The London School of Economics and Political Science

Essays on Capacity Underutilization and Demand Driven Business Cycles

Tiancheng Sun

A thesis submitted to the Department of Economics
for the degree of Doctor of Philosophy

Jun 19th, 2020

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of roughly 43,000 words.

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Wouter den Haan, for his invaluable guidance and mentoring. My work and PhD life benefited a lot from his continuous support. I am also very grateful to my mentors John Moore and Ricardo Reis for their insightful advice and generous help.

I appreciate helpful comments and inspirations from participants at LSE Macro work-in-progress seminars and Edinburgh research seminar. Particularly, I have benefited a lot from fruitful conversations with David Baqaee, Petr Sedlacek, Nobuhiro Kiyotaki, Mike Elsby, Axel Gottfries, and Andrew Clausen.

My thank also goes to my LSE friends, especially Chao He, Yu Yi, and Hong Wang, for our discussions and mutual support during my PhD life.

Last but by no means least, I thank my parents for their lifelong love, and I thank my fiancée Lei Liu, who inspired me a lot and gave me generous support in my PhD adventure.

Abstract

In Chapter 1, I build a macroeconomic model that features chronic excess capacity. In my model, if one firm expands its capacity while other firms do not, it "steals" profitable demand from others. This capacity competition externality can cause an over-investment in capacity. I show that with the existence of chronic excess capacity, capital resources can be slack and consumption demand shocks can generate realistic business cycles. If consumption demand increases, more capacity will be utilized, heating up the capacity competition: firms invest with haste until the capacity utilization rate falls back to normal. If consumption demand decreases, more capacity will be left idle, cooling down the capacity competition: firms dis-invest with haste until the capacity utilization rate goes back to normal.

In Chapter 2, I show that the above results cannot be obtained in models with efficient utilization of capital or capacity. In these models, there is no capacity competition externality. None of these models could feature chronic excess capacity nor capital resource slackness. Thus, the response of output to demand shocks is limited and it is difficult to obtain demand driven business cycles in these models.

In Chapter 3, I study what kind of goods market structure features the capacity competition externality that can cause chronic excess capacity. The following assumptions are identified. First, if a firm expands its capacity while other firms do not, it can "steal" demand from others. Second, firms can charge a sufficiently high price to make a positive net profit. These two assumptions imply a negative capacity competition externality and are sufficient to cause long-term capacity underutilization at the firm-level. Third, if the invested capital has no positive externality that can potentially offset the negative externality, the capacity competition externality will be dominant and the economy will exhibit chronic excess capacity. I present several different ways to micro-found this kind of goods market structure, demonstrating the generality of the results obtained in the previous chapters.

Contents

List of Figures

List of Tables

Chapter 1

Excess Capacity and Demand Driven Business Cycles

1.1 Introduction

Capacity utilization varies substantially over the business cycles and capacity on average is never fully utilized. Figure 1.1 shows the capacity utilization rate in the U.S. published by the Federal Reserve Board. During a recession, the capacity utilization rate can be as low as 65%. During a boom, it is still likely to be less than 90%. At the micro-level, the Quarterly Survey of Plant Capacity Utilization (QSPC) shows that a large share of plants in the U.S. operate below capacity (Boehm et al., 2017). Similar results are found in the firm-level data for Switzerland (Köberl and Lein, 2011). Like unemployment, the underutilization of capacity suggests that there are always some slack resources in the economy and more so during recessions.

The vast majority of firms covered in the QSPC cites insufficient demand as the main reason for capacity underutilization (Boehm et al., 2017). Intuitively, when demand goes up, more capacity is utilized; when demand goes down, more capacity is left idle. The existence of resource slackness and the observation that demand is important for capacity utilization raise the possibility that demand is important for business cycles.

Recently, demand driven business cycles have received renewed interests, reviving the ideas put forward strongly by Keynes. The standard New Keynesian (NK) literature makes monetary demand shocks matter by assuming sticky prices. This paper shows that real demand shocks can drive business cycles in an economy where prices are perfectly flexible

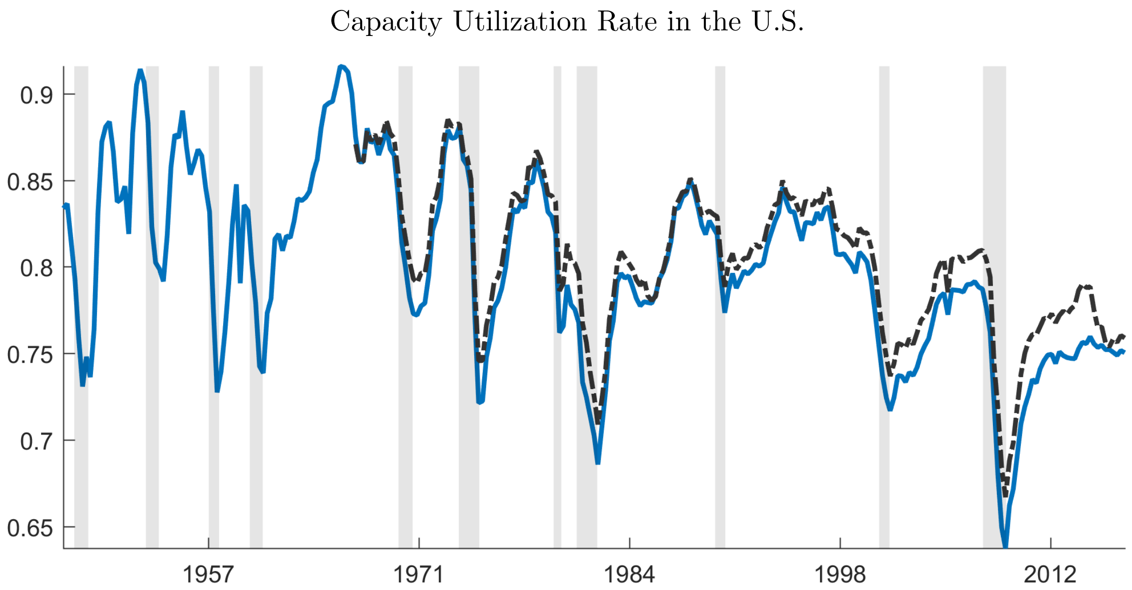

Figure 1.1: The data is from the Board of Governance of the Federal Reserve System. The solid line is for manufacturing. The dash-dotted line is for total industries. Shaded areas indicate the National Bureau of Economic Research (NBER) dated recessions.

but capacity is generally in excess. In my model, firms invest in capacity to compete for buyers. Long term capacity underutilization is an equilibrium result of this capacity competition. Furthermore, when capacity is not fully utilized, capital resources are slack and the real marginal cost curve is flat. As a result, output can be highly responsive to real demand shocks.

The model has two main assumptions that deviate from the standard real business cycle (RBC) model. First, I assume that the production technology is Leontief. Thus, firms can produce with a constant real marginal cost until they are constrained by capacity. Second, I assume that when buyers search for capacity to satisfy their demand, they need to process some price information and are subject to an information processing cost as in the rational inattention literature (e.g., Mattsson and Weibull, 2002 and Matějka and McKay, 2015). If the unit cost of processing information is zero, buyers will conduct a directed search and purchase only the cheapest goods. In this case, the goods market becomes perfectly competitive. If the unit cost of processing information is infinite, buyers will not process any price information, but conduct an undirected search for capacity. In this case, firms that have a larger capacity are more likely to be visited. In general, the unit cost of processing information is positive but finite and the behavior of the buyers is somewhere between the directed search and the undirected search. Thus, firms that charge a lower price and have a larger capacity are more likely to be visited.

Although all goods are perfect substitutes, firms can charge a positive net markup as in the standard Dixit-Stiglitz market structure because buyers are not fully attentive to prices. Furthermore, as a result of the rational inattention, the behavior of buyers is somewhat undirected. Hence, the demand allocated to each firm is not only a function of the relative price but also a function of the relative capacity. If a firm expands its capacity relative to that of other firms, it "steals" demand from its competitors. Each unit of demand stolen is profitable as firms are able to charge a positive net markup. Thus, capacity expansion in one firm has a negative externality on others. There is a great pressure on firms to expand capacity: no one wants to be left behind. Chronic excess capacity is an equilibrium result of this capacity competition externality.

The inclusion of chronic excess capacity is important for the model dynamics. As output has not yet reached the capacity limit, capital resources are slack and the real marginal cost curve is flat locally around the steady state according to the assumed Leontief technology. Hence, output can be highly responsive to demand shocks and the real wage rate is acyclical.

This contrasts sharply with the standard RBC model, in which capital resources are tight. When demand rises, the increase in output is limited because capital as a production factor is scarce, and the real wage rate falls as the marginal productivity of labor falls. Similarly, when demand drops, the decrease in output is limited because any capital resource freed up is a valuable production factor that increases the marginal productivity of labor and the real wage rate.

There is a large literature that incorporates variable capital utilization into an otherwise standard RBC model. Capital may not be fully utilized because firms are subject to a convex capital utilization cost.[1] However, I show that *aggregate capacity* is still fully utilized in steady state, even though *capital* may not. Furthermore, capital resources are tight as an increase in capital can still reduce the real marginal cost by lowering the capital utilization rate. After all, capital is not utilized precisely because it is too costly to be utilized. Thus, the standard variable capital utilization model does not feature chronic excess capacity nor capital resource slackness.

As a result, when demand increases, more capital needs to be utilized and the real marginal cost rises in the standard variable capital utilization model. The upward sloping real

[1] Greenwood et al. (1988) and Basu and Kimball (1997) assume that the depreciation rate of capital is increasing and convex in terms of the utilization of capital. Kydland and Prescott (1988), Burnside et al. (1993), and Bils and Cho (1994) assume that the overtime premium paid to workers is increasing and convex in terms of the utilization of capital. Christiano et al. (2005) and Smets and Wouters (2007) assume that the resources consumed by firms to utilize capital is increasing and convex in terms of the utilization of capital.

marginal cost curve dampens the response of output to demand shocks and causes a countercyclical real wage rate. The evidence documented in the literature, however, suggests that the real wage rate should not be too countercyclical (e.g., Bils, 1985 and Solon et al., 1994).

One can try to reduce the capital resource tightness by reducing the convexity of the utilization cost function. This could make the real wage rate less countercyclical and consumption more responsive. However, the volatility of investment under consumption demand shocks does not increase.[2] The reason is the following. Under consumption demand shocks, firms want to adjust their capital precisely because capital resources are tight so that there is a pro-cyclical upward pressure on the real marginal cost that firms can adjust their capital to mitigate. Hence, removing capital resource tightness not only removes the curb on capital adjustment but also removes the impetus for capital adjustment. Consequently, if capital resources are less tight, consumption becomes more responsive but investment does not, and the relative volatility of investment to consumption falls. Thus, it is difficult to get a large relative volatility of investment to consumption without causing a strongly countercyclical real wage rate under consumption demand shocks.

