm and the associated stock-adjustment equation are derived explicitly from firms' cost-minimization problems. We show that the general-equilibrium model can explain the key stylized facts of aggregate inventory behavior, such as the procyclical inventory investment and the countercyclical inventory-to-sales ratio. However, in contrast to KT (2007a) and Wen (2008), which are based either on the (S,s) inventory strategy or the stockout-avoidance policy, our model predicts that procyclical inventory investment may significantly amplify the volatility of aggregate output. More importantly, it may also help propagate aggregate shocks by generating hump-shaped output dynamics, suggesting that inventory investment may indeed play a key role in the business cycle. This potential role of inventories in propagating the business cycle was first studied by Metzler (1941) but has not been emphasized in the recent theoretical literature.

The intuition behind our results is as follows. The incentives for firms to bunch production and use inventories to smooth sales under idiosyncratic cost shocks imply procyclical inventory investment and more variable production than sales at the firm level. However, the ratio of aggregate inventory stock to sales is countercyclical despite procyclical inventory investment, because a general-equilibrium trade-off between inventories and capital prevents aggregate inventories from rising sharply in the initial periods of a boom: Under a positive aggregate productivity shock, the returns to capital are high, so the demand for capital outweighs the demand for inventories initially. This causes more firms to stockout to meet the rise in final-goods demand so as to economize on the high interest costs of holding inventories. Hence, aggregate inventories do not increase one-for-one with sales in the initial periods of the shock, leading to a countercyclical inventory-to-sales ratio. However, because inventories facilitate sales, inventory investment will eventually accelerate in subsequent periods as the interest rate declines and the production capacity peaks (which relaxes firms' borrowing constraints). The rising inventory stock encourages more sales, which in turn translates into more production capacity and further relaxes firms' borrowing constraints. Such a contemporaneous tradeoff but a dynamic reinforcement between capital and inventories give rise to hump-shaped inventory investment and highly persistent aggregate demand. Thus, in contrast to a standard RBC model without inventories, aggregate consumption smoothing in our model is achieved not only through sharp rises in capital investment in the initial periods but also gradual increases in inventory investment in subsequent periods. This makes sense because imperfectly competitive firms hold inventories to smooth sales. This gradual stock-adjustment process creates an inventory-accelerator mechanism and helps to propagate and amplify the business cycle.

[10] A virtue of our approach is that it is analytically tractable with closed-form solutions for firms' inventory decision rules, in contrast to the (S,s) general-equilibrium inventory models (e.g., Fisher and Hornstein, 2000; and Khan and Thomas, 2007a). Our strategy to make the model analytically tractable is inspired by the general-equilibrium approach of Wen (2008).

A similar stock-adjustment mechanism does not lead to more volatile and persistent aggregate demand in the two-sector (S,s)-inventory model of KT (2007a), because in their model a positive productivity shock to the intermediate-goods sector diverts resources from the final-goods sector, which dampens the rise in final sales and capital accumulation. In our model, firms are heterogeneous in marginal costs of production. Thus, in each period the less-efficient firms (which face high marginal costs) do not accumulate inventories and opt to meet demand by stocking out, while only the more-efficient firms accumulate inventories and opt to produce more than sales. Since a positive aggregate productivity shock raises the real interest rate and thereby increases the number of firms that want to stockout, it strengthens final sales and dampens the initial rise in aggregate inventory investment along the extensive margin, making inventory investment hump-shaped through the stock-adjustment mechanism. This generates more persistence in aggregate demand, therefore increasing the volatility of GDP.

Sales smoothing at the firm level is a key micromechanism that renders inventories destabilizing in our general-equilibrium model. Under sales smoothing, the conventional partial-equilibrium argument about the destabilizing role of inventories becomes approximately valid: *Given* sales, procyclical inventory investment implies a higher variance of production; hence, output is more variable than it would be without inventories. Although this partial-equilibrium argument ignores general-equilibrium feedbacks of inventories on sales, it nonetheless holds approximately true in our model because sales smoothing makes sales a less-ideal buffer to dampen the destabilizing effects of inventories.[11]

While useful for understanding dynamic interactions between inventories and the business cycle, the simplicity of our benchmark model limits its quantitative fit in several respects. First, it understates the volatility of inventory investment. Second, it can match the observed average inventory-to-sales ratio only if the implied markup is at least 20% or higher, or if firms are not (severely) borrowing constrained. Third, when the parameters are calibrated to generate hump-shaped dynamics in aggregate employment and output, the benchmark model significantly understates the dispersion of prices across firms. Although assuming alternative distributions for firms' idiosyncratic shocks may mitigate some of the problems, it cannot address all problems at once and may also create new problems. Therefore, further research is still needed to firmly establish the Metzler-type inventory-accelerator mechanism in general equilibrium.

The rest of the paper is organized as follows. Section 2 presents the benchmark model of heterogeneous firms and shows how to derive closed-form decision rules for production, sales, and inventory investment under borrowing constraints. Section 3 studies the general equilibrium of

[11] On the other hand, the procyclical movements of inventory-depleting firms along the extensive margin further amplify the destabilizing effect of inventories under sales smoothing, and such movements are key to generating a countercyclical stock-to-sales ratio and hump-shaped dynamics. If the extensive margin were constant or countercyclical, inventories would still be destabilizing to GDP in the model (because of sales smoothing), but the hump-shaped dynamics would disappear and the inventory-to-sales ratio might become procyclical.

the model. Section 4 calibrates the model and studies its quantitative implication for aggregate inventory dynamics and the business cycle. Section 5 extends the benchmark model along several dimensions and investigates the robustness of the results. These extensions and robustness analyses provide further insights into the working mechanism of the inventory model. Section 6 concludes the paper. An analytical solution method for solving the model with persistent idiosyncratic shocks is provided in the Appendix.

2 The Benchmark Model

This is a model of output inventories. The framework can be easily extended to input inventories.[12] There are two types of goods in the economy: final goods and intermediate goods. The final goods sector is perfectly competitive and uses a continuum of intermediate goods to produce output according to the technology,

$$Z_t = \left[\int_0^1 y_t(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}. \tag{1}$$

Given prices of intermediate goods, $p_t(i)$, the inverse demand function of intermediate good i is given by

$$p_t(i) = y_t(i)^{-\frac{1}{\sigma}} Z_t^{\frac{1}{\sigma}}, \tag{2}$$

where the final-good price has been normalized to 1. Each type of intermediate good i is supplied by a monopolist firm, which produces output according to

$$x_t(i) = A_t \varepsilon_t(i) k_t(i)^{\alpha} n_t(i)^{1-\alpha}, \tag{3}$$

where $\varepsilon_t(i)$ is the inverse of an idiosyncratic cost shock to firm i, A_t is an aggregate total factor productivity (TFP) shock, $k_t(i)$ is capital, and $n_t(i)$ is labor. The factor markets are competitive, so intermediate-goods firms take the real rental rate of capital (r_t) and the real wage (w_t) as given. The factor demand functions are given by $k_t(i) = \alpha \frac{x_t(i)}{r_t+\delta_k} \phi_t(i)$ and $n_t(i) = (1-\alpha) \frac{x_t(i)}{w_t} \phi_t(i)$, where $\phi_t(i)$ denotes the marginal cost of firm i and $r_t + \delta_k$ the user's cost of capital. These factor demand functions imply $\frac{1}{A_t} \left(\frac{r_t+\delta_k}{\alpha} \right)^a \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha} = \varepsilon_t(i) \phi_t(i)$. We can define $\Phi_t \equiv \frac{1}{A_t} \left(\frac{r_t+\delta_k}{\alpha} \right)^a \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}$ as the aggregate marginal cost and it satisfies $\phi_t(i) = \frac{\Phi_t}{\varepsilon_t(i)}$.

Defining $s_t(i)$ as the stock of inventories that firm i decides to hold in period t, the firm's program is to maximize the expected sum of future profits by solving

[12] See Humphreys, Maccini, and Schuh (2001) for partial-equilibrium analysis of both input and output inventories.

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{\Lambda_t}{\Lambda_0} \left\{ p_t(i) y_t(i) - \frac{\Phi_t}{\varepsilon_t(i)} x_t(i) \right\} \tag{4}$$

subject to

$$y_t(i) + s_t(i) = x_t(i) + (1 - \delta_s) s_{t-1}(i) \tag{5}$$

$$s_t(i) \geq 0 \tag{6}$$

$$p_t(i) y_t(i) - \frac{\Phi_t}{\varepsilon_t(i)} x_t(i) \geq 0, \tag{7}$$

where Λ_t is the marginal utility of consumption of the household and δ_s the depreciation rate of inventories. Equation (5) is the resource constraint, equation (6) is a nonnegativity constraint on inventory stock, and equation (7) is a nonnegativity constraint on dividends.[13]

Because the revenue function, $p_t(i) y_t(i) = y_t(i)^{1-\frac{1}{\sigma}} Z_t^{\frac{1}{\sigma}}$, is concave in sales and the cost function is linear in production (due to constant returns to scale), firms have incentives to smooth both sales and production costs by accumulating inventories, which can maximize average profits and reduce average costs through intertemporal substitution of production activities. For example, when $\varepsilon_t(i)$ is large (marginal cost is low), firm i can produce more than sales (up to the point of a zero profit) and use inventories to substitute for future production when the next-period marginal cost may be high. On the other hand, when $\varepsilon_t(i)$ is small (marginal cost is high), the firm can use inventories to satisfy sales without raising production costs.

Denoting $\{\lambda_t(i), \pi_t(i), \mu_t(i)\}$ as the Lagrangian multipliers of constraints (5)-(7), respectively, the first-order conditions of $\{x_t(i), y_t(i), s_t(i)\}$ are given by

$$\frac{\Phi_t}{\varepsilon_t(i)} (1 + \mu_t(i)) = \lambda_t(i) \tag{8}$$

[13] A nonnegative dividend constraint is a standard assumption in the investment literature with heterogeneous firms and incomplete markets (see, e.g., Browna and Haeglerb, 2004; and Miao, 2005). Gertler and Gilchrist (1994) document empirically that small firms are more likely to be borrowing constrained than large firms, and they play a surprisingly prominent role in the fluctuations of aggregate inventory demand over the business cycle. Borrowing constraints are needed in our model because the linear cost function may induce a low-cost firm to produce infinite output and hold infinite inventories by paying infinitely negative dividends to households. Hence, the firm's optimization problem is not well defined unless we impose a borrowing limit to rule out infinitely negative dividends. However, this assumption alone does not lead to a countercyclical aggregate inventory-to-sales ratio under aggregate shocks for two reasons: (i) Idiosyncratic cost shocks are the main reason for firms to hold inventories in our model, so borrowing constraints determine only the *average* inventory-to-sales ratio across firms but not the dynamic responses of this ratio to aggregate shocks. The dynamics of the aggregate inventory-to-sales ratio are critically affected by shifts in the distribution of firms along the extensive margin (i.e., changes in the number of inventory-holding firms). (ii) For this reason, allowing for negative profits can increase the steady-state inventory level more than it can increase steady-state sales; consequently, the volatility of inventory stock relative to that of sales under aggregate shocks may be reduced and the aggregate stock-to-sales ratio may become even more (instead of less) countercyclical with negative dividends. Hence, the nonnegative dividend assumption does not play exactly the same role as the assumption of increasing marginal costs in Bils and Kahn (2000), although it is consistent with increasing marginal costs at the firm level. Other types of constraints to prevent infinitely negative dividends (such as decreasing returns to scale production technology and adjustment costs in inventory investment) are possible but are not as tractable. See the robustness analysis on relaxing this assumption in Section 5.

$$\left(\frac{\sigma-1}{\sigma}\right) y_t(i)^{-\frac{1}{\sigma}} Z_t^{\frac{1}{\sigma}} \left(1+\mu_t(i)\right) = \lambda_t(i) \tag{9}$$

$$\lambda_t(i) = \beta(1-\delta_s) E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i) + \pi_t(i). \tag{10}$$

Equations (8) and (9) imply $y_t(i) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} Z_t \left(\frac{\varepsilon_t(i)}{\Phi_t}\right)^{\sigma}$, which determines the optimal amount of sales. Consequently, the monopoly price is a markup over the marginal cost,

$$p_t(i) = \frac{\sigma}{\sigma-1} \frac{\Phi_t}{\varepsilon_t(i)}. \tag{11}$$

Notice that, in the absence of idiosyncratic uncertainty, the incentives to hold inventories are diminished because equation (10) implies $\pi_t = \lambda_t - \beta(1-\delta_s)E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}$, which is greater than 0 in the steady state. That is, aggregate shocks do not induce inventory investment near the steady state. This is why Kydland and Prescott (1982) assume inventory as a factor of production in their representative-agent model.

The decision rules for inventories are characterized by a cutoff strategy. Consider two possibilities as follows:

Case A. $\varepsilon_t(i) \geq \varepsilon_t^*$. In this case, the marginal cost of production is low. Suppose $s_t(i) > 0$, then $\pi_t(i) = 0$. In such a case, equation (10) implies $\lambda_t(i) = \beta(1-\delta_s)E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i)$, and equation (8) implies

$$\frac{\Phi_t}{\varepsilon_t(i)} \leq \beta(1-\delta_s) E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i), \tag{12}$$

because $\mu_t(i) \geq 0$. This implies $\varepsilon_t(i) \geq \frac{\Phi_t}{\beta(1-\delta_s)E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i)} \equiv \varepsilon_t^*$, which defines the cutoff value ε_t^* and the relationship

$$\beta(1-\delta_s) E_t \frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i) \equiv \frac{\Phi_t}{\varepsilon_t^*}. \tag{13}$$

Notice that the cutoff is independent of i under the assumption of $i.i.d$ idiosyncratic shocks. Equation (8) then further implies $1 + \mu_t(i) = \frac{\varepsilon_t(i)}{\varepsilon_t^*}$. Hence, we conclude that $\mu_t(i) > 0$ if $\varepsilon_t(i) > \varepsilon_t^*$. In such a case, the nonnegative profit constraint binds, $p_t(i) y_t(i) = \frac{\Phi_t}{\varepsilon_t(i)} x_t(i)$, which together with equation (11) implies

$$x_t(i) = \frac{\sigma}{\sigma-1} y_t(i) > y_t(i), \tag{14}$$

suggesting that inventory investment is strictly positive. That is, in the case of a large enough idiosyncratic productivity shock (or small enough cost shock), the firm produces more than sales and opts to hold the excess supply as inventories.

Case B. $\varepsilon_t(i) < \varepsilon_t^*$. In this case, the marginal cost of production is high. Suppose $p_t(i)y_t(i) > \frac{\Phi_t}{\varepsilon_t(i)}x(i)$, then $\mu_t(i) = 0$. Then equations (8) and (13) imply $\frac{\Phi_t}{\varepsilon_t(i)} = \lambda_t(i) = \frac{\Phi_t}{\varepsilon_t^*} + \pi_t(i)$, which implies $\pi_t(i) = \frac{\Phi_t}{\varepsilon_t(i)} - \frac{\Phi_t}{\varepsilon_t^*} > 0$. Hence, we have $s_t(i) = 0$. In such a case, the firm opts to stockout and the resource identity implies $x_t(i) = y_t(i) - (1-\delta_s)s_{t-1}(i)$. That is, in the case of a high marginal cost, the firm cuts back production and uses existing inventories to satisfy sales.