By contrast, in my model, capital resources are slack locally around the steady state and the real wage rate is acyclical. Despite the capital resource slackness, when future demand is expected to increase, firms have a strong incentive to increase their capital because each unit of capacity installed is expected to attract more demand. Thus, in my model, removing capital resource tightness does not remove the impetus for capital adjustment. Hence, the response of investment to consumption demand shocks can be very large. With some capital and investment adjustment costs, the relative volatility of investment to consumption in my model can be consistent with that in the U.S. data.

To assess the importance of demand for business cycles quantitatively, I allow for the possibility that business cycles are driven both by demand shocks and by labor productivity shocks. In particular, I consider three types of real demand shocks: consumption demand shocks that change the marginal utility of consumption relative to the marginal dis-utility of labor, investment demand shocks that change the subjective discount factor, and exogenous expenditure shocks that change the government expenditure. I use Bayesian estimation techniques to estimate my capacity underutilization model.

I find that the capacity underutilization model attributes most of the business cycle fluctu-

[2] Under investment demand shocks, consumption, investment, and hours will move in different directions. This is known as the Barro-King curse in the literature (Barro and King, 1984).

ations to demand shocks. Labor productivity shocks account for only 2% of the variation in output and 13% of the variation in hours. Among the three types of demand shocks, consumption demand shocks are dominant. Consumption demand shocks alone explain more than 72% of the variance in consumption, 78% of the variance in investment, and 60% of the variance in hours, and generate the correct business-cycle co-movement among consumption, investment, hours, and the Solow residual. The results resonate with two recent empirical papers which suggest that business cycles are mainly driven by a single type of demand shock (e.g., Andrle et al., 2017 and Angeletos et al., 2018).

Summing up, this paper has two major contributions. First, by extending the standard neoclassical framework with rationally inattentive buyers who tend to search for capacity undirectly, I present a new model that can explain long-term capacity underutilization as observed in the data. Second, when viewed through the lens of the model, real demand shocks have the potential to be the main driving force of business cycles, and when the model is estimated to match the U.S. macro data, that turns out to be the case.

The rest of the paper is organized as follows. The next subsection discusses the related literature. Section 1.2 establishes a basic capacity underutilization (CU) model. Section 1.3 studies the properties of the basic CU model. Section 1.4 compares the basic CU model with a standard variable capital utilization model. Section 1.5 extends the basic CU model to a full CU model, estimates the full CU model using Bayesian estimation techniques, and discusses the quantitative results. Section 1.6 concludes.

Related Literature

Cooley et al. (1995) and Gilchrist and Williams (2000) assume that plants are subject to idiosyncratic productivity shocks. Low productivity plants will be left idle to save labor cost. Fagnart et al. (1997) and Fagnart et al. (1999) assume that firms are subject to idiosyncratic demand shocks. Extra capacity will be held by firms as a precaution to save the cost in case demand is high. Bai et al. (2012) incorporate search and matching frictions into the goods market. They use a competitive search framework as in Moen (1997). The capacity utilization rate measures the tightness of the goods market. Having some capacity underutilized saves the cost for buyers to purchase goods from a tight market. In all these models, however, plant or capacity is underutilized for a cost-saving reason. Hence, there is no capacity competition externality. From an aggregate perspective, capacity is not in excess, capital resources are tight, and the dynamic properties of these models are similar to those of the standard variable capital utilization model (see Chapter 2 for further discussions).

In addition, the precautionary capacity model developed by Fagnart et al. (1997) has a difficult time to match the micro-level data on capacity utilization. In their model, the idiosyncratic demand shocks are so large that the capacity constraint is frequently binding and a large proportion of firms are running at full capacity each period. Otherwise, the extra capacity held by firms as a precaution would be very small. The standard deviation of the year-over-year (YoY) quarterly sales growth rate at the firm-level needs to be about 60% to generate a capacity utilization rate of 87% on average, and each quarter 47% of firms are running at full capacity. In the real world, however, the average firm-level volatility of the YoY quarterly sales growth rate is about 15% to 25% (e.g., Comin and Philippon, 2005, Buch et al., 2009, Kelly et al., 2013, and De Veirman and Levin, 2018); and roughly less than 20% of the firms or plants surveyed report running at full capacity each quarter (e.g., Köberl and Lein, 2011 and Boehm et al., 2017). My capacity underutilization model has a better chance to fit the micro-level data. Because of a capacity competition pressure, all firms in my model are willing to hold extra capacity even when there is no demand uncertainty at all.

Michaillat and Saez (2015) also incorporate search and matching frictions into the goods market to explain capacity underutilization. They use a random matching framework. Hence, firms can use capacity to steal profitable demand as in my model. Their paper, however, focuses on monetary demand shocks in an economy with nominal rigidity. Since prices are negotiated after matching takes place, Michaillat and Saez (2015) have the freedom to choose the bargaining protocol and they choose it to be such that all prices are fixed. By contrast, my paper focuses on real demand shocks in an economy with perfectly flexible prices which are determined by profit maximizing firms before matching takes place.

The literature on industrial organization documents the possibility of established firms holding excess capacity to deter entry (e.g., Spence, 1977, Dixit, 1980, and Bulow et al., 1985). Established firms invest in capacity to protect demand from being stolen by potential entrants. In my model, firms invest in capacity to steal demand from others. Both mechanisms allow capacity to have a positive effect on demand. Because of the complex strategic interactions involved, the entry deterrence mechanism is rarely incorporated into macroeconomic models. However, the capacity competition mechanism proposed by this paper can be easily incorporated into the existing macroeconomic framework.

1.2 A Basic Model of Capacity Underutilization

This section presents a basic capacity underutilization (CU) model. The model has two features. First, the production technology is Leontief. Second, buyers have a limited capability to process price information when they search for capacity to satisfy their demand.

1.2.1 Technology

There is a unit mass of identical firms indexed by $j \in [0,1]$. All goods produced are perfect substitutes and can be used either as consumption or investment. At the beginning of each period, each firm has some capital stock inherited from the last period. The law of motion for capital is standard:

$$k_{j,t+1} = k_{j,t}\,(1-\delta) + i_{j,t}, \tag{1.1}$$

where $i_{j,t}$ is the investment made by firm j at time t, $k_{j,t}$ is the capital stock of firm j at the beginning of the period, and parameter $\delta \in (0,1]$ is the depreciation rate.

The production technology is assumed to be Leontief:

$$y_{j,t} = \min\left\{\frac{l_{j,t}}{\alpha_v}, Ak_{j,t}\right\}, \tag{1.2}$$

where $y_{j,t}$ is the amount of goods produced by firm j, $l_{j,t}$ is the variable labor hired by firm j, parameter $\alpha_v > 0$ is the variable labor required per unit of output, and parameter $A > 0$ is the productivity of capital.

Capacity is defined as the output level at which the short run average total cost (SRAC) curve is tangent to the long run average total cost (LRAC) curve (e.g., Morrison, 1985). Along the LRAC curve, one minimizes its average total cost by adjusting both the variable factors, such as labor, and the short run fixed factors, such as capital. Along the SRAC curve, one minimizes the same average total cost, adjusting only the variable factors. At the point of tangency, where full capacity utilization is achieved, the average total cost could not be further minimized even when short run fixed factors can be adjusted. Capacity defined in this way is also known as the economic measure of capacity in the literature (e.g., Nelson, 1989).

In practice, capacity is considered as the maximum level of output that a firm can produce within a given period of time under a *realistic* working schedule, taking into account normal downtime (e.g., Corrado and Mattey, 1997). If the production technology is of constant returns to scale, the LRAC curve is flat. In this case, the economic measure of capacity is the output level at which the SRAC curve achieves its minimum, capturing partly the practical notion that extraordinary efforts are required to produce beyond capacity. Eiteman and Guthrie (1952) conduct a survey and find that the short run average total cost curve is typically downward sloping until a point near or at capacity. This suggests that the practical notion of capacity is roughly consistent with the theoretical definition of capacity when production technology exhibits constant returns to scale.

With a Leontief technology, the SRAC curve of firm j is given by:

$$\mathbf{SRAC}\left(y_{j,t}, k_{j,t}; w_t\right) = w_t \alpha_v + (r + \delta) \frac{k_{j,t}}{y_{j,t}}, \tag{1.3}$$

where w_t is the real wage rate, $r > 0$ is the real interest rate in steady state, and $y_{j,t} \in (0, Ak_{j,t}]$ must be positive but no larger than the production limit. Note that the SRAC curve is downward sloping in output until it reaches the production limit.

The LRAC curve is the minimum of the SRAC when capital can be adjusted:

$$\mathbf{LRAC}\left(w_t\right) = w_t \alpha_v + \frac{r + \delta}{A}, \tag{1.4}$$

which is flat as the Leontief technology exhibits constant returns to scale.

Capacity $\bar{y}_{j,t}$ is the output level at which the SRAC curve reaches its minimum and is tangent to the flat LRAC curve. We have

$$\bar{y}_{j,t} = Ak_{j,t}. \tag{1.5}$$

which shows that capacity is simply the maximum output that can be produced within a given period of time. Because the amount of capital stock is predetermined, capacity may not be fully utilized when demand is not high enough.

I find that it is both theoretically appealing and empirically plausible to start with a simple Leontief technology. Theoretically, the concept of capacity is naturally clear with Leontief technology, consistent with both the theoretical definition and the practical notion. Thus, people that explicitly model capacity utilization often assume that firms produces with some Leontief technology at least in the short run (see, e.g., Fagnart et al., 1997, Fagnart et al., 1999, and Boehm et al., 2017). This is also a standard assumption in the management science and operations research literature.[3]

[3]For some papers in the management science and operations research literature on capacity utilization,

Empirically, Leontief technology is not uncommon. For example, drivers and cars are perfect complements in a taxi company; cooks and cookers are perfect complements in a restaurant. At the micro-level, firms are likely to operate with fixed input-output coefficients especially in the short run because these coefficients are much dictated by the technologies embodied in capital and are carefully designed by modern engineers (Eiteman, 1947). The observation that input-output coefficients are fixed in the short run also motivates the putty-clay technology introduced by Johansen (1959). Indeed, numerous empirical evidence based on accounting, engineering, or questionnaire data suggest that marginal cost at the micro-level is typically constant at least up until some point close to capacity (see Walters, 1963, for a literature survey).