The decision rules of the firm can thus be summarized by the following policy functions:

$$y_t(i) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} Z_t \left(\frac{\varepsilon_t(i)}{\Phi_t}\right)^{\sigma} \tag{15}$$

$$x_t(i) = \begin{cases} \frac{\sigma}{\sigma-1} y_t(i) & \text{if } \varepsilon_t(i) \geq \varepsilon_t^* \\ y_t(i) - (1-\delta_s)s_{t-1}(i) & \text{if } \varepsilon_t(i) < \varepsilon_t^* \end{cases} \tag{16}$$

$$s_t(i) = \begin{cases} \frac{1}{\sigma-1} y_t(i) + (1-\delta_s)s_{t-1}(i) & \text{if } \varepsilon_t(i) \geq \varepsilon_t^* \\ 0 & \text{if } \varepsilon_t(i) < \varepsilon_t^* \end{cases}. \tag{17}$$

Since the shadow value of inventory satisfies

$$\lambda_t(i) = \begin{cases} \frac{\Phi_t}{\varepsilon_t^*} & \text{if } \varepsilon_t(i) \geq \varepsilon_t^* \\ \frac{\Phi_t}{\varepsilon_t(i)} & \text{if } \varepsilon_t(i) < \varepsilon_t^* \end{cases}, \tag{18}$$

equation (13) becomes

$$\frac{\Phi_t}{\varepsilon_t^*} = \beta(1-\delta_s)E_t \frac{\Lambda_{t+1}}{\Lambda_t}\left[\int_{\varepsilon_{t+1}<\varepsilon_{t+1}^*} \frac{\Phi_{t+1}}{\varepsilon_{t+1}} dF(\varepsilon) + \int_{\varepsilon_{t+1}\geq\varepsilon_{t+1}^*} \frac{\Phi_{t+1}}{\varepsilon_{t+1}^*} dF(\varepsilon)\right], \tag{19}$$

which determines the endogenous cutoff value ε_t^* and, consequently, the optimal target inventory level in the model.[14] The left-hand side of equation (19) is the shadow value (opportunity cost) of holding inventory when the firm's productivity is high. The right-hand side is the expected rates of return by carrying one unit of inventory to the next period. In the case of low productivity ($\varepsilon_{t+1} < \varepsilon_{t+1}^*$), the firm opts to stockout ($s_{t+1}(i) = 0$) by keeping production low and the shadow value of inventory is $\frac{\Phi_{t+1}}{\varepsilon_{t+1}(i)}$. In the case of high productivity ($\varepsilon_{t+1} \geq \varepsilon_{t+1}^*$), the firm opts to carry the inventory forward and the shadow value is again $\frac{\Phi_{t+1}}{\varepsilon_{t+1}^*}$. Since the probability of stockout

[14] The probability of stockout in the model is given by $F(\varepsilon_t^*)$. Firms choose a target inventory and level of sales to determine the optimal probability of stockout under cost shocks.

is determined by the cutoff value ε_t^*, the firm chooses ε_t^* so that the marginal cost of holding inventory in period t equals the expected next-period marginal gains. In other words, equation (19) is the Euler equation for determining the inventory target by dynamic programming.

This Euler equation has an important implication: The cutoff ε_t^* is procyclical. Under a positive aggregate productivity shock, the cutoff ε_t^* increases because the real interest rate rises, suggesting that more firms choose to stockout to meet demand. This promotes final sales (i.e., consumption growth and capital accumulation) and reduces aggregate inventory investment along the extensive margin (i.e., the number of firms under Case B increases), resulting in a countercyclical aggregate inventory-to-sales ratio. This reflects a contemporaneous general-equilibrium trade-off but dynamic positive feedback between inventory investment and capital accumulation: Inventories facilitate sales, but producing inventories requires capital. Hence, after a positive aggregate productivity shock, it is optimal to expand production capacity by accumulating capital first and gradually build up inventory stock over time to meet final demand (sales).

The decision rule for production (equation (16)) states that production is larger than sales when the cost of production is below the cutoff value ($\varepsilon_t(i) \geq \varepsilon_t^*$), and it is less than sales when cost is high ($\varepsilon_t(i) < \varepsilon_t^*$). Such a decision rule confirms our earlier intuition that firms opt to bunch production and use inventories to smooth sales and maximize the average profits.

The decision rule for inventory accumulation (equation (17)) states that inventory investment, $s_t(i) - (1 - \delta_s)s_{t-1}(i)$, is procyclical with sales when $\varepsilon_t(i) \geq \varepsilon_t^*$,[15] suggesting that average (or aggregate) inventory investment is procyclical and production is more volatile than sales. This is consistent with the stylized empirical fact. However, despite the procyclical inventory investment, the aggregate inventory stock-to-sales ratio in the model is countercyclical. That is, a 1% increase in aggregate sales corresponds to less than a 1% increase in aggregate inventory stock. This is mainly the consequence of the contemporaneous general-equilibrium trade-off between capital accumulation and inventory investment, which makes the number of firms willing to accumulate inventories countercyclical under a procyclical real interest rate, thus dampening the initial rise in aggregate inventory investment on the extensive margin.

Bils and Kahn (2000) argue that the countercyclical inventory-to-sales ratio implies procyclical marginal costs and countercyclical markup. According to this argument, firms opt to increase inventories by less than sales in a boom because the marginal cost of production is procyclical. In our model, the aggregate marginal cost (Φ) is constant (to be shown shortly); yet, the model predicts a countercyclical aggregate inventory-to-sales ratio and procyclical inventory investment. This is not inconsistent with the argument of Bils and Kahn (2000) because the nonnegativity constraint on profits, equation (7), effectively imposes an upper bound on the capacity of firm-level

[15] On the other hand, its contemporaneous correlation with sales is zero when $\varepsilon < \varepsilon^*$.

production beyond which the marginal cost of firm-level production can be interpreted as infinity. However, this nonnegative-profit constraint alone does not lead to a countercyclical aggregate inventory-to-sales ratio in our model, because this constraint mainly affects the average steady-state inventory-to-sales ratio across firms and does not determine the dynamic responses of this ratio to aggregate TFP shocks. In fact, allowing negative profits may increase the steady-state inventory level more than it can increase steady-state sales; consequently, the percentage change (volatility) of aggregate inventory stock around its steady state relative to that of aggregate sales may be reduced (instead of increased), so the aggregate stock-to-sales ratio may become even more (instead of less) countercyclical.[16]

3 General Equilibrium

3.1 Aggregation

By the law of large numbers, the final output equation, $Z_t = \left[\int y_t(i)^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$, implies that the marginal cost is constant,

$$\Phi = \left(\frac{\sigma - 1}{\sigma}\right) \left[\int \varepsilon^{\sigma-1} dF(\varepsilon)\right]^{\frac{1}{\sigma-1}}, \tag{20}$$

where $F(\varepsilon)$ denotes the cumulative distribution function of the random variable ε. Define $Y \equiv \int_0^1 y(i)di$, $K \equiv \int_0^1 k(i)di$, $N \equiv \int_0^1 n(i)di$, $X \equiv \int_0^1 x(i)di$, and $S \equiv \int_0^1 s(i)$. Based on these definitions, the level of aggregate sales is given by

$$Y_t = PZ_t, \tag{21}$$

where the constant $P \equiv \left[\int \varepsilon(i)^{\sigma} dF(\varepsilon)\right] \left[\int \varepsilon(i)^{\sigma-1} dF(\varepsilon)\right]^{\frac{\sigma}{1-\sigma}}$ can be interpreted as an aggregate measure of the relative price of the final good in terms of intermediate goods. Note $P = 1$ if $\varepsilon(i)$ is constant across firms. Using the firm-level production decision rules, the level of aggregate production is given by

$$X_t = Y_t \frac{\left[\int_{\varepsilon<\varepsilon_t^*} \varepsilon^{\sigma} dF(\varepsilon) + \frac{\sigma}{\sigma-1} \int_{\varepsilon \geq \varepsilon_t^*} \varepsilon^{\sigma} dF(\varepsilon)\right]}{\int \varepsilon^{\sigma} dF(\varepsilon)} - (1 - \delta_s) S_{t-1} F(\varepsilon_t^*), \tag{22}$$

[16] See the robustness analysis in Section 5. Because firms are heterogeneous in our model, the effects of aggregate shocks and those of idiosyncratic shocks on inventory dynamics are different. In particular, idiosyncratic shocks affect only the average stock-to-sales ratio in the steady state, but not the dynamics of this ratio off the steady state. Such an important distinction does not exist in the representative-agent model of Bils and Kahn (2000). The key in generating a countercyclical aggregate stock-to-sales ratio in our model is the general-equilibrium trade-off between inventory investment and capital accumulation, which promotes final sales and reduces aggregate inventory investment along the extensive margin after aggregate productivity shocks.

and the aggregate stock of inventories is given by

$$S_t = \frac{1}{\sigma - 1} \frac{\int_{\varepsilon \geq \varepsilon_t^*} \varepsilon^\sigma dF(\varepsilon)}{\int \varepsilon^\sigma dF(\varepsilon)} Y_t + (1 - \delta_s) S_{t-1} \left[1 - F(\varepsilon_t^*)\right]. \tag{23}$$

Equation (23) resembles the familiar stock-adjustment model of inventories widely used in the empirical literature (see, e.g., Blinder 1986b). However, the crucial difference here is that the speed of adjustment in our model is time varying and depends in general equilibrium on other aggregate variables through the optimal cutoff ε_t^*.[17] According to this stock-adjustment equation, the optimal inventory stock is a weighted cumulative sum of current and past sales with the weights depending negatively on the time-varying cutoff ε_t^*; hence, it is smoother than sales (since the cutoff is procyclical) and implies a countercyclical aggregate stock-to-sales ratio. Such a prediction is consistent with the empirical fact emphasized by Bils and Kahn (2000).[18]

It is easy to check that these aggregate relations (22) and (23) satisfy the aggregate resource identity,

$$Y_t + S_t = (1 - \delta_s) S_{t-1} + X_t. \tag{24}$$

The factor demand functions imply $(r_t + \delta_k) K_t = \alpha \Phi M_t$ and $w_t N_t = (1 - \alpha) \Phi M_t$, where M_t is defined as

$$M_t \equiv \int_0^1 \frac{x_t(i)}{\varepsilon_t(i)} di. \tag{25}$$

Since $\Phi \equiv \frac{1}{A_t} \left(\frac{r_t + \delta_k}{\alpha}\right)^\alpha \left(\frac{w_t}{1-\alpha}\right)^{1-\alpha}$, these aggregate factor demand functions imply the aggregate production function,

$$M_t = A_t K_t^\alpha N_t^{1-\alpha}. \tag{26}$$

Using the definition of M_t and the firm-level production decision rule, we also have

$$M_t = Y_t \frac{\left[\int_{\varepsilon < \varepsilon_t^*} \varepsilon^{\sigma-1} dF(\varepsilon) + \frac{\sigma}{\sigma-1} \int_{\varepsilon \geq \varepsilon_t^*} \varepsilon^{\sigma-1} dF(\varepsilon)\right]}{\int \varepsilon^\sigma dF(\varepsilon)} - (1 - \delta_s) S_{t-1} \int_{\varepsilon < \varepsilon_t^*} \varepsilon^{-1} dF(\varepsilon). \tag{27}$$

Notice that when $\varepsilon_t(i)$ is constant across firms, we have $M_t = \frac{X_t}{\varepsilon}$ by comparing equations (27) and (22).

[17] Schuh (1996) shows that accounting for time variation in the inventory adjustment speed improves the fit of a traditional inventory model.

[18] A similar stock-adjustment equation also arises in the models of KT (2007a) and Wen (2008). Although borrowing constraints play an indirect role in giving rise to the stock-adjustment equation in our model, they are not the key for generating a countercyclical aggregate inventory-to-sales ratio. In fact, it is the *procyclical* movements in the cutoff that hold the key for the countercyclical stock-to-sales ratio in our model. However, borrowing constraints do play an indirect role in supporting the inventory-accelerator mechanism in our model because the general-equilibrium trade-off between capital and inventories would break down if firms could produce infinite output by borrowing infinite amount of capital from households. See Section 5 for the robustness analysis on relaxing borrowing constraints.

3.2 Household

To close the model, we add a representative household. The household's role is to supply labor and accumulate capital. The household derives utility from consumption (C_t) and disutility of working (N_t) by solving

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \left(\log C_t - a \frac{N_t^{1+\gamma}}{1+\gamma} \right) \tag{28}$$

subject to

$$C_t + K_{t+1} \leq (1 + r_t) K_t + w_t N_t + \Pi_t, \tag{29}$$

where Π denotes aggregate profit income distributed from firms. Denoting Λ_t as the Lagrangian multiplier for the budget constraint, the first-order conditions for $\{C_t, N_t, K_{t+1}\}$ are given, respectively, by

$$C_t^{-1} = \Lambda_t \tag{30}$$

$$a N_t^{\gamma} = w_t \Lambda_t \tag{31}$$

$$\Lambda_t = \beta E_t (1 + r_{t+1}) \Lambda_{t+1}. \tag{32}$$

The aggregate profit income is given by

$$\Pi_t = \int_0^1 \left[y_t(i)^{1-\frac{1}{\sigma}} Z_t^{\frac{1}{\sigma}} - \frac{\Phi}{\varepsilon_t(i)} x_t(i) \right] di = Z_t - w_t N_t - (r_t + \delta_k) K_t, \tag{33}$$

where $Z_t = \frac{Y_t}{P}$ is the aggregate final output given in equations (1) and (21). This implies that in equilibrium the household's budget constraint is given by

$$C_t + K_{t+1} - (1 - \delta_k) K_t = Z_t, \tag{34}$$

which is also the equilibrium market-clearing condition for the final good.

The aggregate variables that need to be determined are $\{C_t, K_{t+1}, N_t, X_t, M_t, Y_t, S_t, \varepsilon_t^*\}$. Define the functions $g_1(\varepsilon^*) \equiv \int_{\varepsilon<\varepsilon^*} \varepsilon^{\sigma-1} dF(\varepsilon)$, $g_2(\varepsilon^*) \equiv \int_{\varepsilon \geq \varepsilon^*} \varepsilon^{\sigma-1} dF(\varepsilon)$, $h(\varepsilon^*) \equiv \int_{\varepsilon \geq \varepsilon^*} \varepsilon^{\sigma} dF(\varepsilon)$, $Q(\varepsilon^*) \equiv \int_{\varepsilon<\varepsilon^*} \varepsilon^{-1} dF(\varepsilon)$, and $\tilde{\varepsilon} \equiv \int \varepsilon^{\sigma} dF(\varepsilon)$. The system of equations that solve these variables is given by

$$\frac{1}{\varepsilon_t^*} = \beta (1 - \delta_s) E_t \frac{C_t}{C_{t+1}} \left[Q(\varepsilon_{t+1}^*) + \frac{1 - F(\varepsilon_{t+1}^*)}{\varepsilon_{t+1}^*} \right] \tag{35}$$

$$M_t = Y_t \frac{\left[g_1(\varepsilon_t^*) + \frac{\sigma}{\sigma-1} g_2(\varepsilon_t^*) \right]}{\tilde{\varepsilon}} - (1 - \delta_s) Q(\varepsilon_t^*) S_{t-1} \tag{36}$$

$$S_t = \frac{1}{\sigma - 1} Y_t \frac{h(\varepsilon_t^*)}{\tilde{\varepsilon}} + (1 - \delta_s) S_{t-1} \left[1 - F(\varepsilon_t^*)\right] \tag{37}$$

$$Y_t + S_t = (1 - \delta_s) S_{t-1} + X_t \tag{38}$$

$$a N_t^{1+\gamma} = (1 - \alpha) \Phi \frac{M_t}{C_t} \tag{39}$$

$$1 = \beta E_t \left(\alpha \Phi \frac{M_{t+1}}{K_{t+1}} + 1 - \delta_k \right) \frac{C_t}{C_{t+1}} \tag{40}$$

$$C_t + K_{t+1} - (1 - \delta_k) K_t = \frac{1}{P} Y_t \tag{41}$$

$$M_t = A_t K_t^{\alpha} N_t^{1-\alpha}, \tag{42}$$

where equation (35) is the Euler equation for the optimal cutoff ε_t^*, equation (36) characterizes aggregate production decisions as a function of aggregate sales and existing inventory stock, equation (37) is the equilibrium law of motion for inventory accumulation (i.e., the optimal stock adjustment equation for inventories), equation (38) is the aggregate resource identity for firms' output (that relates total sales and inventory investment to production), equation (39) represents the household's optimal labor supply, equation (40) the household's optimal consumption, equation (41) the household's budget identity, and equation (42) the aggregate production function. Note equations (38) and (41) imply the following final-goods resource constraint that links the final demand to intermediate goods production and inventories,

$$C_t + K_{t+1} - (1 - \delta_k) K_t + \frac{1}{P} \left[S_t - (1 - \delta_s) S_{t-1}\right] = \frac{1}{P} X_t, \tag{43}$$

where $P \equiv \frac{Y_t}{Z_t}$ is the relative price of the final goods. Equation (37) implies the steady-state inventory-to-sales ratio,

$$\frac{S}{Y} = \frac{1}{1 - (1 - \delta_s)(1 - F)} \frac{1}{\sigma - 1} \frac{h(\varepsilon^*)}{\tilde{\varepsilon}}. \tag{44}$$

4 Impulse Responses

4.1 Definition of GDP

Because of imperfect competition and inventories, GDP must be measured carefully in the model economy. By the value-added approach, GDP in the model economy is the sum of the final-good sector's value added and the intermediate-goods sector's value added. The final-good sector's value added is its revenue minus the costs of intermediate goods used in production, $Z_t - \int_0^1 p_t(i) y_t(i) di$, which is zero under perfect competition; whereas the intermediate-goods sector's value added is $\int_0^1 p_t(i)x_t(i)di$. Hence, $GDP_t = \int_0^1 p_t(i)x_t(i)di$. Alternatively, by the expenditure approach, GDP is the sum of Z (aggregate consumption plus business investment) and inventory investment (by market values), $GDP_t = Z_t + \int_0^1 p_t(i)[s_t(i) - (1-\delta_s)s_{t-1}(i)]di$. Since inventory investment equals production minus sales, by the zero-profit condition we have $GDP_t = \int_0^1 p_t(i)x_t(i)di$. Hence, both approaches give the same results. Substituting the relationship $p_t(i) = \frac{\sigma}{\sigma-1}\frac{\Phi_t}{\varepsilon_t(i)}$ into GDP gives

$$GDP_t = \frac{\sigma}{\sigma-1}\Phi\int_0^1 \frac{x_t(i)}{\varepsilon_t(i)}di = \frac{\sigma}{\sigma-1}\Phi M_t, \tag{45}$$

where M_t is the aggregate supply of intermediate goods defined in equation (26). That is, GDP is proportional to aggregate production of intermediate goods. As the markup disappears ($\sigma \to \infty$), we have $\Phi = 1$ and $GDP_t = M_t$.

4.2 A Benchmark Calibration

Assume $\varepsilon(i)$ is drawn from a Pareto distribution, $F(\varepsilon) = 1 - \left(\frac{1}{\varepsilon}\right)^\eta$, with the shape parameter $\eta > 0$ and the support $\varepsilon \in (1, \infty]$. We assume Pareto distribution because an extensive empirical literature documents that many characteristics of firms (such as firm size, productivity, employment, sales, R&D expenditures) follow Pareto distributions.[19] Such empirical facts have motivated much of the heterogeneous-firm literature to assume Pareto distributions for firms' idiosyncratic productivity shocks (see, e.g., Helpman, Melitz, and Yeaple, 2004). However, we examine the robustness of our results to this assumption in Section 5 where alternative distributions, such as lognormal and uniform distributions, are considered.

The functions defined before equation (35) are not integrable under Pareto distribution if $\eta \le \sigma$, so we further assume $\eta > \sigma$ to make these integrations meaningful. With this assumption, the functions $\{g_1(\varepsilon^*),\ g_2(\varepsilon^*),\ h_1(\varepsilon^*),\ h_2(\varepsilon^*),\ q_1(\varepsilon^*),\ q_2(\varepsilon^*)\}$ are given by $g_1(\varepsilon^*) = \frac{\eta}{\eta+1-\sigma}\left[1 - (\varepsilon^*)^{\sigma-\eta-1}\right]$, $g_2(\varepsilon^*) = \frac{\eta}{\eta+1-\sigma}(\varepsilon^*)^{\sigma-\eta-1}$, $h(\varepsilon^*) = \frac{\eta}{\eta-\sigma}(\varepsilon^*)^{\sigma-\eta}$, $q_1(\varepsilon^*) = \frac{\eta}{\eta+1}\left[1-(\varepsilon^*)^{-1-\eta}\right]$, and $q_2(\varepsilon^*) = \frac{\eta}{\eta-1}\varepsilon^{*1-\eta}$. Similarly, we have $\tilde{\varepsilon} = \frac{\eta}{\eta-\sigma}$, the relative price $P = \left[\frac{\eta}{\eta-\sigma}\right]\left[\frac{\eta}{\eta+1-\sigma}\right]^{\frac{\sigma}{1-\sigma}}$, and the aggregate marginal cost $\Phi = \left(\frac{\sigma-1}{\sigma}\right)\left[\frac{\eta}{\eta+1-\sigma}\right]^{\frac{1}{\sigma-1}}$.

The time period is one quarter. We set $\beta = 0.99$, $\alpha = 0.35$, $\delta_k = 0.035$, $\delta_s = 0.005$, $\sigma = 10$ (implying a 10% markup), and $\gamma = 0$ (indivisible labor).[20] We set a in the utility function to imply

[19] See, e.g., Axtell (2001) and Luttmer (2007) and the references therein.
[20] Our main results do not hinge on this particular assumption of the labor supply function.

steady-state hours worked $\bar{N} = 0.24$ (i.e., 40 hours per week).[21] The aggregate technology shock is assumed to follow an $AR(1)$ process in log, $\hat{A}_t = \rho \hat{A}_{t-1} + \nu_t$, with persistence $\rho = 0.9$. The standard deviation of v_t is set to $\sigma_v = 0.007$ (Kehoe and Perri, 2002), which is consistent with most estimations of the Solow residual based on Cobb-Douglas production function (e.g., King and Rebelo, 1999). Accoring to King and Rebelo (1999), a value of $\sigma_v = 0.007$ implies that standard RBC models are not able to explain the variations in GDP. However, our model with inventories is able to match the variance of U.S. GDP because inventories can amplify the business cycle. In particular, by properly choosing the value of η, which measures the variance of firms' idiosyncratic cost shocks and determines the strength of incentives for firms to hold inventories, our model can exactly match the standard deviation of GDP. For this reason, we pick $\eta = 10.13$ so that the predicted standard deviation of GDP in the model roughly matches $\sigma_{gdp} = 0.024$ for the U.S. economy. The key parameter values of the model are summarized in Table 1.

Table 1. Parameter Values

α	β	δ_k	δ_s	γ	σ	η	ρ	σ_v
0.35	0.99	0.035	0.005	0	10	10.13	0.9	0.007

We apply the log-linearization method to study the model's aggregate dynamics around the steady state. The impulse responses of GDP, consumption, labor, investment, inventory investment, and the inventory-to-sales ratio ($\frac{S_t}{Y_t}$) to a 1 standard deviation (SD) aggregate technology shock are graphed in Figure 1 (thick solid lines). The horizontal axis represents time (i.e., the number of quarters after the shock), and the vertical axis represents percentage changes of each variable relative to its steady state.

Figure 1 shows that, under aggregate technology shocks, aggregate inventory investment (panel E) is strongly procyclical and far more volatile than GDP (panel A); at the peak of the responses, inventory investment is more than 10 times as volatile as GDP. However, the inventory stock-to-sales ratio (panel F) is countercyclical, suggesting that the stock fails to track sales one-for-one despite the strongly procyclical changes in inventory investment. Hence, the model is able to explain the key stylized facts of inventory behavior emphasized by the empirical literature (e.g., Bils and Kahn, 2000). In addition, consumption (panel B) is less volatile and investment (panel C) is more volatile than GDP, as in a standard RBC model.

[21] This parameter has no effects on the dynamics of the model around the steady state.

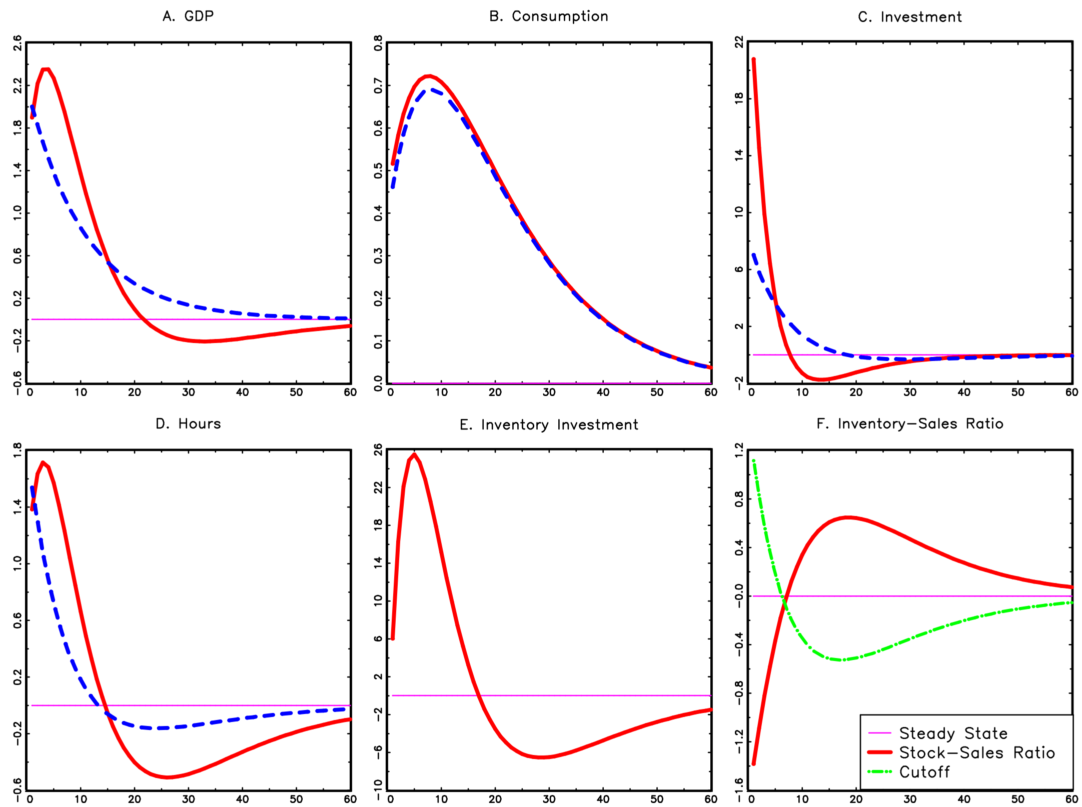

Figure 1. Impulse Responses to a 1% TFP Shock

Most notable in Figure 1 are the hump-shaped impulse responses of labor (panel D) and GDP to a technology shock. This dynamic pattern of aggregate employment and output is a defining feature of the business cycle emphasized in the literature as a litmus test for quantitative business-cycle models (see, e.g., Cogley and Nason, 1995). The fact that our inventory model is able to generate such hump-shaped output dynamics is striking. The general-equilibrium inventory models of KT (2007a) and Wen (2008) do not deliver this result. However, as the variance of the idiosyncratic shocks decreases (or the value of η increases), the hump-shaped output dynamics are weakened rapidly. This is because as η increases, the distribution of idiosyncratic shocks becomes more and more degenerate so the incentives for holding inventories diminish quickly. In the limit ($\eta \to \infty$), inventories completely disappear and the model is reduced to a standard Dixit-Stiglitz RBC model without hump-shaped dynamics.

As a comparison, the thick dashed lines in Figure 1 (panels A, B, C, D) show the responses of our control model, which has no inventories. The control model is a version of our inventory model with $\varepsilon(i) = 1$, which is also equivalent to a limiting version of the inventory model with $\eta \to \infty$. The figure shows that the control model is significantly less volatile than the inventory model in terms of GDP (panel A), capital investment (panel C), and employment (panel D). Under our benchmark calibrations, aggregate output is about 40% more volatile with inventories than without;

similarly, aggregate employment and aggregate investment are also significantly more volatile with inventories than without. However, aggregate consumption is only slightly more volatile with inventories than without.[22] As the variance of idiosyncratic shocks decreases, the model's impulse responses gradually converge to those of the control model.

Table 2. Selected Inventory Statistics*

	Mean ($\frac{S}{Y}$)	Second Moments	$x = \Delta S_t$	$x = \frac{S_t}{Y_t}$
		σ_x/σ_{output}	17.8 (16.7)	1.19 (0.86)
U.S. Data	0.55 (0.61)	$cor(x, output)$	0.55 (0.61)	-0.50 (-0.48)
		$cor(x, sales)$	0.57 (0.41)	-0.59 (-0.51)
		σ_x/σ_{output}	11.0	0.55
Model A	0.24	$cor(x_t, output)$	0.91	-0.75
		$cor(x_t, sales)$	0.60	-0.89
		σ_x/σ_{output}	10.80	0.61
Model B	0.53	$cor(x_t, output)$	0.88	-0.79
		$cor(x_t, sales)$	0.62	-0.92

*U.S. data (1958:1-2000:4): Total manufacturing inventory stock (S), sales (Y), and output ($\Delta S + Y$). Numbers in parentheses are NIPA data for aggregate inventory stock, sales, and GDP. Model A's calibration: $\sigma = 10, \eta = 10.13$. Model B's calibration: $\sigma = 6, \eta = 6.25$.

Table 2 reports some selected population moments of inventory dynamics implied by the benchmark model (Model A) and the U.S. economy.[23] We also report the results for the case where $\sigma = 6$ (implying a markup of 20%) in the lower panel of Table 2 (Model B).[24] With $\sigma = 6$, we need to recalibrate η (= 6.25) to generate a sufficiently volatile GDP that matches the data. The impulse responses of Model B remain hump-shaped and look similar to those of Model A. Notice that the benchmark model with $\sigma = 10$ understates the inventory-to-sales ratio substantially. However, with 20% markup (Model B), the model can match the data quite well.[25] Also note that the benchmark

[22] With inventories, consumption can be either slightly more volatile or less volatile, depending on the elasticity of labor supply. However, the hump-shaped dynamics are not very sensitive to the elasticity of labor supply. That is, we can also obtain hump-shaped impulse responses with log function of leisure.