1.2.2 Buyers and the Goods Market Structure

The purchasing process takes two steps. First, a household or a firm decides how many goods shall be consumed or invested in period t based on the aggregate price P_t. Second, the household or the firm sends out a buyer to purchase the goods. All buyers are identical. If a buyer chooses to purchase from firm j, the payoff for the buyer is a strictly decreasing function of the real price ${}^{P_{j,t}}/{}_{P_t}$ charged by the firm:[4]

$$v_t(j) = -\ln\left(\frac{P_{j,t}}{P_t}\right). \tag{1.6}$$

In principle, buyers want to purchase only the cheapest goods that yields the highest payoff. However, as in Mattsson and Weibull (2002), I assume that it is costly for buyers to implement the desired outcome because they need to process some price information in order to direct their actions towards the desired.

Without exerting any information processing effort, buyers can only purchase blindly and randomly. The purchasing behavior in this case is called the *default* purchasing behavior, which is the most inattentive behavior of buyers. I assume that the default purchasing behavior is an *undirected search* for capacity, a behavior that results in a random matching

see Florian and Klein (1971), Kalish (1983), and Deng and Yano (2006). For some textbooks on operations management, see Stevenson (2002), Kumar and Suresh (2006), and Gupta and Starr (2014).

[4]It will become clear later that the logarithm functional form helps generate a demand curve of constant elasticity. However, the results of this paper will not be affected as long as the payoff function is strictly decreasing in the real price of the goods purchased (see section 3.3 of Chapter 3). In addition, although buyers purchase on behalf of households and firms, the buyers have their own payoffs. This assumption is convenient as it greatly simplifies the aggregation problem by allowing buyers to be homogeneous across potentially different households and firms. In section 3.3 of Chapter 3, I show that a homogeneous purchasing problem can be obtained directly as a sub-problem of households and firms, if the information processing cost is proportional to the amount of goods purchased and is paid in terms of goods or services. In this case, the payoff of purchasing a unit of goods from firm j decreases linearly rather than log-linearly in the real price charged by the firm. However, the results of this paper remain unchanged.

between the demand from buyers and the capacity supplied by firms.

The rationale behind the above assumption is as follows. Each matching process can be described by a probability density function f_t. $f_t(x)$ gives the likelihood that a unit of capacity $x \in [0, \bar{y}_t]$ is matched with a unit of demand, where $\bar{y}_t \equiv \int_0^1 \bar{y}_{j,t} dj$ is the total amount of capacity supplied by firms. If a buyer pays no attention to prices, the matching between the demand from the buyer and the capacity supplied by firms should be the *most disordered*. The disorder of the matching process can be measured by the entropy of f_t, which is given by

$$-\int_0^{\bar{y}_t} \ln f_t(x) f_t(x) dx. \tag{1.7}$$

For all distributions with a support limited to the interval $[0, \bar{y}_t]$, the maximum entropy distribution is the uniform distribution: $f_t(x) = 1/\bar{y}_t$. Hence, the most disordered matching process is that each unit of capacity has an equal probability to be matched.

Thus, without exerting any information processing effort, the probability density that a buyer purchases from firm j is proportional to the capacity of the firm:

$$n_t^*(j) \equiv \frac{\bar{y}_{j,t}}{\bar{y}_t}, \tag{1.8}$$

and n_t^* is called the default probability density function as it describes the default purchasing behavior.

If a buyer wants to deviate from the default n_t^*, she has to process some price information and incurs some information processing cost. Let n_t be the probability density function eventually obtained by a buyer. $n_t(j)$ gives the likelihood that the buyer purchases from firm j. As in the rational inattention literature, I assume that the information processing cost is proportional to the amount of information processed measured by the relative entropy of n_t with respect to n_t^*. Relative entropy is also known as Kullback–Leibler (KL) divergence, which is non-negative, convex, and obtains its minimum value zero if $n_t = n_t^*$:

$$D_{KL}(n_t || n_t^*) \equiv \int_0^1 n_t(j) \ln\left(\frac{n_t(j)}{n_t^*(j)}\right) dj. \tag{1.9}$$

Intuitively, the more different n_t is from n_t^*, the more information is needed to be processed.

The buyer's problem is to maximize the expected payoff net of the information processing cost:

$$\max_{n_t \geq 0} \int_0^1 v_t(j) n_t(j) dj - \Lambda \left(\int_0^1 n_t(j) \ln\left(\frac{n_t(j)}{n_t^*(j)}\right) dj\right), \tag{1.10}$$

subject to

$$\int_0^1 n_t(j)\,dj = 1, \tag{1.11}$$

where $\Lambda > 0$ is the unit cost of processing information.

Note that n_t^* is exogenous to the buyer's problem because the default purchasing behavior is *not* a rational choice of the buyer.

Mattsson and Weibull (2002) show that the unique solution to the above problem is:

$$\forall j \in [0,1]: \; n_t(j) = \frac{n_t^*(j) e^{\frac{v_t(j)}{\Lambda}}}{\int_0^1 n_t^*(j) e^{\frac{v_t(j)}{\Lambda}} dj}. \tag{1.12}$$

Equation (1.12) says that the optimal probability of purchasing from firm j is proportional to the default probability $n_t^*(j)$ and moderated by the payoff of purchasing from firm j. The higher the payoff is, the higher the probability that firm j would be chosen by the buyer. We can substitute the buyer's payoff function (1.6) for $v_t(j)$ in equation (1.12) and obtain:

$$\forall j \in [0,1]: \; n_t(j) = \frac{n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda}}}{\int_0^1 n_t^*(j) P_{j,t}^{-\frac{1}{\Lambda}} dj}. \tag{1.13}$$

If all prices are equal, buyers will be indifferent between choosing any two firms and have no incentive to do any costly information processing. In this case, buyers will simply follow the undirected search for capacity and will be distributed across firms according to the default probability density function n_t^*.

If prices are different, firms that charge a relatively low price can attract additional buyers. However, because of a positive information processing cost, not all buyers will purchase the goods of the lowest price. Even the goods of the highest price will still be purchased by some buyers, as we can see from equation (1.13) that $n_t(j)$ is always positive as long as $P_{j,t}$ remains finite. Intuitively, since buyers have a limited capability of processing price information, they rationally choose to be partially inattentive to prices rather than to purchase always the cheapest goods. Hence, buyers allow themselves to make some "mistakes" with positive probabilities in order to save the information processing cost.

Since all buyers are identical, demand aggregation is a trivial task. The demand for the goods produced by firm j is the multiplication of the total demand y_t from buyers and the share of buyers that purchase from firm j: $y_{j,t} = n_t(j)\,y_t$. Substitute the buyer's solution

(1.13) for $n_t(j)$; we have

$$y_{j,t} = \frac{n_t^*(j) P_{j,t}^{-\varepsilon}}{\int_0^1 n_t^*(j) P_{j,t}^{-\varepsilon} dj} y_t = \frac{k_{j,t} P_{j,t}^{-\varepsilon}}{\int_0^1 k_{j,t} P_{j,t}^{-\varepsilon} dj} y_t, \tag{1.14}$$

where $\varepsilon \equiv \Lambda^{-1}$ is the price elasticity of demand. The second equality comes from equations (1.5) and (1.8). The former says that capacity is proportional to capital. The latter says that the default probability of purchasing from firm j is proportional to the capacity of the firm.

The aggregate price P_t must satisfy the following aggregation condition:

$$P_t y_t = \int_0^1 P_{j,t} y_{j,t} dj, \tag{1.15}$$

which ensures that the money spent on aggregate goods is equal to the money earned by firms. Since all goods are perfect substitutes, P_t is simply the average price of the goods purchased by buyers: $P_t = \int_0^1 P_{j,t} n_t(j)\, dj$.

With a positive information processing cost, buyers act *as if* the same goods produced by different firms are imperfect substitutes.[5] The lower the information processing cost is, the more competitive the market is. In one extreme where $\Lambda \to 0$ and $\varepsilon \to \infty$, buyers are fully attentive to prices, the search process is fully directed, and the goods market is perfectly competitive. In the other extreme where $\Lambda \to \infty$ and $\varepsilon \to 0$, buyers pay no attention to prices, the search process is fully undirected, and the demand that goes to each firm no longer depends on relative prices. Hence, the model spans both directed search and undirected search as well as the intermediate cases where the search is partially directed. The parameter Λ controls the degree to which the search is directed.

In addition to prices, the default purchasing behavior n_t^* plays an important role in determining the relative size of demand. n_t^* is endogenously affected by the relative capacity of the firm (see equation (1.8)). Because of this feature, a firm not only can lower its relative price but also can expand its relative capacity to compete for buyers. For a given amount of total demand from buyers, if a firm expands its capacity while other firms do not, it "steals" demand from others. Thus, in addition to the usual price competition, there is a capacity competition among firms.

For example, Starbucks can expand its market share by opening more brick-and-mortar coffee stores than its competitor Costa. A printing store can expand its market share

[5] This may explain why in reality firms that sell near homogeneous goods, such as oil, steel, and sugar, can still charge a markup to compensate for their fixed costs.

by installing more printing machines than other printing stores. Intuitively, firms with a larger capacity are more likely to be visited by buyers who are not fully attentive to prices and thus search for capacity in a somewhat undirected way. In the standard Dixit-Stiglitz monopolistic competition model, however, capacity has no effect on the allocated demand. Therefore, the demand curve in my model is fundamentally different from that in the standard Dixit-Stiglitz setup, though they do share some similarities.

1.2.3 Households

There is a unit mass of identical households. Consider a representative household who maximizes her expected lifetime utility

$$\max_{\{c_t,l_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u\left(z_{c,t}, c_t, l_t\right), \tag{1.16}$$

subject to the budget constraint of the household

$$c_t = w_t l_t + d_t, \tag{1.17}$$

where $\beta \in [0, 1)$ is the subjective discount factor, $z_{c,t}$ is a preference parameter that varies with time, c_t is the amount of consumption goods that the household asks her buyer to purchase, l_t is the labor supply, w_t is the real wage rate, and d_t is the amount of dividends received from firms. I assume that the labor market is perfectly competitive. Hence, the household takes the real wage rate w_t as exogenous.

The functional form of the one-period utility is assumed to be

$$u\left(z_{c,t}, c_t, l_t\right) = \begin{cases} \phi e^{z_{c,t}} \frac{\left(c_t/\left(\phi e^{z_{c,t}}\right)\right)^{1-\gamma}-1}{1-\gamma} - \bar{\omega} l_t, & \gamma \neq 1 \\ \phi e^{z_{c,t}} \ln\left(c_t/(\phi e^{z_{c,t}})\right) - \bar{\omega} l_t, & \gamma = 1 \end{cases} \tag{1.18}$$

where $\phi > 0$ is a scaling parameter, $\gamma^{-1} > 0$ is the elasticity of inter-temporal substitution, and $\bar{\omega} > 0$ is the marginal dis-utility of labor. The assumption that the marginal dis-utility of labor is a constant follows from the indivisible labor theory proposed by Hansen (1985) and Rogerson (1988).