[23] The average inventory-to-sales ratio (S/Y) for the U.S. economy is based on the manufacturing sector's shipment (Y) and total inventory stock (S) (including finished goods, work in progress, and raw materials) for the period 1958:1-2000:4. This value is 0.55, which is consistent with that reported by Kahn, McConnell, and Perez-Quiros (2002) based on real nonfarm inventories and final sales of goods (including both durables and nondurables). The other U.S. statistics are based on NIPA data, where sales (Y) is defined as GDP minus inventory investment (ΔS). The inventory stock is measured as the cumulative sum of inventory changes with the initial value equals $0.65 \times GDP_{1958:1}$, so that the estimated inventory stock shares a balanced growth rate with GDP. The resulting stock-to-sales ratio is 0.61, slightly higher than that of the manufacturing sector but lower than the figure of 0.72 reported by KT (2007a). The second moments of the data are estimated based on HP-filtered data. To apply the HP filter to inventory changes (ΔS_t), we follow the method in Wen (2008) by obtaining a HP-filtered inventory stock series first, and then derive inventory changes that are consistent with the log-linearization method and the rate of depreciation assumed in the model.

[24] The value of σ is related to the degree of firms' monopoly power. Estimates of markups typically fall in the 10-20% percent range, implying values of σ in the 6-10 range (see, e.g., Rotemberg and Woodford, 1995; and Basu and Fernald, 1997).

[25] To generate a high enough inventory-to-sales ratio, the model requires η to be small, but the value of η is bounded

model can generate very volatile and procyclical inventory investment, with a volatility more than 10 times that of GDP and a correlation with GDP about 0.9. Despite this strongly procyclical inventory investment, the inventory-to-sales ratio is countercyclical. Its correlation with GDP is -0.75 and its correlation with sales is about -0.9. The results are similar for Model B. These predictions are qualitatively consistent with the data.[26]

Equation (16) suggests that production at the firm level rises with sales for both low- and high-cost firms, and equation (17) suggests that inventory investment rises with sales for low-cost firms (i.e., firms with $\varepsilon(i) \geq \varepsilon_t^*$); hence, average inventory investment, sales, and production across firms comove with each other. This implies that, under aggregate productivity shocks, (i) aggregate inventory investment is procyclical and (ii) aggregate production is more variable than sales.

Equation (35) is the key equation to understanding the countercyclical inventory-to-sales ratio and hump-shaped dynamics. This equation implies that the cutoff ε_t^* comoves with consumption growth (or the real interest rate). Under a positive, persistent aggregate productivity shock, aggregate consumption rises gradually with a hump, implying that the cutoff increases in the initial periods of the boom and declines afterward (because of negative consumption growth). That is, the cutoff (ε_t^*) is procyclical (see the dot-dashed lines in panel F of Figure 1).[27] Because of the sharp rise in the cutoff on impact, aggregate inventory investment does not rise as rapidly as aggregate sales initially because fewer firms make inventory investments when the cutoff is high, leading to a countercyclical aggregate inventory-to-sales ratio. Since equation (35) is derived independently from the nonnegative profit constraint, borrowing constraints are thus not the cause for the countercyclical inventory-to-sales ratio in the model. In fact, as we show in Section 5, relaxing the borrowing constraint in equation (7) may lead to an even more countercyclical stock-to-sales ratio.[28]

A higher cutoff in the initial periods of the boom implies that more firms opt to stockout, which strengthens final sales and enables more rapid capital accumulation by households than the case without inventories.[29] This explains the dramatic increases in capital investment in the initial

below by σ. However, Section 5 shows that the steady-state stock-to-sales ratio can be significantly increased without lowering the value of σ if the borrowing constraint (equation (7)) is relaxed or if we assume different distributions for $\varepsilon_t(i)$.

[26] However, the simplicity of our benchmark model limits its quantitative fit in several respects. Most notably, even though inventory investment is 10 times more volatile than that of GDP in the model, it still understates the volatility of inventory investment and the volatility of inventory-to-sales ratio.

[27] The dot-dashed line is rescaled by multiplying $\eta = 10.13$ to the original value. The number of inventory-accumulating firms in our model is given by $1 - F(\varepsilon_t^*)$. With Pareto distribution, we have $1 - F(\varepsilon_t^*) = (\varepsilon_t^*)^{-\eta}$. Hence, a 1% increase in the cutoff leads to -10.13% decline in the number of inventory-investing firms. Thus, the rescaled dot-dashed line captures the actual impact of the cutoff on aggregate inventory stock.

[28] This seemingly counterintuitive result occurs because borrowing constraints only restrict the inventory stock at the firm level but do not determine the changes in the distribution of inventory-holding firms under aggregate productivity shocks. In other words, borrowing constraints determine mainly the steady-state aggregate inventory-to-sales ratio instead of the dynamic impulse response of the ratio.

[29] Under a positive aggregate technology shock, the marginal products of capital and labor both increase, giving rise to the initial boom in the economy. This is true with or without inventories. However, inventories facilitate sales—not only will low-cost firms want to produce and sell more intermediate goods to the final-good sector, but high-cost firms can also sell more because of previously accumulated inventories. Thus, aggregate investment increases more than it

phase of the boom in the inventory model. This, in turn, raises the future marginal product and demand of labor and allows firms to produce more inventories with greater production capacity (and more relaxed borrowing constraints) in the subsequent periods. This results in a cumulative process of expansion. Sooner or later the technology shock will be exhausted so that the engine of inventory accumulation loses steam. Once the economy starts to contract, sales fall and firms opt to decumulate inventories so that the relative prices of intermediate goods rise to prevent revenue from declining sharply. In particular, because intermediate-goods prices are much higher when firms deplete inventories, imperfectly competitive firms opt to reduce production faster than they would in the control model. This causes the economy to overshoot its steady state from above, giving rise to a boom-bust–like propagation mechanism.

More specifically, the movements in aggregate inventories in our model come from two margins — an intensive margin determined by individual firms' inventory investment and an extensive margin determined by the fraction of firms making inventory investment. Idiosyncratic cost shocks (in conjunction with a concave profit function in sales) provide the incentive for individual firms to hold inventories to smooth sales, so low-cost firms opt to accumulate inventories. However, holding inventories is costly because of depreciation and positive real interest rates. There is thus a contemporaneous trade-off between holding inventories and accumulating capital, which implies that a positive aggregate productivity shock will change the equilibrium distribution of firms so that there are fewer "low-cost" firms that opt to accumulate inventories (but with a higher intensity) and more "high-cost" firms that opt to stockout. This adjustment along the extensive margin strengthens final sales but dampens the initial rise in aggregate inventory investment. Consequently, capital investment can rise sharply and inventory investment increases only slowly with a lag. However, since the purpose of accumulating capital is to produce more inventories to satisfy future sales, inventory investment must eventually accelerate as the real interest rate declines (and the borrowing constraint relaxes) over time. Associated with this acceleration is the sharp decrease of the cutoff below the steady state in the middle of the boom (see the dot-dashed line in panel F of Figure 1), indicating significant increases in the fraction of firms making inventory investment as the capital stock (production capacity) peaks. Hence, the rapid accumulation of capital in the initial periods facilitates more inventory investment in the subsequent periods, which further enhances sales and capital accumulation in the future. This dynamic (intertemporal) reinforcement between capital and inventory accumulations (with the help of endogenous labor supply) gives rise to an inventory-accelerator mechanism and the hump-shaped dynamics in the model.[30]

would if inventories did not exist.

[30] In a counterfactual experiment, we note that once capital adjustment costs are introduced so that the contemporaneous tradeoff and intertemporal reinforcement between capital and inventories are dampened, the hump-shaped output dynamics then disappear and the volatility of GDP decreases significantly. However, the variance of GDP is still larger with inventories than without because the sales-smoothing mechanism is still working. This suggests that borrowing constraints do not play a direct key role in giving rise to the hump-shaped dynamics (because capital

To quantify the destabilizing effects of inventories on GDP, Table 3 reports the business-cycle statistics implied by the inventory model, the control model, and the models of KT (2007a) (including their control model). In the table, C denotes aggregate consumption, I capital investment, and N total hours.[31] We normalize the SD of GDP to 1 in both models. Panel A in the table reports SDs relative to each model's own GDP, and panel B reports SDs relative to the counterpart variables in the corresponding control model. The most informative information is in panel B. The table shows that GDP in our model is 39% more volatile with inventories than without. This value is even higher if the statistics are based on HP-filtered samples (numbers in brackets). In contrast, GDP in the KT model is only slightly (1.02 times) more volatile with inventories than without. The main reason for this sharp difference between the two models is that the components of aggregate sales (especially investment) in our model are far more volatile with inventories than without, whereas those in the KT model are significantly less volatile with inventories than without. For example, capital investment in our model is 2.24 times more volatile with inventories, but that in the KT model is 14% less volatile with inventories. This is because inventories promote and reinforce sales in our model while they dampen sales in the KT model.

Table 3. Business Cycles with and without Inventories*

	GDP	*C*	*I*	*N*
A: Standard deviations relative to GDP				
Our model	1.0	0.47 [0.24]	4.28 [6.84]	0.73 [0.78]
KT model	1.0	0.35	6.32	0.72
B: Standard deviations relative to control				
Our model	1.39 [1.50]	1. 03 [1.18]	2.24 [3.01]	1.76 [1.52]
KT model	1.02	0.92	0.86	1. 1

*Numbers are exact population moments, and numbers in brackets (and the KT model) are based on HP-filtered samples (with smoothing parameter 1600 and sample size 5000).

That inventory investment promotes and reinforces sales in our model can be seen from equation (37), which can be rearranged to

$$Y_t = \frac{\tilde{\varepsilon}\,(\sigma-1)}{h(\varepsilon_t^*)}\left[S_t - (1-\delta_s)S_{t-1}\right] + \frac{\tilde{\varepsilon}\,(\sigma-1)}{h(\varepsilon_t^*)}(1-\delta_s)F(\varepsilon_t^*)S_{t-1}, \tag{46}$$

where $h(\varepsilon_t^*) = \frac{\eta}{\eta-\sigma}\left(\varepsilon_t^*\right)^{\sigma-\eta}$ under Pareto distribution. This equation shows that aggregate sales depend positively on current aggregate inventory investment, $S_t - (1-\delta_s)\,S_{t-1}$, and the lagged

adjustment costs do not directly interfere with borrowing constraints). Nonetheless, borrowing constraints may have indirectly facilitated this propagation mechanism because they may strengthen firms' strategic motives for capital accumulation to help relax the capacity constraints in subsequent periods.

[31] The statistics for our model are exact population moments. We also report in the brackets statistics based on HP-filtered samples (with the smoothing parameter $\lambda = 1600$ and sample size 5000).

inventory stock of high-cost firms, $F(\varepsilon_t^*)S_{t-1}$. These positive relationships are further amplified when the cutoff is procyclical because $\frac{\partial h}{\partial \varepsilon^*} < 0$ and $\frac{\partial F}{\partial \varepsilon^*} > 0$. This means that an increase in the fraction of high-cost firms (which opt to stockout because of a high real interest rate) strengthens sales. Consequently, the variance of aggregate production in our model is significantly amplified by inventories, whereas it is essentially unaffected by inventories in the KT model because in that model the reduction in the volatility of sales offsets the potentially positive impact of procyclical inventory changes on GDP.

5 Extensions and Robustness Analysis

5.1 Relaxing Borrowing Constraints

To prevent infinite production and unlimited inventory accumulation under the production-cost–smoothing motive, the benchmark model assumes that firms cannot pay negative dividends to households. Here we relax this assumption by allowing firms to incur negative profits, so constraint (7) becomes

$$p_t(i)y_t(i) - \frac{\Phi_t}{\varepsilon_t(i)}x_t(i) \geq -b_0, \tag{47}$$

where the parameter $b_0 \geq 0$ measures the extent of firms' external financing by households. With this modification, a firm's decision rules for optimal production and inventory investment become

$$x_t(i) = \begin{cases} \frac{\sigma}{\sigma-1}y_t(i) + \frac{\varepsilon_t(i)}{\Phi_t}b_0 & \text{if } \varepsilon_t(i) \geq \varepsilon_t^* \\ y_t(i) - (1-\delta_s)s_{t-1}(i) & \text{if } \varepsilon_t(i) < \varepsilon_t^* \end{cases} \tag{48}$$

$$s_t(i) = \begin{cases} \frac{1}{\sigma-1}y_t(i) + \frac{\varepsilon_t(i)}{\Phi_t}b_0 + (1-\delta_s)s_{t-1}(i) & \text{if } \varepsilon_t(i) \geq \varepsilon_t^* \\ 0 & \text{if } \varepsilon_t(i) < \varepsilon_t^* \end{cases}. \tag{49}$$

Clearly, the possibility of external financing ($b_0 > 0$) allows a firm to produce and accumulate more inventories in the case of low idiosyncratic marginal cost ($\varepsilon \geq \varepsilon^*$). Consequently, the stock-adjustment equation (23) will have an additional positive term, $\frac{b_0}{\Phi}\int_{\varepsilon \geq \varepsilon_t^*} \varepsilon dF(\varepsilon)$, on its right-hand side and the aggregate stock-to-sales ratio in equation (44) becomes

$$\frac{S}{Y} = \frac{1}{1-(1-\delta_s)(1-F)}\left[\frac{1}{\sigma-1}\frac{h(\varepsilon^*)}{\int \varepsilon^\sigma dF(\varepsilon)} + \frac{b}{\Phi}\int_{\varepsilon \geq \varepsilon^*} \varepsilon dF(\varepsilon)\right], \tag{50}$$

where $b \equiv \frac{b_0}{Y}$ denotes a firm's external financing limit-to-sales ratio in the steady state. Clearly, the aggregate inventory-to-sales ratio increases with the borrowing limit (b) in the steady state.

Table 4 confirms that allowing for negative profits increases the steady-state inventory-to-sales ratio.[32] For example, when $\sigma = 10$, $\eta = 10.15$, and $b = 0.30$, the implied stock-to-sales ratio is 0.61, which is 10% larger than the average stock-to-sales ratio of the manufacturing sector but exactly the same as the aggregate inventory-to-sales ratio for the U.S. economy (see Table 2). More importantly, relaxing borrowing constraints does not necessarily make the inventory-to-sales ratio less countercyclical. On the contrary, it may make the ratio more countercyclical. For example, when the value of b increases from 0 to 0.10, the inventory-to-sales ratio increases from 0.24 to 0.36, and its correlation with GDP drops from -0.22 to -0.41. The main intuition is that allowing negative profits may increase the steady-state inventory level more than it can increase steady-state sales; consequently, the volatility of inventory stock relative to that of sales may be reduced instead of increased, leading to a more countercyclical stock-sales ratio. In addition, although firms with $\varepsilon(i) > \varepsilon^*$ increase inventory investment, the number of inventory-investing firms declines at the same time because ε^* increases, dampening the rise in aggregate inventories on the extensive margin. Hence, the assumption of nonnegative dividends in the benchmark model does not play the same role as the assumption of increasing marginal costs does in Bils and Kahn (2000), although it is consistent with increasing marginal costs.[33]

Table 4. Effects of Debt on Inventory Dynamics ($\sigma = 10, \eta = 10.15$)

$\frac{b_0}{Y}$ Ratio (b)	Inventory-to-Sales Ratio	Correlation with GDP	
		ΔS_t	$\frac{S_t}{Y_t}$
0	0.24	0.87	-0.22
0.05	0.30	0.84	-0.34
0.10	0.36	0.81	-0.41
0.15	0.43	0.80	-0.42
0.30	0.61	0.79	-0.40

Because inventories magnify and propagate the impact of aggregate shocks in our model through an inventory-accelerator mechanism, allowing $b > 0$ will strengthen the model's propagation mechanism and make the model more volatile. The left panel in Figure 2 shows that, as the value of b increases, the variance of GDP rises accordingly because output becomes more hump-shaped (we defer discussion of the middle and right panels until the next subsection).[34] The intuition is that the ability to borrow raises the level of inventories across firms, thus allowing firms to increase sales more than they could previously; which strengthens capital accumulation in the initial periods of

[32] The statistics reported in Table 4 are exact population moments implied by our model. In generating Table 4, we have relaxed the value of η from 10.13 to 10.15 but have kept the other structural parameters in Table 1 unchanged. This allows us to examine a wider range of the possible values of b without overstating the stock-to-sales ratio.

[33] Bils and Kahn (2000) cast their arguments in a representative-agent framework where there is no distinction between aggregate shocks and idiosyncratic shocks.

[34] The value of η assumed in generating the graphs in the left window in Figure 2 is $\eta = 10.15$, consistent with that in Table 4.

a boom and reinforces the inventory-accelerator mechanism of the model.[35]

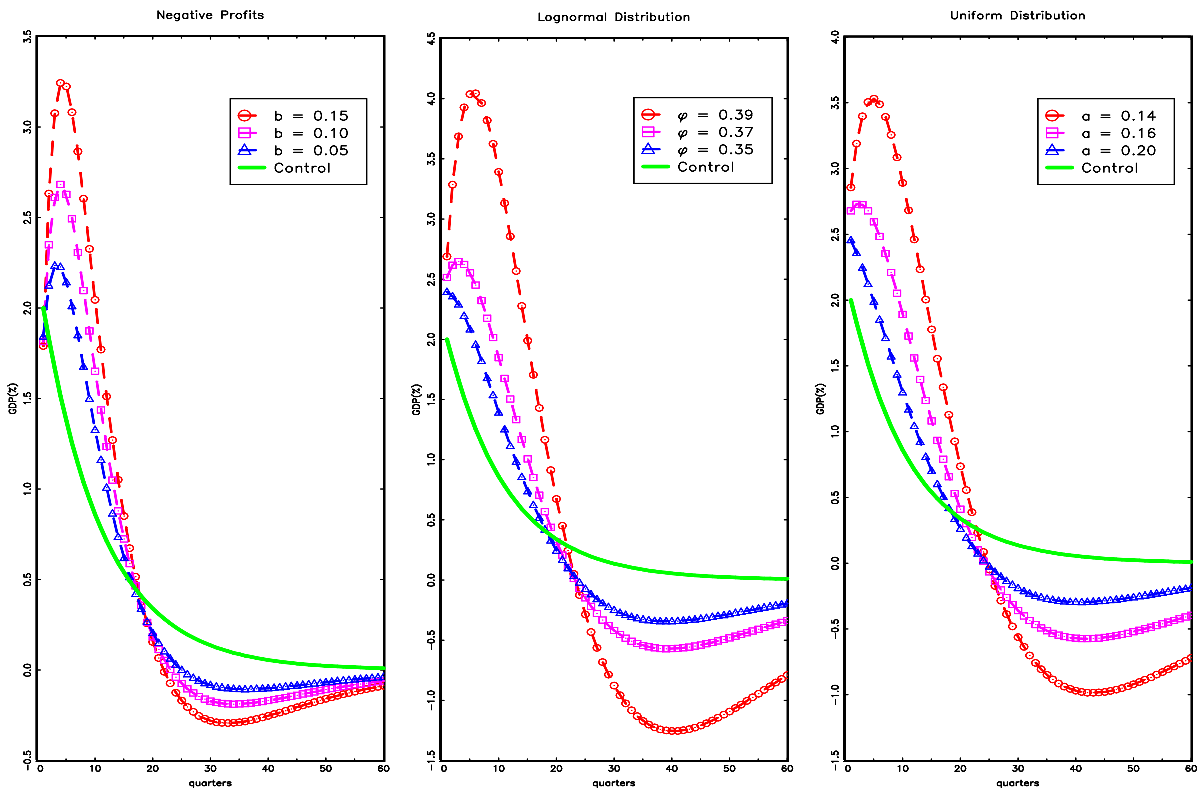

Figure 2. Impulse Responses of GDP to a 1% TFP Shock

5.2 Alternative Distributions

In the previous analyses, we assumed Pareto distribution for idiosyncratic productivity shocks. Pareto distribution has the property that it is long-tailed and fits well the empirical pattern of the distribution of firm size and productivity. This long-tail property has important implications for firms' optimal inventory behaviors, especially for the cutoff variable, ε_t^*. Because inventories are held to facilitate future sales when the marginal costs may be high in the next period, firms have less incentive to hold inventories to smooth sales knowing that it is always likely to draw a very low cost shock (or very high productivity shock) tomorrow. They thus opt to choose a higher cutoff so that in equilibrium the number of firms making positive inventory investment is relatively low in the steady state. This leads to a relatively low aggregate stock-to-sales ratio under Pareto distribution. Suppose we change the distribution function to lognormal, $\log\varepsilon(i) \sim \mathcal{N}\left(\mu, \varphi^2\right)$, with mean μ and SD φ. Without loss of generality, we assume $\mu = -\frac{\varphi^2}{2}$, so that the expected value $E\left[\varepsilon\right] = 1$. With this distribution, the number of inventory-holding firms can become significantly

[35]This suggests that borrowing constraints are not the main friction that gives rise to the hump-shaped dynamics in the model.

larger in the steady state than that under the Pareto distribution. The reason is that lognormal distribution is more skewed toward lower realizations of $\varepsilon(i)$ and is less fat-tailed for high values of $\varepsilon(i)$ compared with Pareto distribution. Hence, firms expect to have relatively higher chances to draw low productivity (or high marginal costs), especially if the value of φ is relatively large. As a result, firms opt to reduce the cutoff ε^* and make positive inventory investment even in a relatively high-cost state. This leads to a much higher aggregate stock-to-sales ratio. For example, holding $\sigma = 10$, if we set $\varphi = 0.122$ so that the SD of $\varepsilon(i)$ is 0.122, the same as that under Pareto distribution with the shape parameter $\eta = 10.13$, the inventory-to-sales ratio is 0.51, significantly higher than that under Pareto distribution. Assuming a uniform distribution with $E[\varepsilon] = 1$ and $\varepsilon(i) \in [a, 2-a]$ can also give rise to a much lower cutoff and a much higher inventory-to-sales ratio than previously if a is close enough to 0. For example, given $\sigma = 10$, if $a = 0.15$, the predicted inventory-to-sales ratio is 1.8, more than seven times larger than that under Pareto distribution.

Again, because inventories are destabilizing in our model, the tendency to lower the cutoff ε^* and thus hold more inventories makes the economy more volatile. More importantly, with alternative distributions the model can still generate hump-shaped output dynamics, provided that the variance of $\varepsilon(i)$ is sufficiently large. The middle and right panels in Figure 2 depict the impulse responses of aggregate output under alternative distributions for various values of φ and a. Since the intensity of inventory accumulation depends positively on the variance of $\varepsilon(i)$, both the magnitude and the length of the hump-shaped impulse response functions increase with φ (lognormal case) and decreases with a (uniform case), as expected.

On the other hand, holding φ or a constant, increasing the value of σ (i.e., reducing the markup) stabilizes GDP because firms have weaker incentives to hold inventories to smooth sales when the profit function is less concave in sales. Inventories will disappear completely when $\sigma \to \infty$.[36] This suggests that the destabilization force of inventories to aggregate output also depends crucially on the monopoly power of firms (or the demand elasticity). Therefore, an alternative explanation for the "Great Moderation" of the U.S. economy since the mid-1980s may be that we have had a more competitive economy instead of an improved inventory management technology, since both can lead to reduced inventories.

5.3 Dispersion and Parameter Sensitivity

Although we have shown that inventories may significantly destabilize the economy in a model of imperfect competition with production-cost–smoothing motives, our results are nonetheless sensitive to a number of parameters in the model, most notably the distribution and the variance of the

[36]With lognormal or uniform distributions, the functions defined before equation (35), $\{g_1(\varepsilon^*), g_2(\varepsilon^*), h(\varepsilon^*), q(\varepsilon^*), ...\}$, are integrable for any values of σ. Hence, unlike the case of Pareto distribution, the effects of σ can now be disentangled from those of the variance of the distributions.

idiosyncratic shocks. Recall that we have calibrated the value of η in the benchmark model so that the the variance of GDP in the model matches its counterpart in the U.S. economy. In the middle and right panels of Figure 2, however, we allow parameter values for φ and a so that the impulse responses of GDP are hump-shaped. An important question is how realistic are such values? One concern is that the variance of $\varepsilon(i)$ required to generate hump-shaped impulse responses may be too large to be justified empirically.

We address this question by looking at the implied dispersions in productivity, prices, and sales across firms in our model when the variance of $\varepsilon(i)$ is chosen so that the SD of GDP in the model matches that in the data. The price of good i is given by equation (11), $p(i) = \frac{\sigma}{\sigma-1}\Phi\frac{1}{\varepsilon(i)}$, and sales are given by $y_t(i) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} Z_t \left(\frac{\varepsilon_t(i)}{\Phi_t}\right)^{\sigma}$. Following the empirical literature, we measure price dispersion by the SD of relative price changes across firms, $\omega_p = \sqrt{\int\left(\frac{p(i)}{\bar{p}}-1\right)^2 dF}$, where $\bar{p} = E\left[p(i)\right]$ is the mean; and we measure sales dispersion and productivity dispersion analogously by $\omega_y = \sqrt{\int\left(\frac{y(i)}{\bar{y}}-1\right)^2 dF}$ and $\omega_\varepsilon = \sqrt{\int\left(\frac{\varepsilon(i)}{\bar{\varepsilon}}-1\right)^2 dF}$, respectively. With Pareto distribution, we have $\omega_\varepsilon = \sqrt{\frac{1}{\eta(\eta-2)}}$, $\omega_p = \sqrt{\frac{1}{\eta(\eta+2)}}$, but ω_y does not exist[37]; with lognormal distribution, we have $\omega_\varepsilon = \omega_p = \varphi$ and $\omega_y = \sigma\varphi$; and with uniform distribution, we have $\omega_\varepsilon = \sqrt{\frac{2}{3}\frac{(2-a)^3-a^3}{(2-a)^2-a^2}-1}$, $\omega_p = \sqrt{\left(\frac{2-2a}{\ln\frac{2-a}{a}}\right)^2\frac{1}{(2-a)a}-1}$, and $\omega_y = \sqrt{\frac{(2-2a)(1+\sigma)^2}{\left((2-a)^{1+\sigma}-a^{1+\sigma}\right)^2}\frac{(2-a)^{1+2\sigma}-a^{1+2\sigma}}{1+2\sigma}-1}$.

The implied dispersions for productivity, prices, and sales in the model are reported in Table 5. The empirical estimates of productivity dispersion are around $0.84 \sim 1.67$ and sales dispersion around $0.6 \sim 0.75$, as reported by Bernard, Eaton, Jensen, and Kortum (2003, p. 1283 and table 2) for U.S. firms. The empirical range of price dispersion is around $0.3 \sim 0.37$, as reported by Reinsdorf (1994) for major U.S. cities.

[37] With $\eta < 2\sigma$, price dispersion is not well defined. This is because the second moment of $\varepsilon(i)^\sigma$ does not exist under Pareto distribution if $\eta < 2\sigma$.

Table 5. Dispersion of Productivity (ω_ε), Prices (ω_p), and Sales (ω_y)

	σ_{gdp}	$\frac{S}{Y}$	ω_ε	ω_p	ω_y
U.S. Data	0.024	0.55	0.84∼1.67	0.3∼0.37	0.6∼0.75
Control	0.016	0.00	0.00	0.00	0.00
Pareto ($\eta = 10.13$)	0.024	0.24	0.11	0.09	NA
($\eta = 6.25, \sigma = 6$)	0.024	0.53	0.19	0.14	NA
Lognormal ($\varphi = 0.35$)	0.024	1.28	0.35	0.35	3.50
Uniform ($a = 0.202$)	0.024	1.55	0.46	0.68	2.03

*SDs of GDP (σ_{gdp}) are based on HP-filtered samples. $\frac{S}{Y}$ denotes stock-to-sales ratio.

Several results emerge from Table 5: (1) Regardless of the type of distributions, the model with inventories can generate enough volatility in GDP to match that in the U.S. economy, whereas the control model without inventories explains only 0.67% of the data (assuming the SD of technology innovation is 0.007).[38] However, under lognormal and uniform distributions, the hump-shaped dynamics do not emerge unless the variance of $\varepsilon(i)$ are further increased (see Figure 2).[39] (2) As noted before, Pareto distribution tends to understate the inventory-to-sales ratio ($\frac{S}{Y}$) unless with large markups (20%) or external borrowing, and lognormal and uniform distributions tend to overstate the stock-to-sales ratio even with small markups and no borrowing. (3) To match the volatility of the U.S. economy, the assumed variance of productivity shocks ($\varepsilon(i)$) under various distributions does not exceed the empirical counterparts estimated by Bernard, Eaton, Jensen, and Kortum (2003) for U.S. firms; hence, the values of $\{\eta, \varphi, a\}$ assumed in Table 5 are conservative in this regard. However, under lognormal and uniform distributions, the implied price dispersion and sales dispersion are too large; and under Pareto distribution, sales dispersion is not well defined (thus the benchmark model cannot be evaluated along this dimension).

In addition to the above sensitivity results, we also note that the predicted SD of GDP is quite sensitive to the dispersion of productivity across firms, especially under Pareto distribution. For example, under Pareto distribution, when η increases from 10.13 to 10.15 (indicating the dispersion of prices declines by less than 0.5%), the implied SD of GDP drops by 20%, while the implied stock-to-sales ratio decreases by less than 0.5%. However, this is less the case under alternative distributions. For example, under uniform distribution, when a increases from 0.16 to 0.18, the price dispersion declines by 6%, the SD of GDP drops by 16%, and the stock-to-sales ratio decreases by 6%. Although we do not have an empirical base to judge these values, we believe that GDP is

[38] This problem of standard RBC models is discussed in detail by King and Rebel (1999).

[39] For example, under uniform distribution, we need $a \geq 0.16$ to generate a visible hump in GDP. However, we only need $a = 0.202$ to generate a sufficiently large SD in GDP to match the data.

too sensitive to the dispersion of productivity under Pareto distribution.