The preference parameter $z_{c,t}$ is interpreted as consumption demand. An increase in $z_{c,t}$ is an increase in the marginal utility of consumption relative to the marginal dis-utility of labor. I assume that $z_{c,t}$ follows an AR(1) process $z_{c,t} = \rho_c z_{c,t-1} + e_{c,t}$, where $\rho_c \in [0, 1)$ is a persistence parameter and $e_{c,t}$ is a shock to the consumption demand.

The first order conditions (FOCs) of the household's problem are

$$\bar{\omega} = \lambda_t w_t, \tag{1.19}$$

$$\lambda_t = \left(\frac{c_t}{\phi e^{z_{c,t}}}\right)^{-\gamma}, \tag{1.20}$$

where λ_t is the Lagrangian multiplier for the household's budget constraint (1.17) and can be interpreted as the shadow price of goods measured in terms of utils, or simply, the price in utils. The optimal labor supply condition (1.19) shows that the Frisch elasticity of labor supply is infinite. However, an interior solution is still possible because an increase in the supplied labor relaxes the budget constraint of the households (1.17) and causes a decrease in the marginal utility of consumption.

1.2.4 Firms

Each firm aims to maximize its firm value, which is the present value of the firm's dividend flows $d_{j,t}$:

$$\max_{\{P_{j,t},d_{j,t},y_{j,t},i_{j,t},k_{j,t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} d_{j,t}\right), \tag{1.21}$$

subject to the demand curve (1.14), the resource constraint of the firm

$$d_{j,t} + i_{j,t} = \left(\frac{P_{j,t}}{P_t} - w_t \alpha_v\right) y_{j,t}, \tag{1.22}$$

the capacity constraint

$$y_{j,t} \leq A k_{j,t}, \tag{1.23}$$

and the law of motion for capital (1.1). Because firms are owned by households, the stochastic discount factor that discounts the value at time $t+1$ to time t is given by $\beta \lambda_{t+1}/\lambda_t$ and the real interest rate in steady state is $r = \beta^{-1} - 1$.

To prevent firms from charging an infinitely high price, the unit cost of processing information Λ should not be too large. I assume that Λ is less than one; thus, the price elasticity of demand is larger than one: $\varepsilon = \Lambda^{-1} > 1$.

The FOCs of the firm's problem are

$$\frac{P_{j,t}}{P_t} = \frac{\varepsilon}{\varepsilon - 1} \left(w_t \alpha_v + \mu_{j,t}\right), \tag{1.24}$$

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(\frac{P_{j,t+1}}{P_{t+1}} \frac{1}{\varepsilon} A u_{j,t+1} + A \mu_{j,t+1} + 1 - \delta \right) \right), \tag{1.25}$$

where $\mu_{j,t} \geq 0$ is the Lagrangian multiplier of the capacity constraint (1.23) and $u_{j,t+1} \equiv {}^{y_{j,t+1}}/_{\bar{y}_{j,t+1}} = {}^{y_{j,t+1}}/_{Ak_{j,t+1}}$ is the capacity utilization rate of the firm.

Equation (1.24) says that the firm sets its price according to a constant markup rule. If the firm operates below its capacity limit, the price charged by the firm will be proportional to the firm's marginal cost: $w_t \alpha_v$. Once the demand hits the firm's capacity limit, the price will be raised up so as to equate the demand $y_{j,t}$ to the capacity: $y_{j,t} = \bar{y}_{j,t} = Ak_{j,t}$.

Equation (1.25) shows that there are two reasons for a firm to invest in capital. First, capital investment relaxes the capacity constraint. This value is captured by the term $A\mu_{j,t+1}$, which is the shadow value of relaxing the capacity constraint. Second, capital investment makes the goods produced by the firm more likely to be purchased by buyers. This value is captured by the multiplication of the demand attracted per unit of capital $Au_{j,t+1}$ and the profit contributed by each unit of demand attracted $\left({}^{P_{j,t+1}}/_{P_{t+1}}\right) \varepsilon^{-1}$.

1.2.5 Symmetric Equilibrium

In a symmetric equilibrium, if a variable is of the form $x_{j,t}$, we have $x_{i,t} = x_{j,t}$ for all $i, j \in [0, 1]$. After obtaining the FOCs, we can omit the subscripts that index a variable to a particular firm. The symmetric equilibrium is a stable stochastic process of nine variables (c_t, λ_t, i_t, w_t, μ_t, l_t, u_t, k_{t+1}, and y_t) that satisfies the household's FOCs (1.20)-(1.19), the firm's pricing condition

$$1 = \frac{\varepsilon}{\varepsilon - 1} \left(w_t \alpha_v + \mu_t \right), \tag{1.26}$$

the firm's investment condition

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(\frac{1}{\varepsilon} A u_{t+1} + A \mu_{t+1} + 1 - \delta \right) \right), \tag{1.27}$$

the measure of the capacity utilization rate

$$u_t = \frac{y_t}{Ak_t}, \tag{1.28}$$

the complementary slackness condition for the capacity constraint

$$\mu_t \left(Ak_t - y_t \right) = 0, \tag{1.29}$$

where $\mu_t \geq 0$ and $Ak_t - y_t \geq 0$, the amount of labor hired according to the Leontief production function

$$l_t = \alpha_v y_t, \tag{1.30}$$

the law of motion for capital

$$k_{t+1} = k_t (1 - \delta) + i_t, \tag{1.31}$$

and the aggregate resource constraint

$$c_t + i_t = y_t, \tag{1.32}$$

which follows from the budget constraint of the household (1.17) and the resource constraint of the firm (1.22).

1.3 Properties of the Basic Capacity Underutilization Model

In this section, I show the following properties of the basic CU model. First, the decentralized equilibrium is generally inefficient because buyers are not fully attentive to prices. Second, if the degree of inattention is large enough, capacity will be underutilized in steady state and the economy will exhibit chronic excess capacity. Third, locally around the steady state where capacity is in excess, capital resources are slack. Finally, because of the capital resource slackness, the real wage rate is acyclical and output is highly and much more responsive to demand shocks than in the standard RBC model.

1.3.1 Inefficiency and Capacity Competition Externality

Given the technology and the preference of my basic CU model, the efficient allocation can be obtained by solving a corresponding social planner's problem:

$$\max_{\{c_t, i_t, y_t, l_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u (z_{c,t}, c_t, l_t) , \tag{1.33}$$

subject to the Leontief production function (1.2), the law of motion for capital (1.31), and the aggregate resource constraint (1.32).

The solution to the social planner's problem is a stable stochastic process of eight variables (c_t, λ_t, i_t, μ_t, l_t, u_t, k_{t+1}, and y_t) that satisfies the optimal consumption condition (1.20), the optimal trade-off between the benefit and the cost of production

$$\lambda_t = \bar{\omega}\alpha_v + \lambda_t \mu_t, \tag{1.34}$$

the Euler's equation that gives the optimal investment

$$1 = \beta \mathbb{E}_t \left(\frac{\lambda_{t+1}}{\lambda_t} \left(A\mu_{t+1} + 1 - \delta \right) \right), \tag{1.35}$$

the complementary slackness condition for the capacity constraint (1.29), the required labor followed from the Leontief production function (1.30), the law of motion for capital (1.31), and the aggregate resource constraint (1.32).

The conditions that characterize the efficient allocations are the same as those that characterize the decentralized equilibrium except for conditions (1.34) and (1.35). In the decentralized equilibrium, the trade-off between the benefit and the cost of production is given by the firm's pricing condition (1.26) combined with the household's labor supply condition (1.19):

$$\lambda_t = \frac{\varepsilon}{\varepsilon - 1} \left(\bar{\omega}\alpha_v + \lambda_t \mu_t \right); \tag{1.36}$$

and the firm's investment condition is given by equation (1.27).

Hence, the equilibrium of the basic CU model is efficient if and only if $\varepsilon \to \infty$ or $\Lambda \to 0$. In this case, the basic CU model is reduced to a standard RBC model with a Leontief production technology, or a Leontief-RBC model for short, where buyers are fully attentive to prices and the goods market is perfectly competitive.

In general, as long as buyers are not fully attentive to prices ($\varepsilon = \Lambda^{-1} > 1$ or $\Lambda \in (0, 1)$), the equilibrium of the basic CU model is inefficient.

First, the rational inattention of buyers allows firms to enjoy some market power. Hence, the real wage rate is depressed and the price in utils λ_t is inflated by the markup charged by firms. As a result, consumption level tends to be inefficiently low.

Second, when buyers are not fully attentive to prices, they search for capacity in a somewhat undirected way.[6] Hence, firms can expand their relative capacity to compete for buyers. If a firm expands its capacity while other firms do not, the firm is stealing demand from others. Furthermore, each unit of demand stolen is profitable because firms

[6] In a symmetric equilibrium, since all firms charge the same price, the behavior of the buyers is completely undirected as there is no need to process any price information. This, however, does not mean that the presence of information processing cost is not important because it allows firms to charge a positive net markup and to steal demand from each other.

can take advantage of the inattentive buyers to charge a positive net markup. Thus, capacity expansion has a negative externality. From an individual firm's perspective, capacity expansion not only relaxes the capacity constraint but also steals profitable demand from its competitors (see equation (1.27)). From an aggregate perspective, however, the return of capacity is only derived from the value of relaxing the capacity constraint (see equation (1.35)). Because of this capacity competition externality, firms tend to over-invest in capacity.

The first mechanism is usual, which also shows up in the standard Dixit-Stiglitz setup. The second mechanism characterizes the basic CU model and is important for us to understand why there is long-term capacity underutilization in the economy.

1.3.2 Long-term Capacity Underutilization and Chronic Excess Capacity

Let us focus on the steady state of the basic CU model. If a variable is of the form x_t, its value in steady state is denoted by x.

If the unit cost of processing information is large enough $\Lambda \in \left(\frac{r+\delta}{A}, 1\right)$, according to the firm's investment condition (1.27), we have:

$$u = \frac{r+\delta}{A\Lambda} - \frac{\mu}{\Lambda} \leq \frac{r+\delta}{A\Lambda} < 1, \tag{1.37}$$

which says that there is capacity underutilization in steady state.

If $\Lambda \to 0$, however, the basic CU model is reduced to the Leontief-RBC model and the allocation of the economy becomes efficient. In this case, according to the optimal investment condition (1.35), we have

$$\mu = \frac{r+\delta}{A} > 0, \tag{1.38}$$

which means that capacity must be fully utilized in steady state: $y = Ak$.

The comparison between the basic CU model and the Leontief-RBC model highlights how a single modification on the assumption of the behavior of buyers can cause a substantial change.

In the Leontief-RBC model, buyers pay full attention to prices. Thus, capacity has no effect on demand. The only reason for firms to invest in capital is to relax the capacity constraint. If the capacity constraint is not binding in steady state, the marginal value of relaxing the capacity constraint is zero: $A\mu = 0$. The only value of capital disappears but holding capital is costly. Hence, the firm will reduce its capital holding until its capacity constraint is binding in steady state.

In the basic CU model, however, buyers are not fully attentive to prices, and thus, search for capacity in a somewhat undirected way. Hence, firms can expand their capacity to compete for buyers and steal demand from each other. The limited capability of buyers to process price information ($\Lambda \in (0,1)$) also provides firms with a monopolistic power, which allows firms to earn a profit for each unit of demand stolen: $\varepsilon^{-1} = \Lambda > 0$. This monopolistic profit is a lure for firms to hold extra capacity. Even though capacity is not fully utilized, as long as the monopolistic profit rate is large enough, firms will not want to reduce their capacity to lose demand to others, justifying the existence of long-term capacity underutilization in the basic CU model.

From an aggregate perspective, however, the only value of capital is to save cost by relaxing the capacity constraint. If firms hold more capacity than that could be justified by minimizing the aggregate total cost, we say that capacity is in excess.

Definition 1.1. (*Excess capacity*) Capacity is said to be in excess if and only if aggregate capacity is underutilized.

Definition 1.2. (*Aggregate capacity*) Aggregate capacity is the aggregate output level at which the aggregate short run average total cost (SRAC) curve is tangent to the aggregate long run average total cost (LRAC) curve. The aggregate SRAC is given by

$$\mathcal{SRAC}\left(y_t, k_t; w_t\right) = \frac{\mathcal{C}\left(y_t, k_t; w_t\right) + \left(r + \delta\right) k_t}{y_t}, \tag{1.39}$$

where $\mathcal{C}$ is the aggregate variable cost as a function of aggregate demand and capital. The aggregate LRAC is the minimum of the aggregate SRAC when aggregate capital can be adjusted.

In the basic CU model, the aggregate variable cost function is given by

$$\mathcal{C}\left(y_t, k_t; w_t\right) = \min_{\{l_{j,t}, y_{j,t}, k_{j,t} \geq 0\}_{j \in [0.1]}} w_t \int_0^1 l_{j,t} dj, \tag{1.40}$$

subject to the aggregation condition $y_t = \int_0^1 y_{j,t} dj$, the Leontief production technology (1.2), and the capital stock available $\int_0^1 k_{j,t} dj \leq k_t$. Since all firms are identical and all variable costs are paid by firms, the aggregate variable cost function is the same as the variable cost function at the firm-level: $\mathcal{C}\left(y, k; w\right) = w\alpha_v y$, for all $y \leq Ak$.

Thus, in a symmetric equilibrium, if capacity is underutilized at the *firm-level*, *aggregate* capacity is also underutilized. By Definition 1.1, the economy exhibits chronic excess capacity when the unit cost of processing information Λ is large enough.

In the rest of the paper, I assume that the consumption demand shocks are small and that Λ is large such that the capacity constraint (1.23) never binds.[7] With this assumption, I can ignore the occasionally binding capacity constraint to focus on the *local* dynamic properties of the basic CU model around the steady state where capacity is in excess.

1.3.3 Capital Resource Slackness

To explain how the existence of unused capacity affects the model dynamics, I introduce the concept of capital resource tightness (or slackness), which captures the scarcity of capital as a production factor.

Definition 1.3. (*Capital resource tightness*) If a marginal decrease in capital stock makes the current output level infeasible or leads to an increase in the real marginal cost (MC), I say that capital resources are *tight*. If a marginal decrease in capital stock has no effect on the real MC, I say that capital resources are *slack*.

Capital resource tightness can be measured by the negative capital elasticity of the real MC. If the production technology is of constant returns to scale, the measure of capital resource tightness is also equal to the output elasticity of the real MC, i.e., the steepness of the real MC curve. Let ζ be the measure of capital resource tightness in steady state. We have

$$\zeta \equiv -\frac{\partial \ln \mathcal{C}_y (y, k; w)}{\partial \ln k} = \frac{\partial \ln \mathcal{C}_y (y, k; w)}{\partial \ln y}, \tag{1.41}$$

where $\mathcal{C}_y$ gives the real MC.

The upper left panel of Figure 1.2 illustrates the relationship between the real MC and capital, and the upper right panel of Figure 1.2 shows the real MC curve.

In the basic CU model, aggregate capacity is underutilized in steady state. Hence, the real MC curve is flat and capital resources are slack locally around the steady state. In the Leontief-RBC model, where aggregate capacity is fully utilized, the real MC curve is vertical and capital resources are infinitely tight locally around the steady state. Finally,

[7] It might be interesting to note that if the capacity constraint never binds, one can regard my model as equivalent to a standard monopolistic competition model where the production function is simply linear in labor, unaffected by any physical capital, and firms accumulate a "marketing capital" that affects demand. This analogy breaks down, of course, if the capacity constraint binds occasionally.

the dotted lines in Figure 1.2 are for the standard RBC model with a Cobb-Douglas (CD) production technology, or the CD-RBC model for short, in which capital resources are tight but not infinitely tight.

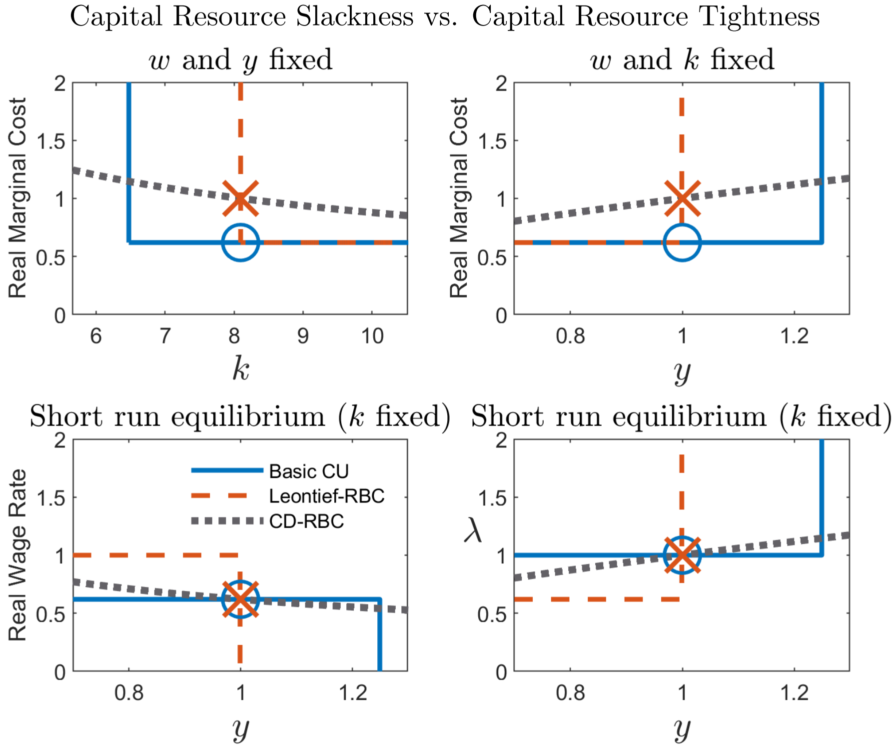

Figure 1.2: The circle marker is the steady state of the basic CU model and the cross marker is the steady state of the Leontief-RBC model and the CD-RBC model.

1.3.4 Acyclical Real Wage Rate and Large Responses to Demand

Capital resource slackness implies an acyclical real wage rate and allows output to be highly responsive to demand shocks, while capital resource tightness implies a countercyclical real wage rate and dampens the response of output to demand shocks.

In the Leontief-RBC model, for example, capital resources are infinitely tight. When demand increases, output cannot increase, but the marginal value of relaxing the capacity constraint μ_t increases, creating a pressure for the real MC to increase. According to the firm's pricing condition (1.26), this pressure is transmitted to a decrease in the real wage rate. The lower left panel of Figure 1.2 shows this relationship between the real wage rate and output in a short run equilibrium. The countercyclical real wage rate will then cause a pro-cyclical price in utils λ_t according to the household's labor supply condition (1.19). The lower right panel of Figure 1.2 shows this short run equilibrium relationship. As is clear from the household's consumption condition (1.20) and the firm's investment condition (1.27), the pro-cyclical price in utils λ_t reduces the responses of consumption and investment to demand shocks.

By contrast, in the basic CU model, capital resources are slack. If there is a higher demand, more output can be produced without causing any pressure for the real MC to increase. The real wage rate w_t and the price in utils λ_t are both locally constant under demand shocks (see the solid lines of the lower two panels of Figure 1.2). Hence, the responses of consumption and investment to demand shocks are not reduced.

To see this clearly, I log-linearize the household's consumption condition (1.20) and the firm's investment condition (1.27) of the basic CU model around the steady state where capacity is in excess. Let $\hat{x}_t \equiv \ln x_t - \ln x$ be the log-deviation of x_t from its steady state x. The log-linearized consumption condition is

$$0 = z_{c,t} - \hat{c}_t, \tag{1.42}$$

and the log-linearized investment condition is

$$0 = \mathbb{E}_t \left(\hat{u}_{t+1} \right) = \mathbb{E}_t \left(\hat{y}_{t+1} - \hat{k}_{t+1} \right), \tag{1.43}$$

where $\hat{k}_{t+1}$ follows a log-linearized capital law of motion: $\hat{k}_{t+1} = (1 - \delta) \hat{k}_t + \delta \hat{i}_t$.

Equation (1.42) shows that consumption responds one-to-one to changes in consumption demand. Equation (1.43) shows that if there is an expected increase in future demand, investment shoots up so that capacity can match with the increased demand in the next period. Although capital is not scarce as a production factor, firms still have a strong desire to invest because the amount of demand that can be attracted by each unit of capacity invested increases with the expected aggregate demand $\mathbb{E}_t (\hat{y}_{t+1})$. Since capital resources are slack, all the induced investment can and will be made immediately in the first period. Therefore, the response of investment to a persistent increase in consumption demand will be very large in the basic CU model.