5.4 Persistent Idiosyncratic Shocks

We assumed *i.i.d* idiosyncratic cost shocks in the above analyses. This assumption, however, may not be entirely realistic and may have important consequences on our results. Here we relax this assumption by allowing for a second type of idiosyncratic cost shock, $\xi_t(i)$, which is serially correlated. Thus we modify firms' production function to

$$x_t(i) = A_t \xi_t(i) \varepsilon_t(i) k_t(i)^{\alpha} n_t(i)^{1-\alpha}, \tag{51}$$

where $\varepsilon_t(i)$ is the original *i.i.d* shocks with a compact support. To keep the model analytically tractable, we assume that $\xi(i)$ takes a finite number of discrete values. The original shock $\varepsilon(i)$ is still needed because only shocks with a continuum support can give rise to a time-varying cutoff, which affects aggregate inventories along the extensive margin; this margin is crucial for our inventory-accelerator mechanism. On the other hand, when the shocks are persistent, we need to take the joint distribution of inventory and the idiosyncratic shocks into consideration; the discrete nature of $\xi(i)$ makes the joint distribution tractable. So for simplicity, we assume that $\xi_t(i)$ follows a discrete $AR(1)$ process with serial correlation ρ_ξ: In each period t, there is a constant probability ρ_ξ such that $\xi_t(i) = \xi_{t-1}(i)$ and probability $1 - \rho_\xi$ that $\xi_t(i)$ is redrawn from the following discrete invariant distribution, $\Pr\left[\xi_t(i) = \xi_j\right] = \tau_j$, for $j = 1, 2, ..., J$; where $\left\{\xi_j, \tau_j\right\}$ are constant. That is, $\xi_t(i)$ may either remain at its last-period value (with probability ρ_ξ) or take one of J discrete new values if it is redrawn (with probability $1 - \rho_\xi$). Without loss of generality, assume the mean $\sum \tau_j \xi_j = 1$. It is easy to see that $E_t\left[\xi_{t+1}(i)|\xi_t(i)\right] = \rho_\xi \xi_t(i) + (1 - \rho_\xi)$; hence, we have $E_t\left[\xi_{t+1}(i) - 1|\xi_t(i)\right] = \rho_\xi[\xi_t(i) - 1]$, or $E_t\left[\xi_{t+\ell}(i) - 1|\xi_t(i)\right] = \left(\rho_\xi\right)^{\ell}[\xi_t(i) - 1]$ for $\ell = 1, 2, ..., \infty$.

With the second idiosyncratic shock $\xi_t(i)$, the derivation of aggregate dynamics becomes more complicated because of tedious aggregation. Nonetheless, the model remains analytically tractable. We report our main results here and provide the details of the analyses in the Appendix. We calibrate the distribution of $\xi_t(i)$ by setting $J = 2$, $\xi_1 = 0.97$, $\xi_2 = 1.03$, $\tau_1 = \tau_2 = 0.5$, and $\rho_\xi = 0.9$. The other structural parameters remain the same as in Table 1. The impulse responses of the model to an aggregate productivity shock are depicted in Figure 3. The figure shows that adding persistent idiosyncratic shocks to the model does not diminish the hump-shaped dynamics of our model; in fact, it enhances the inventory-accelerator mechanism and amplifies the destabilizing role of inventories to the economy. The intuition is that firms with a high realization of $\xi_t(i)$ expect their marginal costs to remain low with high probability; hence, they reduce the cutoff (ε_{jt}^*) and thus have a stronger incentive than otherwise to accumulate inventories, which more than offsets the actions of the firms with a low realization of $\xi_t(i)$. This reinforces the original inventory-accelerator mechanism.

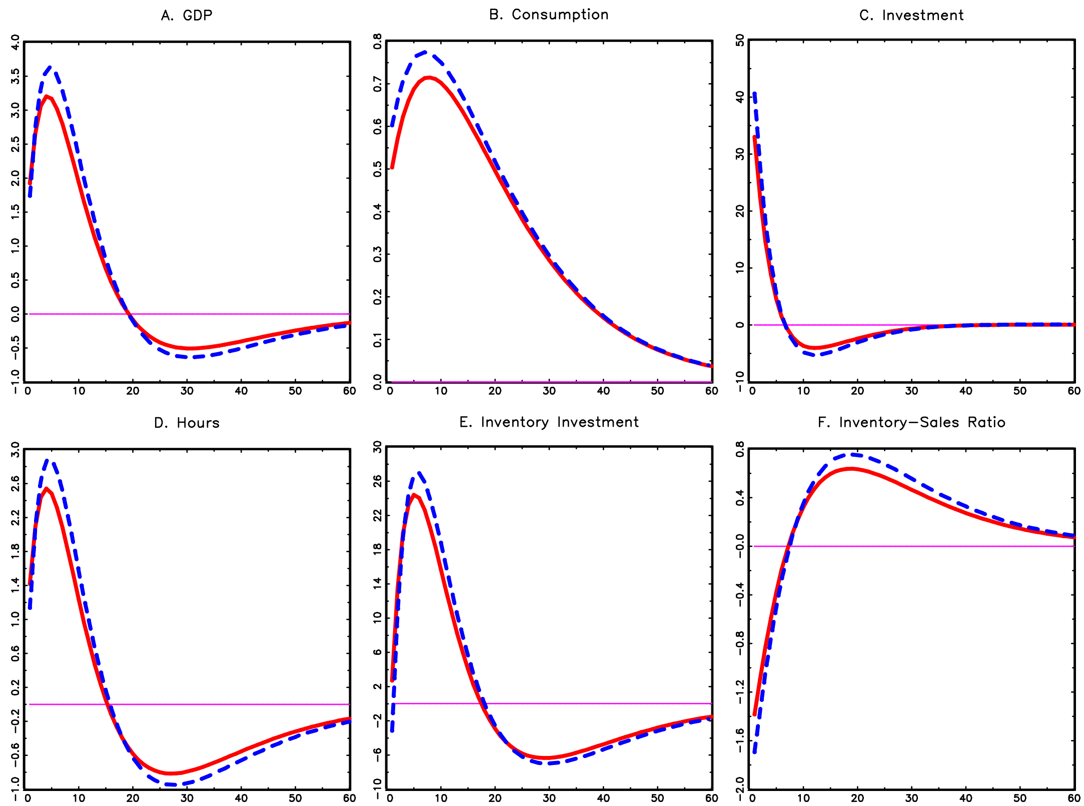

Figure 3. Impulse Responses to a 1% TFP Shock (— Benchmark, - - - Persistent shock)

5.5 Sales Smoothing and the Conventional Wisdom

Sales smoothing at the firm level is a key micromechanism that makes inventories destabilizing in our general-equilibrium model. Under sales smoothing, the conventional partial-equilibrium argument about the destabilizing role of inventories becomes approximately valid. Namely, if movements in final sales are taken as given, procyclical inventory investment would then imply that changes in inventories must raise the cyclical volatility of GDP. Although this argument ignores the general-equilibrium feedback of inventories on sales, it nonetheless holds approximately true in our model because sales smoothing makes sales a less ideal buffer to dampen the destabilizing effects of inventories.[40]

[40] On the other hand, the procyclical cutoff (the extensive margin) in our model amplifies the destabilizing role of inventories under sales smoothing and is key to generating a countercyclical stock-to-sales ratio and hump-shaped dynamics in aggregate demand. For example, in a counterfactual experiment, we delete equation (33) from our model (which determines the cutoff) and let the cutoff ε_t^* be exogenously set so that it responds to technology shocks negatively. The results show that inventories are still destabilizing to GDP but mainly during the impact period (because of sales smoothing); also, the hump-shaped dynamics disappear completely and the inventory-to-sales ratio becomes procyclical. In such a case, the lack of a hump-shaped propagation mechanism reduces the destabilizing effect of inventories substantially compared with the case of a procyclical cutoff.

We believe this intuition also applies to other general-equilibrium inventory models. For example, when we introduce habit formation and capital adjustment costs into the model of KT (2007a), so that final sales in their model are severely smoothed and much less able to serve as a buffer for dampening inventory fluctuations through the endogenous general-equilibrium feedback of inventories on sales, the volatility of aggregate output can rise significantly with inventories. More specifically, suppose we change the utility function of the KT model to $\log(c_t - \theta_c \bar{c}_{t-1}) + \eta\left(1 - n_t^h\right)$, where $\bar{c}$ measures the consumption level of average households in the economy, and add capital-adjustment costs $\frac{\theta_k}{2}\left(k_{t+1} - \bar{k}\right)^2$, where $\bar{k}$ is the steady-state level of the capital stock. When we set the habit-formation parameter $\theta_c = 0.8$ and the adjustment-cost parameter $\theta_k = 0.1$, the SD of GDP in the KT model increases by about 10% more with inventories than without.[41] This suggests that the sales-smoothing mechanism under imperfect competition in our model is important in driving our results and that the general-equilibrium effects of inventories on final sales emphasized by KT (2007a) are indeed crucial for understanding whether inventories are destabilizing or not.[42]

Notice that sales smoothing does not imply that inventories stabilize sales. On the contrary, it implies that inventories are less able to dampen (stabilize) sales through the price mechanism. For this reason, the main difference between our model and those of KT (2007a) and Wen (2008) can be summarized as follows. In our model, inventories enhance sales (especially capital spending) through a general-equilibrium trade-off between inventory investment and capital accumulation, which postpones aggregate inventory investment through adjustments along the extensive margin. In the model of KT (2007a), inventories stabilize sales through resource relocation from the final-goods sector to the intermediate-goods sector, thereby dampening final demand. In Wen's (2008) model, inventories under the stockout-avoidance motive lead to a procyclical asset value of inventories, which discourages sales in a boom and encourages sales in a recession, therefore reducing the volatility of aggregate demand more than raising the variance of aggregate production. These differences are driven mainly by the distinct microlevel motives for firms to hold inventories. These diverse motives for holding inventories lead to different general-equilibrium effects of inventories on final sales, and thus on the variance of aggregate output.

6 Conclusion

This paper provides a general-equilibrium model of inventories with microfoundations. In the model, idiosyncratic cost shocks can induce monopolistic firms to bunch production and smooth

[41] Computer programs for solving the KT model with habit formation and adjustment costs are available upon requests.

[42] This is why inventories become less destabilizing in our model as we increase the value of σ (i.e., the inverse of the markup) because a higher value of σ makes a firm's profit function less concave in sales and thus reduces firms' incentives for sales smoothing.

sales by holding inventories. The intertemporal substitution of production activities at the firm level is based on a production-cost–smoothing motive emphasized in the empirical and partial-equilibrium literature by Eichenbaum (1989) and others. However, our model rationalizes the ad hoc cost functions and the associated target-adjustment equations assumed in this empirical literature. The predictions of our general-equilibrium model are broadly consistent with the stylized facts of aggregate inventory behavior, such as procyclical inventory investment and countercyclical inventory-to-sales ratio. Our analysis also reveals that inventory investment motivated by sales-smoothing may not only amplify aggregate shocks but also propagate them. This finding confirms (within a special context) a long-standing conjecture in the history of economic thought (e.g., Metzler, 1941) that inventories can serve as an accelerator of the business cycle.

In light of the provocative findings of Khan and Thomas (2007a) and Wen (2008), our analysis suggests that whether inventories are stabilizing or destabilizing to the aggregate economy depends crucially on the sources of firm-level uncertainties and incentives for holding inventories. If idiosyncratic marginal-cost shocks dominate idiosyncratic demand shocks, for example, then the sales-smoothing motive studied in this paper is more important than the stockout-avoidance motive studied by Wen (2008); hence, inventories are destabilizing; otherwise, inventories may be stabilizing. In other words, in contrast to the partial-equilibrium tradition of Blinder (1981, 1986a, 1990) and others in the earlier literature, the insight gained from general-equilibrium analysis (Khan and Thomas, 2007a; Wen, 2008; and this paper) is that the destabilizing nature of inventories does not hinge completely on whether inventory investment is procyclical (or whether production is more variable than sales), but crucially on the specific motives for holding inventories and the dynamic impact of such motives on aggregate sales.

References

[1] Axtell, R., 2001, Zipf distribution of US firm sizes, Science 293, 1818–1820.

[2] Basu, S. and J. Fernald, 1997, Returns to scale in U.S. production: Estimates and implications, Journal of Political Economy 105, 249–83.

[3] Bernard, A., J. Eaton, B. Jensen, and S. Kortum, 2003, Plants and productivity in international trade, American Economic Review 93(4), 1268-1290.

[4] Bils, M. and J. Kahn, 2000, What inventory behavior tells us about business cycles? American Economic Review 90(3), 458-481.

[5] Blanchard, O., 1983, The production and inventory behavior of the American automobile industry, Journal of Political Economy 91(3), 365-400.

[6] Blinder, A., 1981, Retail inventory behavior and business fluctuations, Brookings Papers on Economic Activity 12(2), 443-520.

[7] Blinder, A., 1986a, Can the production smoothing model of inventory behavior be saved? Quarterly Journal of Economics 101(3), 431-453.

[8] Blinder, A., 1986b, More on the speed of adjustment in inventory models, Journal of Money, Credit, and Banking 18(3), 355-365.

[9] Blinder, A., 1990, Inventory Theory and Consumer Behavior, University of Michigan Press.

[10] Blinder, A. and L. Maccini, 1991, Taking stock: A critical assessment of recent research on inventories, Journal of Economic Perspectives 5(1), 73-96.

[11] Browna, W. and U. Haeglerb, 2004, Financing constraints and inventories, European Economic Review 48, 1091-1123.

[12] Cogley, T. and J. Nason, 1995, Output dynamics in real-business-cycle models, American Economic Review 85(3), 492-511.

[13] Durlauf, S. and L. Maccini, 1995, Measuring noise in inventory models, Journal of Monetary Economics 36(1), 65-89.

[14] Eichenbaum, M., 1984, Rational expectations and the smoothing properties of inventories of finished goods, Journal of Monetary Economics 14(1), 71-96.

[15] Eichenbaum, M., 1989, Some empirical evidence on the production level and production cost smoothing models of inventory investment, American Economic Review 79(4), 853-864.

[16] Fisher, J. and A. Hornstein, 2000, (S,s) inventory policies in general equilibrium, Review of Economic Studies 67(1), 117-145.

[17] Gertler, M. and S. Gilchrist, 1994, Monetary policy, business cycles, and the behavior of small manufacturing firms, Quarterly Journal of Economics 109(2), 309-340.

[18] Haltiwanger, J. and L. Maccini, 1988, A model of inventory and layoff behaviour under uncertainty, Economic Journal 98(392), 731-745.

[19] Helpman, E., M. Melitz, and S. R. Yeaple, 2004, Export versus FDI with heterogeneous firms, American Economic Review 94(1), 300-316.

[20] Humphreys, B., L. Maccini, and S. Schuh, 2001, Input and output inventories, Journal of Monetary Economics, 47(2), 347-375.

[21] Iacoviello, M., F. Schiantarelli, and S. Schuh, 2007, Input and output inventories in general equilibrium, Boston College Working Paper.

[22] Irvinea, O. and S. Schuh, 2005a, Inventory investment and output volatility, International Journal of Production Economics 93-94(8), 75-86.

[23] Irvinea, O. and S. Schuh, 2005b, The roles of comovement and inventory investment in the reduction of output volatility, FRB of Boston Working Paper No. 05-9.

[24] Kahn, J., 1987, Inventories and the volatility of production, American Economic Review 77(4), 667-679.

[25] Kahn, J., 2008, Durable goods inventories and the Great Moderation, FRB of New York Staff Report No. 325.

[26] Kahn, J., M. McConnell, and G. Perez-Quiros, 2002, On the causes of the increased stability of the U.S. economy, Economic Policy Review 8(1), 183-202.

[27] Kehoe, P. and F. Perri, 2002, International business cycles with endogenous incomplete markets, Econometrica 70(3), 907-928.

[28] Khan, A. and J. Thomas, 2007a, Inventories and the business cycle: An equilibrium analysis of (S,s) policies, American Economic Review 97(4), 1165-1188.

[29] Khan, A. and J. Thomas, 2007b, Explaining inventories: A business cycle assessment of the stockout avoidance and (S,s) motives, Macroeconomic Dynamics 11(5), 638-664.