1.3.5 Calibration

To illustrate the above results quantitatively, I calibrate the basic CU model at a quarterly frequency.

We have eight parameters to calibrate. The depreciation rate δ is calibrated to match the average ratio of gross private domestic investment to private fixed assets from 1947 to

Table 1.1: Parameters and Calibration Targets – Basic CU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Elasticity of inter-temporal substitution 1
ϕ	0.8300	Price in utils normalized to 1
$\bar{\omega}$	0.6798	Output normalized to 1
β	0.9747	Capacity utilization rate 0.8
A	0.1544	Investment to output ratio 0.17
α_v	0.9120	Labor underutilization rate 0.088
Λ	0.3800	Labor share of income 0.62

2016 in the National Income and Product Accounts (NIPA) prepared by the U.S. Bureau of Economic Analysis (BEA). The inverse of the elasticity of inter-temporal substitution γ is chosen to be 1, a value that implies a commonly used log utility. One can see from equation (1.42) that the value of γ actually does not affect the dynamic properties of the basic CU model.

The rest of the six parameters are jointly calibrated to achieve the following targets in steady state.

The marginal utility of income in steady state λ is normalized to 1. This target is mostly associated with the scaling parameter ϕ in the representative household's utility function.

The size of output in steady state y is normalized to 1. This target is mostly associated with the dis-utility of labor $\bar{\omega}$, which affects the size of the economy through the supply of labor.

The labor underutilization rate in steady state is defined as one minus the ratio of the labor hours actually utilized to the total labor hours that the representative household can potentially supply: $1 - l$, where the total hours that the representative household could supply is normalized to one. I choose the average of U-5 and U-6 from 1994 to 2016 published by the Bureau of Labor Statistics (BLS) as a target for the labor underutilization rate.[8] This target is mostly associated with the required labor per unit of output α_v, which determines the demand for labor.

The investment to output ratio in steady state i/y is matched to the average ratio of gross private domestic investment to gross domestic product (GDP) from 1947 to 2016 in

[8] According to the BLS, U-5 is defined as total unemployed, plus discouraged workers, plus all other persons marginally attached to the labor force, as a percent of the civilian labor force plus all persons marginally attached to the labor force. U-6 is defined as total unemployed, plus all persons marginally attached to the labor force, plus total employed part time for economic reasons, as a percent of the civilian labor force plus all persons marginally attached to the labor force.

NIPA. This target is mostly associated with the productivity of capital A, which affects the capital to output ratio.

The capacity utilization rate in steady state u is matched to the average of the total industry capacity utilization rate from 1967 to 2016 reported by the Federal Reserve Board (FRB). This target is mostly associated with the subjective discount factor β, which affects the opportunity cost of holding capacity.

The labor share of income in steady state wl/y is matched to the average labor share of income estimated by the BLS from 1946 to 2016. This target is mostly associated with the unit cost of processing information Λ or the demand elasticity $\varepsilon = \Lambda^{-1}$, which affects the size of the monopolistic profit and thus the labor share of income.

Table 1.1 summarizes the calibrated parameter values and their mostly associated targets.[9]

1.3.6 Impulse Responses and Discussions

Figure 1.3 plots the impulse response functions (IRFs) for the basic CU model together with those for the Leontief-RBC model and those for the CD-RBC model. The first two rows of Figure 1.3 show the case where the change in consumption demand is completely transitory ($\rho_c = 0$), while the last two rows of Figure 1.3 show the case where the change in consumption demand is highly persistent ($\rho_c = 0.99$).

Indeed, although the only difference between the Leontief-RBC model and the basic CU model is the assumption on the behavior of buyers, the local dynamics of the two models are drastically different.

In the Leontief-RBC model, the capacity constraint is binding and capital resources are infinitely tight. Output is restricted by capacity and cannot be changed immediately. Hence, consumption and investment must move in opposite directions. If there is a one-off increase in consumption demand ($\rho_c = 0$), a 1% increase in consumption demand can only lead to a 0.84% increase in consumption, and the increase in consumption must be satisfied by a sharp decrease in investment. If the increase in consumption demand is highly persistent ($\rho_c = 0.99$), it is worthwhile to invest so that households can enjoy a higher consumption in the long run. However, the induced investment must be satisfied

[9] For comparison, I also calibrated the Leontief-RBC model and the CD-RBC model. I do not use the observed capacity utilization rate as a calibration target for the RBC models because capacity turns out to be fully utilized in steady state. The other calibration targets are the same as in the basic CU model. See Appendix 4.1 for details.

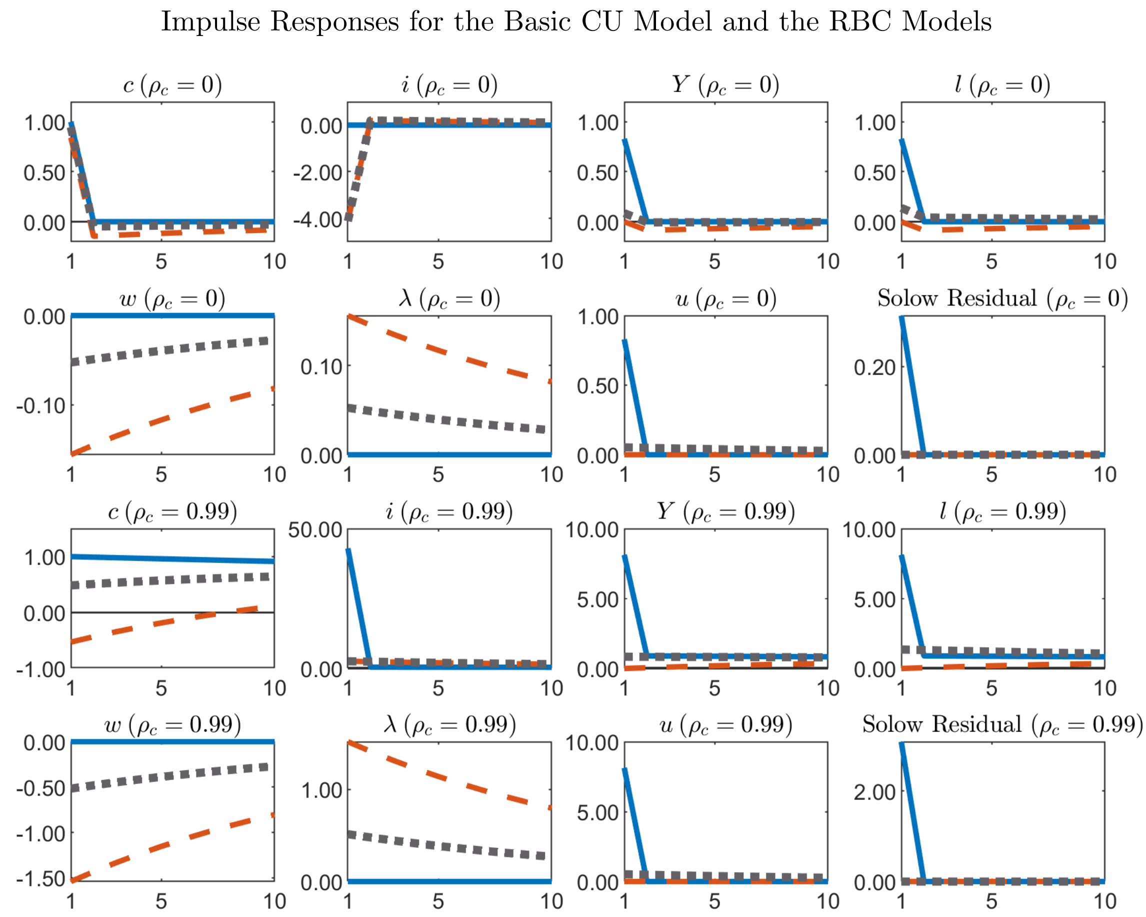

Figure 1.3: Responses to a 1 percentage point increase in demand disturbance e_c. The solid lines are for the basic CU model. The dashed lines are for the Leontief-RBC model. The dotted lines are for the CD-RBC model. All variables are expressed as log deviations from steady state.

by a temporary decrease in consumption. Besides, the real wage rate correlates negatively with demand shocks and the Solow residual is acyclical.[10] In the U.S. data, however, consumption and investment co-move, the real wage rate is weakly pro-cyclical or acyclical, and the Solow residual is strongly pro-cyclical. These illustrate the typical difficulties of demand driving business cycles in an economy where capacity is fully utilized.

In the basic CU model, capacity is in excess and capital resources are slack. Consumption and investment can now move in the same direction. If changes in consumption demand are completely transitory ($\rho_c = 0$), a 1% increase in consumption demand can lead to a 1% increase in consumption and *no* investment has to be sacrificed.[11] If changes in consumption demand are persistent ($\rho_c = 0.99$), a 1% increase in consumption demand

[10] Throughout this paper, the Solow residual is calculated by assuming a CD production function for the final products and a CD labor share of 0.62.

[11] The one-off response of consumption does not mean that consumption smoothing motive is absent. Consumption smoothing motive does not say that a household always wants her consumption to be stable but says that she wants her marginal utility of consumption to be stable. Since the demand shock is a preference shock that increases the marginal utility of consumption $\phi e^{\gamma z_{c,t}} c_t^{-\gamma}$ only in the first period, to smooth out the marginal utility of consumption across time, the best thing to do is to consume more in the first period, offsetting the effect of the increased $z_{c,t}$, and then to consume normally as $z_{c,t}$ goes back to normal. This is exactly what happens in the basic CU model.

can lead to a 1% increase in consumption and also a huge increase in investment. Firms want to increase their investment because the amount of demand that can be stolen by each unit of capacity invested increases. A 1% increase in future consumption encourages roughly a 1% increase in capacity, which in turn requires roughly a $(1/\delta)\,\% = 48\%$ increase in investment. Since capital resources are slack, investment can be made immediately so that capacity is raised up as soon as possible to compete for the increased demand.[12] Because of the induced investment, a 1% persistent increase in consumption demand is amplified to more than 8% increase in output. Capital resource slackness also allows the real wage rate to be acyclical. Finally, as a result of a pro-cyclical capacity utilization, the Solow residual is highly pro-cyclical under demand shocks.

In the CD-RBC model, capital resources are tight but not as tight as in the Leontief-RBC. Hence, the response of output to demand shocks is larger than in the Leontief-RBC model, but still much smaller than in the basic CU model. Particularly, the response of investment is as limited as in the Leontief-RBC model. In addition, the real wage rate correlates negatively with demand shocks and the Solow residual is acyclical.

To sum up, the basic CU model exhibits a positive co-movement between consumption, investment, and hours, a large fluctuation in investment, an acyclical real wage rate, and a pro-cyclical Solow residual, even though all fluctuations are driven by demand shocks.[13] These results are difficult to be obtained in the standard RBC models.

1.4 Comparison with a Standard Variable Capital Utilization Model

In this section, I introduce a standard model that features variable capital utilization (VU). In the standard VU model, capital is not fully utilized because of a convex utilization cost. However, unlike the basic CU model, the standard VU model does not feature any capacity competition externality. There is only a price competition among firms and the only role of capital in the standard VU model is to save the aggregate total cost.

[12] This lack of persistence in the response of investment in the basic CU model follows from capital resource slackness caused by the existence of unused capacity, but does not depend on the linear relationship between the capacity and capital. Suppose that the production function of the firm is given by $y_{j,t} = \min\left\{l_{j,t}/\alpha_v, Ak_{j,t}^{\alpha}\right\}$, where $\alpha \in (0,1)$. In this example, even though capacity is given by a non-linear function of capital, $\bar{y}_{j,t} = Ak_{j,t}^{\alpha}$, we still have that the response of investment to demand shocks concentrated in the first period.

[13] A drawback of the basic CU model is that the response of investment lacks persistence. This issue, however, can be easily resolved by including capital and (or) investment adjustment costs (see section 1.5.1, for further discussions).

Based on this observation, I show that despite *capital* underutilization in the standard VU model, *aggregate capacity* is fully utilized in steady state and capital resources are tight locally around the steady state. Hence, the real wage rate under demand shocks must be countercyclical and the response of output to demand shocks is dampened.

1.4.1 The Standard VU Model

Setup

The setup of the households is exactly the same as in the basic CU model (see section 1.2.3).

The goods market is characterized by the Dixit-Stiglitz monopolistic competition market structure. There is a continuum of identical firms indexed by $j \in [0, 1]$. Each produces a certain type of differentiated goods. The aggregate goods are composed of the differentiated goods:

$$y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{1.44}$$

where $y_{j,t}$ is the amount of differentiated goods produced by firm j, y_t is the amount of aggregate goods, and $\varepsilon > 1$ is the elasticity of substitution among differentiated goods.

Households and firms demand only aggregate goods. The goods demand faced by firm j is given by

$$y_{j,t} = y_t \left(\frac{P_{j,t}}{P_t} \right)^{-\varepsilon}, \tag{1.45}$$

where P_t is the price of aggregate goods: $P_t = \left(\int_0^1 P_{j,t}^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}}$.

Each firm operates with a production technology that is of constant elasticity of substitution (CES) between capital and labor:

$$F\left(\theta_{j,t} k_{j,t}, l_{j,t}\right) = \left(\alpha \left(\frac{l_{j,t}}{\alpha_v} \right)^{\frac{\epsilon_k - 1}{\epsilon_k}} + (1-\alpha) \left(\theta_{j,t} A k_{j,t}\right)^{\frac{\epsilon_k - 1}{\epsilon_k}} \right)^{\frac{\epsilon_k}{\epsilon_k - 1}}, \tag{1.46}$$

where $\theta_{j,t}$ is the capital utilization rate, $\alpha \in (0, 1)$ is the production weight of labor, and $\epsilon_k \in [0, 1]$ is the elasticity of substitution between capital and labor. When $\epsilon_k \to 0$, the CES production function converges to the Leontief production function as in the basic CU model; and when $\epsilon_k \to 1$, the CES production function converges to the usual CD production function.

Following Christiano et al., 2005 and Smets and Wouters (2007), it is costly to utilize capital. The capital utilization cost is paid in terms of goods, increasing and convex in the capital utilization rate, and proportional to the size of capital: $a(\theta_{j,t})\, k_{j,t}$, $a \geq 0$, $a' > 0$, and $a'' > 0$.

1. Cost Minimization. For a given level of output and capital, the firm minimizes its variable cost by choosing its labor input and its capital utilization rate:

$$C(y_{j,t}, k_{j,t}; w_t) = \min_{l_{j,t}, \theta_{j,t} \geq 0} w_t l_{j,t} + a(\theta_{j,t})\, k_{j,t}, \tag{1.47}$$

subject to its production constraint $y_{j,t} \leq F(\theta_{j,t} k_{j,t}, l_{j,t})$. By the envelope theorem and the implicit function theorem, C is non-negative, twice differentiable, and homogeneous of degree one; and $C_y > 0$, $C_k < 0$, and $C_{kk} > 0$.

2. Firm Value Maximization. Taking the variable cost function C as given, the firm maximizes its firm value

$$\max_{\{P_{j,t}, d_{j,t}, y_{j,t}, i_{j,t}, k_{j,t+1}\}} \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} d_{j,t} \right), \tag{1.48}$$

subject to the demand curve (1.45), the resource constraint of the firm

$$d_{j,t} + i_{j,t} = \frac{P_{j,t}}{P_t} y_{j,t} - C(y_{j,t}, k_{j,t}; w_t), \tag{1.49}$$

and the law of motion for capital $k_{j,t+1} = k_{j,t}(1-\delta) + i_{j,t}$.

Symmetric Equilibrium

The symmetric equilibrium is a stable stochastic process of eight variables (c_t, λ_t, w_t, i_t, θ_t, k_{t+1}, l_t, and y_t) that satisfies the household's FOCs (1.20)-(1.19), the firm's pricing condition

$$1 = \frac{\varepsilon}{\varepsilon - 1} C_y(y_t, k_t; w_t), \tag{1.50}$$

the firm's capital utilization condition

$$C_y(y_t, k_t; w_t)\, F_K(\theta_t k_t, l_t) = a'(\theta_t), \tag{1.51}$$

the firm's investment condition

$$1=\beta\mathbb{E}_t\left(\frac{\lambda_{t+1}}{\lambda_t}\left(-C_k\left(y_{t+1},k_{t+1};w_{t+1}\right)+1-\delta\right)\right), \tag{1.52}$$

the production function (1.46), the law of motion for capital

$$k_{t+1}=k_t\left(1-\delta\right)+i_t, \tag{1.53}$$

and the resource constraint

$$c_t+i_t=y_t-a\left(\theta_t\right)k_t=Y_t, \tag{1.54}$$

where Y_t is the final product.[14]

1.4.2 Properties of the Standard VU Model

No Capacity Competition Externality

The only source of inefficiency in the standard VU model is the monopolistic power of firms. If and only if the goods market is perfectly competitive ($\varepsilon\to\infty$), the economy is efficient. However, the price competition is the only competition among firms. Unlike firms in the basic CU model, firms in the standard VU model cannot expand capacity to compete for buyers. It is shown in the rest of this subsection that without a capacity competition externality, the monopolistic distortion alone does not imply chronic excess capacity nor capital resource slackness.

No Chronic Excess Capacity

Claim 1.4. Capacity at the firm-level is fully utilized in steady state.

Proof. Consider a representative firm in the steady state of a symmetric equilibrium. According to the firm's investment condition (1.52), in steady state, the capital level k should be such that solves

$$1=\beta\left(-C_k\left(y,k;w\right)+1-\delta\right), \tag{1.55}$$

[14] In the basic CU model, the Leontief-RBC model, and the CD-RBC model, no intermediate goods are used as a cost of capital utilization; thus, the final product is simply given by the amount of goods produced by firms: $Y_t=y_t=c_t+i_t$.

or, equivalently, minimizes the total cost: $C(y,k;w)+(r+\delta)k$, where $r=\beta^{-1}-1$. By definition, the capacity of the firm $\bar{y}$ is the output level at which the short run average total cost (SRAC) curve is tangent to the *flat* long run average total cost (LRAC) curve. Note that the tangency point is unique because C is strictly convex. Since the point (y,k) is on both the SRAC curve and the LRAC curve, we have $\bar{y}=y$, meaning that capacity is fully utilized in steady state. □

Intuitively, when capacity has no effect on demand, the only reason for firms to invest in capital is to reduce the total cost of production. If capacity is underutilized in steady state, it is capital reduction rather than expansion that reduces the total cost. Hence, firms will decrease their capital stock until their capacity is no larger than sales, restoring the full utilization of capacity in steady state. This is true despite *capital* underutilization. In the standard VU model, capital is not fully utilized because doing so is too costly. Hence, before the full utilization of capital, the average total cost has already reached its long run minimum.

The following claim says that aggregate capacity is also fully utilized in steady state.

Claim 1.5. Aggregate capacity is fully utilized in steady state. Thus, the economy does not exhibit chronic excess capacity.

Proof. The aggregate capacity $\bar{Y}_t$ is the output level at which the aggregate SRAC is tangent to the aggregate LRAC. The aggregate SRAC is given by

$$\mathcal{SRAC}(Y_t,k_t;w_t)=\frac{\mathcal{C}(Y_t,k_t;w_t)+(r+\delta)k_t}{Y_t}, \tag{1.56}$$

where $\mathcal{C}$ is the aggregate variable cost incurred to satisfy the aggregate demand Y_t. When there is a production network and firms have some monopolistic power, the aggregate variable cost function will be distorted and is given by:

$$\mathcal{C}(Y_t,k_t;w_t)=\min_{\{\theta_{j,t},l_{j,t},y_{j,t},k_{j,t}\geq 0\}_{j\in[0,1]}} w_t\int_0^1 l_{j,t}dj+\frac{1}{\varepsilon}\int_0^1 a(\theta_{j,t})k_{j,t}dj, \tag{1.57}$$

subject to the aggregate resource constraint $Y_t\leq\left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}}dj\right)^{\frac{\varepsilon}{\varepsilon-1}}-\int_0^1 a(\theta_{j,t})k_{j,t}dj$, the production function $y_{j,t}\leq F(\theta_{j,t}k_{j,t},l_{j,t})$, and the capital stock available $\int_0^1 k_{j,t}dj\leq k_t$, where ε^{-1} is a "shadow tax" on intermediate goods, capturing the monopolistic distortion. It is easy to verify that $\mathcal{C}$ is strictly convex and that in a symmetric equilibrium $\mathcal{C}_k(Y_t,k_t;w_t)=C_k(y_t,k_t;w_t)$. Hence, in the steady state of a symmetric equilibrium, we have

$$1 = \beta \left(-\mathcal{C}_k \left(Y, k; w \right) + 1 - \delta \right), \tag{1.58}$$

which imples that the aggregate capital in steady state minimizes the aggregate SRAC. By definition, we have $\bar{Y} = Y$, meaning that aggregate capacity is fully utilized in steady state. Thus, the economy does not exhibit chronic excess capacity. □

Capital Resource Tightness

By the envelope theorem and implicit function theorem, one can show that the aggregate variable cost function $\mathcal{C}$ (1.57) is twice differentiable and $\mathcal{C}_{Yk} < 0$, which means that a marginal decrease in capital would cause an increase in the real marginal cost (MC). Intuitively, when there is a decrease in capital, either the remaining capital has to used more intensively or the labor to capital ratio has to increase. Both will cause an increase in the real MC. Thus, capital resources are tight in the standard VU model.