[30] King, R. and S. Rebelo, 1999, Resuscitating real business cycles, In Handbook of Macroeconomics Volume 1b, ed. J. B. Taylor and M. Woodford, 927-1007. Amsterdam: Elsevier Science, North-Holland.

[31] Kryvtsov, O. and V. Midrigany, 2008, Inventories, markups, and real rigidities in menu cost models, New York University Working Paper.

[32] Kydland, F. and Prescott, E., 1982, Time to build and aggregate fluctuations, Econometrica 50, 1345-1370.

[33] Luttmer, E., 2007, Selection, growth, and the size distribution of firms, Quarterly Journal of Economics 122(3), 1103-1144.

[34] Metzler, LA., 1941, The nature and stability of inventory cycles, Review of Economics and Statistics 23(3), 113-129.

[35] Miao, J., 2005, Optimal capital structure and industry dynamics, Journal of Finance LX(6), 2621-2660.

[36] Ramey, V., 1989, Inventories as factors of production and economic fluctuations, American Economic Review, 79(3), 338–354.

[37] Ramey, V., 1991, Nonconvex costs and the behavior of inventories, Journal of Political Economy 99(2), 306-334.

[38] Ramey, V. and D. Vine, 2006, Declining volatility in the U.S. automobile industry, American Economic Review 96(5), 876-1889.

[39] Reinsdorf, M., 1994, New evidence on the relation between inflation and price dispersion, American Economic Review 84(3), 720-731.

[40] Romer, D., 2001, Advanced Macroeconomics (second edition), New York: McGraw-Hill Higher Education.

[41] Rotemberg, J. and M. Woodford, 1995, Dynamic general equilibrium models with imperfectly competitive product markets, In Frontiers of Business Cycle Research, edited by Thomas F. Cooley. Princeton, NJ: Princeton Univ. Press.

[42] Schuh, S., 1996, Evidence on the link between firm-level and aggregate inventory behavior, Federal Reserve Board of Governors Finance and Economics Discussion Paper 96-46.

[43] Wen, Y., 2005, Understanding the inventory cycle, Journal of Monetary Economics 52(8), 1533-55.

[44] Wen, Y., 2008, Inventories, liquidity, and the macroeconomy, Federal Reserve Bank of St. Louis Working Paper 2008-045.

[45] West, K., 1986, A variance bounds test of the linear quadratic inventory model, Journal of Political Economy 94(2), 374-401.

Appendix

This appendix provides details for analytically solving the model with persistent idiosyncratic cost shocks, $\xi_t(i)$. The complication arising from persistent idiosyncratic shocks is aggregation. However, if the persistent idiosyncratic shocks assume discrete values, aggregation is still manageable, although it is more tedious than the case of $i.i.d$ shocks.

The distribution of $\xi_t(i)$ is described in Section 5.3. With the ξ shock introduced, the firm's maximization program becomes

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \Lambda_t \left\{ p_t(i) y_t(i) - \frac{\Phi_t}{\varepsilon_t(i)\xi_t(i)} x_t(i) \right\} \tag{52}$$

subject to

$$y_t(i) + s_t(i) = x_t(i) + (1-\delta_s) s_{t-1}(i) \tag{53}$$

$$s_t(i) \geq 0 \tag{54}$$

$$p_t(i) y_t(i) - \frac{\Phi_t}{\varepsilon_t(i)\xi_t(i)} x_t(i) \geq 0. \tag{55}$$

Denoting $\{\lambda_t(i), \pi_t(i), \mu_t(i)\}$ as the Lagrangian multipliers of constraints (53)-(55), respectively, the first-order conditions of $\{x_t(i), y_t(i), s_t(i)\}$ are given by

$$\frac{\Phi_t}{\varepsilon_t(i)\xi_t(i)} (1 + \mu_t(i)) = \lambda_t(i) \tag{56}$$

$$\left(\frac{\sigma - 1}{\sigma} \right) y_t(i)^{-\frac{1}{\sigma}} Z_t^{\frac{1}{\sigma}} (1 + \mu_t(i)) = \lambda_t(i) \tag{57}$$

$$\lambda_t(i) = \beta(1-\delta_s) E_t \left[\frac{\Lambda_{t+1}}{\Lambda_t} \lambda_{t+1}(i) \right] + \pi_t(i). \tag{58}$$

The optimal sales is then given by

$$y_t(i) = \left(\frac{\sigma - 1}{\sigma} \right)^{\sigma} Z_t \left(\frac{\varepsilon_t(i)\xi_t(i)}{\Phi_t} \right)^{\sigma}, \tag{59}$$

and the monopoly price is given by

$$p_t(i) = \frac{\sigma}{\sigma - 1} \frac{\Phi_t}{\varepsilon_t(i)\xi_t(i)}. \tag{60}$$

Decision Rules for Inventories

Because $\xi_t(i)$ may take J different values in any period, there exist J corresponding cutoffs $\varepsilon^*_{j,t}$, $j = 1, 2, ..., J$. For each value ξ_j in period t, we have the following two cases to consider for a firm's decision rules. We call a firm with $\xi_t(i) = \xi_j$ a type-j firm.

Case A. $\varepsilon_t(i) \geq \varepsilon^*_{j,t}$. Suppose $s_t(i) > 0, \pi_t(i) = 0$. In such a case, equation (58) implies $\lambda_t(i) = \beta(1-\delta_s)E_t \frac{\Lambda_{t+1}}{\Lambda_t}\lambda_{t+1}(i)$, and equation (56) implies

$$\frac{\Phi_t}{\varepsilon_t(i)\xi_j} \leq \beta(1-\delta_s)E_t\frac{\Lambda_{t+1}}{\Lambda_t}\left[\rho_\xi \lambda_{j,t+1}(i) + (1-\rho_\xi)\bar{\lambda}_{t+1}(i)\right], \tag{61}$$

where $\lambda_{j,t+1}(i)$ is the value of $\lambda_{t+1}(i)$ if $\xi_{t+1}(i) = \xi_t(i) = \xi_j$, and $\bar{\lambda}_{t+1}(i) \equiv \sum_{\ell=1}^J \tau_\ell \lambda_{\ell,t+1}(i)$ is the average value of $\lambda_{\ell,t+1}(i)$ over all other possible values of ξ if $\xi_{t+1}(i)$ is redrawn. Because $\mu_t(i) \geq 0$, this implies $\varepsilon_t(i) \geq \frac{\Phi_t}{\beta(1-\delta_s)E_t\frac{\Lambda_{t+1}}{\Lambda_t}\left[\rho_\xi\lambda_{j,t+1}(i)+(1-\rho_\xi)\bar{\lambda}_{t+1}(i)\right]} \equiv \varepsilon^*_{j,t}$, which defines the cutoff value $\varepsilon^*_{j,t}$ and the relationship

$$\beta(1-\delta_s)E_t\frac{\Lambda_{t+1}}{\Lambda_t}\left[\rho_\xi \lambda_{j,t+1}(i) + (1-\rho_\xi)\bar{\lambda}_{t+1}(i)\right] \equiv \frac{\Phi_t}{\xi_j \varepsilon^*_{j,t}}. \tag{62}$$

Equation (56) then further implies $1+\mu_t(i) = \frac{\varepsilon_t(i)}{\varepsilon^*_{j,t}}$. Hence, we conclude that $\mu_t(i) > 0$ if $\varepsilon_t(i) > \varepsilon^*_{j,t}$. In such a case, profit constraint binds, $p_t(i)y_t(i) = \frac{\Phi_t}{\varepsilon_t(i)\xi_j}x_t(i)$, which together with (60) implies

$$x_{j,t}(i) = \frac{\sigma}{\sigma-1}y_{j,t}(i). \tag{63}$$

This implies $x_t(i) > y_t(i)$, or production exceeds sales.

Case B. $\varepsilon_t(i) < \varepsilon^*_{j,t}$. Suppose $p_t(i)y_t(i) > \frac{\phi_t}{\varepsilon_t(i)\xi_j}x(i), \mu_t(i) = 0$. Then equations (56) and (62) imply $\frac{\Phi_t}{\varepsilon_t(i)\xi_j} = \lambda_t(i) = \frac{\Phi_t}{\varepsilon^*_{j,t}\xi_j} + \pi_t(i)$, which implies $\pi_t(i) = \frac{\Phi_t}{\varepsilon_t(i)} - \frac{\Phi_t}{\varepsilon^*_{j,t}} > 0$. Hence, we have $s_t(i) = 0$. In such a case, the firm opts to stockout and the resource identity implies $x_t(i) = y_t(i) - (1-\delta_s)s_{t-1}(i)$.

The decision rules of a type-j firm with $\xi_t(i) = \xi_j$ in period t can thus be summarized by the following policy functions:

$$y_t(\varepsilon_t, \xi_j, s_{t-1}) = \left(\frac{\sigma-1}{\sigma}\right)^\sigma Z_t\left(\frac{\varepsilon_t\xi_j}{\Phi_t}\right)^\sigma \tag{64}$$

$$x_t(\varepsilon_t, \xi_j, s_{t-1}) = \begin{cases} \frac{\sigma}{\sigma-1}y_t(\varepsilon_t, \xi_j, s_{t-1}) & \text{if } \varepsilon_t \geq \varepsilon^*_{j,t} \\ y_t(\varepsilon_t, \xi_j, s_{t-1}) - (1-\delta_s)s_{t-1} & \text{if } \varepsilon_t < \varepsilon^*_{j,t} \end{cases} \tag{65}$$

$$s_t(\varepsilon_t, \xi_j, s_{t-1}) = \begin{cases} \frac{1}{\sigma-1} y_t(\varepsilon_t, \xi_j, s_{t-1}) + (1-\delta_s) s_{t-1} & \text{if } \varepsilon_t \geq \varepsilon^*_{j,t} \\ 0 & \text{if } \varepsilon_t < \varepsilon^*_{j,t} \end{cases}. \tag{66}$$

Since the shadow value of inventory satisfies

$$\lambda_t(\varepsilon_t, \xi_j, s_{t-1}) = \begin{cases} \frac{\Phi_t}{\varepsilon^*_{j,t}\xi_j} & \text{if } \varepsilon_t \geq \varepsilon^*_{jt} \\ \frac{\Phi_t}{\varepsilon_t \xi_j} & \text{if } \varepsilon_t < \varepsilon^*_{jt} \end{cases}, \tag{67}$$

equation (62) becomes

$$\frac{\Phi_t}{\xi_j \varepsilon^*_{j,t}} = \beta(1-\delta_s) E_t \frac{\Lambda_{t+1}}{\Lambda_t} \left[\frac{\rho_\xi \Phi_{t+1}}{\xi_j} E_\varepsilon \max\left\{ \frac{1}{\varepsilon_{t+1}}, \frac{1}{\varepsilon^*_{jt+1}} \right\} + \sum_{\ell=1}^{J} \frac{(1-\rho_\xi)\Phi_{t+1}}{\xi_\ell} \tau_\ell E_\varepsilon \max\left\{ \frac{1}{\varepsilon_{t+1}}, \frac{1}{\varepsilon^*_{\ell t+1}} \right\} \right]. \tag{68}$$

This equation determines the cutoff $\varepsilon^*_{j,t}$. Notice that the optimal cutoff depends negatively on the value of ξ_j because a higher productivity enables more type-j firms to accumulate inventories, other things equal.

Aggregation

By the law of large numbers, the final-goods production function, $Z_t = \left[\int y_t(i)^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}$, and the decision rule for firms' sales, $y_t(i) = \left(\frac{\sigma-1}{\sigma}\right)^\sigma Z_t \left(\frac{\varepsilon_t(i)\xi_t(i)}{\Phi_t}\right)^\sigma$, imply

$$Z_t = \left(\frac{\sigma-1}{\sigma}\right)^\sigma \frac{Z_t}{\Phi_t^\sigma} \left[\int \varepsilon^{\sigma-1} dF(\varepsilon) \right]^{\frac{\sigma}{\sigma-1}} \left[\sum \tau_\ell \xi_\ell^{\sigma-1} \right]^{\frac{\sigma}{\sigma-1}}. \tag{69}$$

This implies that the aggregate marginal cost is constant,

$$\Phi_t = \Phi = \left(\frac{\sigma-1}{\sigma}\right) \left[\int \varepsilon^{\sigma-1} dF(\varepsilon) \right]^{\frac{1}{\sigma-1}} \left[\sum \tau_\ell \xi_\ell^{\sigma-1} \right]^{\frac{1}{\sigma-1}}. \tag{70}$$

Define $Y_t \equiv \int_0^1 y_t(i)di$, $K_t \equiv \int_0^1 k_t(i)di$, $N_t \equiv \int_0^1 n_t(i)di$, $X_t \equiv \int_0^1 x_t(i)di$, and $S_t \equiv \int_0^1 s_t(i)di$. The level of aggregate sales is given by

$$Y_t = \left(\frac{\sigma-1}{\sigma}\right)^\sigma \frac{Z_t}{\Phi_t^\sigma} \left[\int \varepsilon^\sigma dF(\varepsilon) \right] \left[\sum \tau_\ell \xi_\ell^\sigma \right] \equiv P Z_t, \tag{71}$$

where

$$P \equiv \frac{\left[\int \varepsilon^\sigma dF(\varepsilon) \right] \left[\sum \tau_\ell \xi_\ell^\sigma \right]}{\left[\int \varepsilon^{\sigma-1} dF(\varepsilon) \right]^{\frac{\sigma}{\sigma-1}} \left[\sum \tau_\ell \xi_\ell^{\sigma-1} \right]^{\frac{\sigma}{\sigma-1}}} \tag{72}$$

can be interpreted as an aggregate measure of the relative price of the final good in terms of intermediate goods. Note $P = 1$ if ε and ξ are constant across firms.

Define the index function,

$$\mathbf{1}_{\xi_t(i)=\xi_j} = \begin{cases} 1 & \text{if } \xi_t(i) = \xi_j \\ 0 & \text{otherwise} \end{cases}, \tag{73}$$

to facilitate the aggregation of firm-level variables. Note that $\mathbf{1}_{\xi_t(i)=\xi_j} = 1$ with probability τ_j. We have

$$X_t \equiv \int_0^1 x_t(i)di = \int_0^1 \left(\sum x_t(i)\mathbf{1}_{\xi_t(i)=\xi_j}\right) di. \tag{74}$$

Define

$$X_{j,t} \equiv \frac{\int_0^1 x_t(i)\mathbf{1}_{\xi_t(i)=\xi_j} di}{\int_0^1 \mathbf{1}_{\xi_t(i)=\xi_j} di} = \frac{\int_0^1 x_{j,t}(i)di}{\tau_j} \tag{75}$$

as the average production of type-j firms with $\xi_t(i) = \xi_j$. So we have

$$X_t = \sum \tau_j X_{j,t}. \tag{76}$$

Define

$$S_{j,t-1} \equiv \tau_j^{-1} \int_0^1 s_{t-1}(i)\mathbf{1}_{\xi_t(i)=\xi_j} di \tag{77}$$

as the average past inventory stock of all type-j firms from the perspective of period t. Then by equation (65), we have

$$\begin{aligned} X_{jt} &= \tau_j^{-1} \int_0^1 x_{j,t}(i)di \\ &= \int_{\varepsilon \geq \varepsilon_{j,t}^*} \left[\frac{\sigma}{\sigma - 1}\left(\frac{\sigma - 1}{\sigma}\right)^\sigma Z_t \left(\frac{\varepsilon\xi_j}{\Phi_t}\right)^\sigma\right] dF(\varepsilon) \\ &\quad + \int_{\varepsilon < \varepsilon_{j,t}^*} \left(\frac{\sigma - 1}{\sigma}\right)^\sigma Z_t \left(\frac{\varepsilon\xi_j}{\Phi_t}\right)^\sigma dF(\varepsilon) - (1 - \delta_s)S_{j,t-1}F(\varepsilon_{j,t}^*), \end{aligned} \tag{78}$$

or equivalently,

$$X_{jt} = Z_t \left(\frac{\sigma - 1}{\sigma}\right)^\sigma \Phi_t^{-\sigma}\xi_j^\sigma \left[\frac{\sigma}{\sigma - 1}\int_{\varepsilon \geq \varepsilon_{j,t}^*} \varepsilon^\sigma dF(\varepsilon) + \int_{\varepsilon < \varepsilon_{j,t}^*} \varepsilon^\sigma dF(\varepsilon)\right] - (1 - \delta_s)F(\varepsilon_{j,t}^*)S_{j,t-1}. \tag{79}$$

Note that $s_{t-1}(i)$ is independent of $\varepsilon_t(i)$ and $\xi_t(i)$.

If we group type-j firms in period $t+1$ into two groups, one with probability ρ_ξ, the firms do not make a new draw of $\xi_{t+1}(i)$ so that $\xi_{t+1}(i) = \xi_t(i) = \xi_j$; and another group with probability $1-\rho_\xi$, the firms make a new draw for $\xi_{t+1}(i)$ so that $\xi_{t+1}(i) = \xi'_{t+1}(i) = \xi_j$. Then the average inventory stock of type-j firms at the end of period t is given by

$$\begin{aligned} S_{j,t} &= \tau_j^{-1} \int s_t(i) \mathbf{1}_{\xi_{t+1}(i)=\xi_j} di \\ &= \tau_j^{-1}\rho_\xi \int s_t(i)\mathbf{1}_{\xi_{t+1}(i)=\xi_t(i)=\xi_j} di + \tau_j^{-1}(1-\rho_\xi)\int s_t(i)\mathbf{1}_{\xi_{t+1}(i)=\xi'_{t+1}(i)=\xi_j} di \\ &= \tau_j^{-1}\rho_\xi \int s_t(i)\mathbf{1}_{\xi_t(i)=\xi_j} di + \tau_j^{-1}(1-\rho_\xi)\int s_t(i) di \int \mathbf{1}_{\xi'_{t+1}(i)=\xi_j} di, \end{aligned} \tag{80}$$

where the last term derives from the fact that the new draw $\xi'_{t+1}(i)$ is independent of the firm's inventory history. Since $\int \mathbf{1}_{\xi'_{t+1}(i)=\xi_j} di = \tau_j$, we can rewrite the above definition as

$$S_{j,t} = \tau_j^{-1}\rho_\xi \int s_t(i)\mathbf{1}_{\xi_t(i)=\xi_j} di + (1-\rho_\xi) S_t, \tag{81}$$

where $S_t \equiv \int s_t(i)di$. It is easy to check that $S_t = \sum \tau_j S_{jt}$. We need to calculate the average inventories of type-j firms in period t, namely, the first term ($\int s_t(i)\mathbf{1}_{\xi_t(i)=\xi_j} di$) in the above equation. Using the decision rule (66), we have

$$\begin{aligned} &\tau_j^{-1} \int s_t(i)\mathbf{1}_{\xi_t(i)=\xi_j} di \\ =& \int_{\varepsilon \geq \varepsilon^*_{j,t}} \left[\frac{1}{\sigma-1}\left(\frac{\sigma-1}{\sigma}\right)^\sigma Z_t \left(\frac{\varepsilon \xi_j}{\Phi_t}\right)^\sigma \right] dF(\varepsilon) + (1-\delta_s)\tau_j^{-1}\int_0^1 s_{t-1}(i)\mathbf{1}_{\xi_t(i)=\xi_j}\mathbf{1}_{\varepsilon_t(i)\geq \varepsilon^*_{j,t}} di \\ =& Z_t \left(\frac{\sigma-1}{\sigma}\right)^\sigma \Phi_t^{-\sigma}\xi_j^\sigma \frac{1}{\sigma-1}\int_{\varepsilon\geq\varepsilon^*_{j,t}} \varepsilon^\sigma dF(\varepsilon) + (1-\delta_s)[1-F(\varepsilon^*_{j,t})]S_{j,t-1}. \end{aligned} \tag{82}$$

It follows that

$$S_{j,t} = \rho_\xi Z_t \left(\frac{\sigma-1}{\sigma}\right)^\sigma \frac{\Phi_t^{-\sigma}\xi_j^\sigma}{\sigma-1}\int_{\varepsilon\geq\varepsilon^*_{j,t}} \varepsilon^\sigma dF(\varepsilon) + \rho_\xi(1-\delta_s)[1-F(\varepsilon^*_{j,t})]S_{j,t-1} + \left(1-\rho_\xi\right) S_t. \tag{83}$$

We need to check that the following resource identify holds:

$$Y_t + S_t = (1-\delta_s)S_{t-1} + X_t. \tag{84}$$

Since

$$X_t = Z_t \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \Phi_t^{-\sigma} \sum_{j=1}^{J} \tau_j \xi_j^{\sigma} \left[\int_{\varepsilon<\varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon) + \frac{\sigma}{\sigma-1} \int_{\varepsilon \geq \varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon)\right] - (1-\delta_s) \sum_{j=1}^{J} F(\varepsilon_{j,t}^*) \tau_j S_{j,t-1} \tag{85}$$

and

$$S_t = Z_t \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \frac{\Phi_t^{-\sigma}}{\sigma-1} \sum_{j=1}^{J} \tau_j \xi_j^{\sigma} \int_{\varepsilon \geq \varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon) + (1-\delta_s) \sum_{j=1}^{J} [1 - F(\varepsilon_{j,t}^*)] \tau_j S_{j,t-1}, \tag{86}$$

we have

$$Y_t + S_t = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \frac{Z_t}{\Phi_t^{\sigma}} \left[\int \varepsilon^{\sigma} dF(\varepsilon)\right] \left[\sum \tau_j \xi_j^{\sigma}\right] Z_t + Z_t \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \frac{\Phi_t^{-\sigma}}{\sigma-1} \sum_{j=1}^{J} \tau_j \xi_j^{\sigma} \int_{\varepsilon \geq \varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon)$$

$$+ (1-\delta_s) \sum_{j=1}^{J} [1 - F(\varepsilon_{j,t}^*)] \tau_j S_{j,t-1}. \tag{87}$$

On the other hand, we know that

$$X_t + (1-\delta) S_{t-1} = Z_t \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \Phi_t^{-\sigma} \sum_{j=1}^{J} \tau_j \xi_j^{\sigma} \left[\int_{\varepsilon<\varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon) + \frac{\sigma}{\sigma-1} \int_{\varepsilon \geq \varepsilon_{j,t}^*} \varepsilon^{\sigma} dF(\varepsilon)\right]$$

$$- (1-\delta_s) \sum_{j=1}^{J} F(\varepsilon_{j,t}^*) S_{j,t-1} + (1-\delta) \sum_{j=1}^{J} \tau_j S_{jt-1}. \tag{88}$$

Hence, it can be verified that $Y_t + S_t = (1-\delta_s) S_{t-1} + X_t$.

The factor demand functions imply $r_t K_t = \alpha \Phi M_t$ and $w_t N_t = (1-\alpha)\Phi M_t$, where $M_t \equiv \int_0^1 \frac{x_t(i)}{\varepsilon_t(i)\xi_t(i)} di$. Since $\Phi \equiv \frac{1}{A}\left(\frac{r_t}{\alpha}\right)^{\alpha}\left(\frac{w_t}{1-\alpha}\right)^{1-\alpha}$, these aggregate factor demand functions imply the aggregate production function,

$$M_t = A_t K_t^{\alpha} N_t^{1-\alpha}. \tag{89}$$

Using the definition of M_t, we also have

$$\int_0^1 \frac{x_t(i)}{\varepsilon_t(i)\xi_t(i)} di = \int_0^1 \frac{x_t(i)}{\varepsilon_t(i)\xi_t(i)} \left[\sum_{j=1}^{J} \mathbf{1}_{\xi_t(i)=\xi_j}\right] di. \tag{90}$$

Defining

$$M_{j,t} = \frac{1}{\xi_j} \frac{\int_0^1 \frac{x_t(i)}{\varepsilon_t(i)} \mathbf{1}_{\xi_t(i)=\xi_j} di}{\int_0^1 \mathbf{1}_{\xi_t(i)=\xi_j} di} = \frac{1}{\tau_j \xi_j} \int_0^1 \frac{x_t(i)}{\varepsilon_t(i)} \mathbf{1}_{\xi_t(i)=\xi_j} di, \tag{91}$$

we then have

$$M_t = \sum \tau_j M_{j,t}. \tag{92}$$

Using the decision rule of $x_t(i)$, we then have

$$M_{j,t} = \frac{1}{\xi_j} Z_t \left(\frac{\sigma-1}{\sigma}\right)^{\sigma} \Phi_t^{-\sigma} \xi_j^{\sigma} \left[\int_{\varepsilon<\varepsilon^*_{j,t}} \varepsilon^{\sigma-1} dF(\varepsilon) + \frac{\sigma}{\sigma-1}\int_{\varepsilon\geq\varepsilon^*_{j,t}} \varepsilon^{\sigma-1} dF(\varepsilon)\right] - \frac{1}{\xi_j}(1-\delta_s) S_{j,t-1} \int_{\varepsilon<\varepsilon^*_{j,t}} \frac{dF(\varepsilon)}{\varepsilon}. \tag{93}$$

Therefore, we have finished the aggregation problem.

General Equilibrium

To close the model, we add a representative household. The household's problem is identical to that in the benchmark model and the household's budget constraint can be shown to be the same as before: $C_t + K_{t+1} - (1-\delta_k)K_t = Z_t = \frac{Y_t}{P}$. Define $R(\varepsilon^*_{j,t}) = \int \max\left\{\frac{1}{\varepsilon_{t+1}}, \frac{1}{\varepsilon^*_j}\right\} dF$, $G(\varepsilon^*_{j,t}) = \left[\int_{\varepsilon<\varepsilon^*_{j,t}} \varepsilon^{\sigma-1} dF(\varepsilon) + \frac{\sigma}{\sigma-1}\int_{\varepsilon\geq\varepsilon^*_{j,t}} \varepsilon^{\sigma-1} dF(\varepsilon)\right]$, $Q(\varepsilon^*_{j,t}) = \int_{\varepsilon<\varepsilon^*_{j,t}} \frac{1}{\varepsilon} dF(\varepsilon)$, and $H(\varepsilon^*_{j,t}) = \int_{\varepsilon\geq\varepsilon^*_{j,t}} \varepsilon^{\sigma} dF(\varepsilon)$. The aggregate dynamics of the model are thus characterized by the following set of nonlinear equations that solve for $\{C_t, M_t, N_t, K_t, Y_t, X_t, \{\varepsilon^*_{j,t}\}_{j=1}^J, \{S_{jt}\}_{j=1}^J, \{M_{jt}\}_{j=1}^J\}$:

$$\frac{1}{\xi_j \varepsilon^*_{jt}} = \beta(1-\delta_s) E_t \frac{\Lambda_{t+1}}{\Lambda_t}\left[\frac{\rho_\xi}{\xi_j} R\left(\varepsilon^*_{j,t+1}\right) + \sum_{\ell} \frac{(1-\rho_\xi)}{\xi_\ell} \tau_\ell R\left(\varepsilon_{\ell,t+1}\right)\right] \tag{94}$$

$$M_{j,t} = \frac{G(\varepsilon^*_{j,t})}{\left[\int \varepsilon^{\sigma} dF\right]\left[\sum \tau_\ell \xi_\ell^{\sigma}\right]} \xi_j^{\sigma-1} Y_t - (1-\delta_s)\frac{S_{j,t-1}}{\xi_j} Q(\varepsilon^*_{j,t}) \tag{95}$$

$$S_{j,t} = \rho_\xi \frac{Y_t}{\left[\int \varepsilon^{\sigma} dF\right]\left[\sum \tau_\ell \xi_\ell^{\sigma}\right]} \frac{\xi_j^{\sigma}}{\sigma-1} H(\varepsilon^*_{j,t}) + \rho_\xi (1-\delta_s)[1-F(\varepsilon^*_{j,t})] S_{j,t-1} + \left(1-\rho_\xi\right) S_t \tag{96}$$

$$Y_t + \sum \tau_j S_{j,t} = (1-\delta_s)\sum \tau_j S_{jt-1} + X_t. \tag{97}$$

$$M_t = \sum \tau_j M_{jt} \tag{98}$$

$$a N_t^{\gamma} = (1-\alpha)\,\Phi \frac{M_t}{N_t} C_t^{-\theta} \tag{99}$$

$$C_t^{-\theta} = \beta E_t \left(\alpha \Phi \frac{M_{t+1}}{K_{t+1}} + 1 - \delta_k\right) C_{t+1}^{-\theta}. \tag{100}$$

$$C_t + K_{t+1} - (1-\delta_k) K_t = \frac{1}{P} Y_t, \tag{101}$$

$$M_t = A_t K_t^{\alpha} N_t^{1-\alpha}. \tag{102}$$

In the steady state, we have

$$\frac{1}{\xi_j \varepsilon_j^*} = \beta\left(1-\delta_s\right)\left[\frac{\rho_\xi}{\xi_j} R(\varepsilon_j^*) + \sum_{\ell} \frac{(1-\rho_\xi)}{\xi_\ell} \tau_\ell R(\varepsilon_\ell^*)\right]. \tag{103}$$

Assuming Pareto distribution for ε, we have

$$R(\varepsilon_j^*) = \frac{\eta}{\eta+1}[1-\left(\varepsilon_j^*\right)^{-\eta-1}] + \left(\varepsilon_j^*\right)^{-\eta-1} = \frac{\eta + \left(\varepsilon_j^*\right)^{-\eta-1}}{\eta+1}. \tag{104}$$

Further, assuming $J = 2$ and that ξ_j takes two values, $\{\xi_h, \xi_l\}$. The two cutoff $\{\varepsilon_h^*, \varepsilon_l^*\}$ can then be solved jointly from the following two equations:

$$\frac{1}{\xi_h \varepsilon_h^*} \frac{1}{\beta\left(1-\delta_s\right)} = \frac{\eta + (\varepsilon_h^*)^{-\eta-1}}{\eta+1} \frac{\rho_\xi + (1-\rho_\xi)\tau_h}{\xi_h} + \frac{\eta + (\varepsilon_l^*)^{-\eta-1}}{\eta+1} \frac{(1-\rho_\xi)\tau_l}{\xi_l} \tag{105}$$

$$\frac{1}{\xi_l \varepsilon_l^*} \frac{1}{\beta\left(1-\delta_s\right)} = \frac{\eta + (\varepsilon_h^*)^{-\eta-1}}{\eta+1} \frac{(1-\rho_\xi)\tau_h}{\xi_h} + \frac{\eta + (\varepsilon_l^*)^{-\eta-1}}{\eta+1} \frac{(1-\rho_\xi)\tau_l + \rho_\xi}{\xi_l}. \tag{106}$$