When there is a production network and monopolistic distortions, the aggregate variable cost function incorporates a "shadow tax" on intermediate goods (see the second term in equation (1.57)). Hence, capital resource tightness shall not be measured directly by the increase in the real MC after a decrease in capital.[15] Instead, capital resource tightness can be measured by a decrease in the real wage rate (or the marginal product of labor) that keeps the real MC unchanged after a decrease in capital.[16] Because the production function (1.46) is of constant returns to scale, the measure of capital resource tightness is also equal to a decrease in the real wage rate (or the marginal product of labor) that keeps the real MC unchanged after an increase in aggregate output.

In the standard VU model, the magnitude of capital resource tightness in steady state is given by

$$\begin{aligned} \zeta \equiv \frac{-\mathcal{C}_{Yk} k}{\mathcal{C}_{Yw} w} &= \frac{\mathcal{C}_{YY} Y}{\mathcal{C}_{Yw} w} \\ &= \frac{1-\alpha}{\alpha} \frac{\xi}{1 + \epsilon_k \xi - \psi \left(1 - \alpha\right)} \frac{Y}{Y + a\left(\theta\right) k} > 0, \end{aligned} \tag{1.59}$$

where $\xi \equiv \frac{a''(\theta)\theta}{a'(\theta)} > 0$ measures the convexity of the capital utilization cost function in steady state and $\psi \equiv 1 - \varepsilon^{-1}$ is the real MC in steady state.

---

[15]When there is a "shadow tax" on intermediate goods, a percentage increase in the real MC after a percentage decrease in capital underestimates the decrease in the marginal product of labor.

[16]This new measure is the same as the old measure (1.41) when the aggregate variable cost does not contain any "shadow tax".

The more convex the capital utilization cost function is, the tighter the capital resources. The upper panels of Figure 1.4 illustrate how capital resource tightness is affected by the convexity of the capital utilization cost function ξ in a standard VU model with a CD production function ($\epsilon_k = 1$) and a perfectly competitive market ($\varepsilon \to \infty$), or the CD-VU model for short. Intuitively, when the capital utilization cost becomes more convex, an increase in output would cause a larger increase in the marginal cost of utilization. Hence, it becomes more difficult to squeeze output out of the production system.

Similarly, if the relative importance of labor in production α or the substitutability between labor and capital ϵ_k is smaller, it becomes more difficult to squeeze output out of the production system using labor. The parameter α also controls the relative importance of capital utilization in production. A smaller α implies that the production system relies more on using intermediate goods to utilize capital. Hence, more goods have to be produced to satisfy an extra unit of aggregate demand. Because of this amplification effect due to the production network, it becomes more difficult to squeeze a final product out of the production system. Finally, the existence of market power ($\psi < 1$) dis-encourages the use of intermediate goods, reduces the amplification effect due to the production network, and relaxes capital resource tightness.

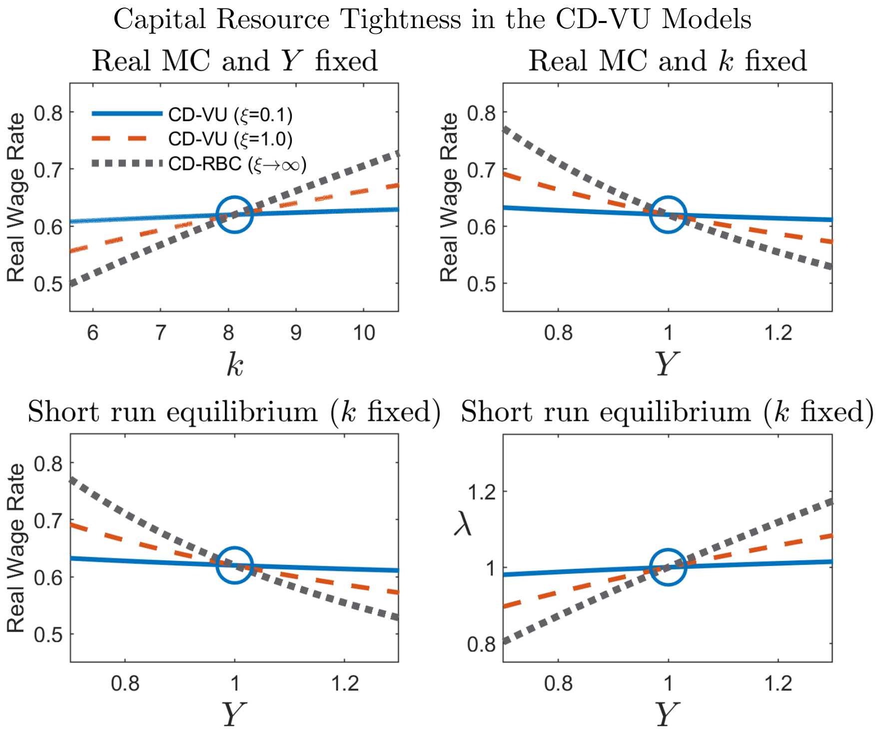

Figure 1.4: The circle marker indicates the steady state of the CD-VU models and ξ measures the convexity of the capital utilization cost.

Countercyclical Real Wage Rate and Limited Responses to Demand

Capital resource tightness implies a countercyclical real wage rate and limits the response of output to demand shocks.

In the standard VU model, since capital resources are tight, an increase in output generates a pressure for the real MC to increase. This pressure will be transmitted to a decrease in the real wage rate according to the firm's pricing condition (1.50). The lower left panel of Figure 1.4 shows this relationship between the real wage rate and output in a short run equilibrium. At a first order approximation, we have

$$\hat{w}_t = -\zeta \left(\hat{Y}_t - \hat{k}_t \right). \tag{1.60}$$

The countercyclical real wage rate will then cause a pro-cyclical price in utils λ_t according to the household's labor supply condition (1.19). The lower right panel of Figure 1.4 shows this short run equilibrium relationship. Finally, the pro-cyclical price in utils reduces the responses of consumption and investment to demand shocks.

To see this point clearly, I log-linearize the household's consumption condition (1.20)

$$\zeta \left(\hat{Y}_t - \hat{k}_t \right) = \hat{\lambda}_t = \gamma \left(z_{c,t} - \hat{c}_t \right), \tag{1.61}$$

and the firm's investment condition (1.52)

$$\zeta \left(\hat{Y}_t - \hat{k}_t \right) = \hat{\lambda}_t = \mathbb{E}_t \left(\left(1 + \beta \frac{wl}{k} \right) \zeta \left(\hat{Y}_{t+1} - \hat{k}_{t+1} \right) \right). \tag{1.62}$$

Equations (1.61) and (1.62) show that the pro-cyclical price in utils λ_t reduces the responses of consumption and investment.

If the magnitude of capital resource tightness converges to zero ($\zeta \to 0$), the real wage rate is almost acyclical and consumption is almost as responsive as in the basic CU model.

One might think that investment can also be almost as responsive as in the basic CU model if capital resource tightness is negligible. This conjecture, however, is not true. Suppose that the magnitude of capital resource tightness is reduced. On one hand, the price in utils λ_t is less pro-cyclical so that the dampening effect becomes weaker (see the left hand side of equation (1.62)). On the other hand, the return of capital also becomes less pro-cyclical so that firms have less incentive to adjust capital (see the right hand side of equation (1.62)). These two effects cancel out.

In fact, firms want to adjust capital precisely because capital resources are tight so that there is a pro-cyclical upward pressure on the real MC that firms can adjust their capital to mitigate, or equivalently, a countercyclical real wage rate that firms can adjust their capital to take advantage of. If capital resources are less tight, firms will have less incentive to adjust capital. In other words, removing the curb on capital adjustment also removes the impetus for capital adjustment. Therefore, the response of investment in the standard VU model is always much limited.

As capital resources become less tight, consumption will be more volatile while investment will not. Thus, the relative volatility of investment to consumption declines. Hence, it is difficult to achieve a large relative volatility of investment to consumption as observed in the U.S. data without causing a strongly countercyclical real wage rate, restricting the role of consumption demand shocks in driving business cycles.

1.4.3 Calibration

To document the above results quantitatively, I calibrate the CD-VU model, where $\epsilon_k = 1$ and $\varepsilon \to \infty$. Since capacity is fully utilized in steady state, I do not use the observed average capacity utilization rate as a calibration target. Also note that capital productivity A is not identified because for each value of $A > 0$, there exists a value of α_v such that the total factor productivity (TFP) $\alpha_v^{-\alpha} A^{1-\alpha}$ is the same. I choose A such that the ratio y/Ak is normalized to 1. Capital utilization rate in steady state θ is normalized to 1. As in Smets and Wouters (2007), the capital utilization cost in steady state $a(\theta)k$ is normalized to 0.[17] The convexity of the capital utilization cost function in steady state ξ is set to 0.1, 1, or infinity, for a sensitivity analysis. The other calibration targets are the same as in the basic CU model.

Table 1.2 lists the calibrated parameter values and their mostly associated calibration targets.

[17] It is theoretically flawed to set $a(\theta)k$ to 0 as it implies a negative utilization cost if the capital utilization rate is smaller than that in steady state. Quantitatively, for any value of $a(\theta)k \geq 0$, there exists a value of $\xi > 0$ such that the magnitude of capital resource tightness ζ remains unchanged; thus, the local dynamics of consumption, investment, and real wage rate are not affected. The local dynamics of hours are also unaffected if markets are perfectly competitive. If there is a market power, a larger value of $a(\theta)k$ would cause hours to be slightly more pro-cyclical under demand shocks. However, in the standard VU model, hours are already too pro-cyclical, implying a countercyclical output to labor ratio. Hence, by following the convention in Smets and Wouters (2007), I do not overstate the difficulties of demand driving business cycles.

Table 1.2: Parameters and Calibration Targets – CD-VU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Elasticity of inter-temporal substitution 1
ϕ	0.8300	Price in utils normalized to 1
$\bar{\omega}$	0.6798	Output normalized to 1
β	0.9747	Investment to output ratio 0.17
α	0.6200	Labor share of income 0.62
α_v	0.9120	Labor underutilization rate 0.088
A	0.1235	y/Ak normalized to 1
θ	1.0000	θ normalized to 1
$a(\theta)$	0.0000	$a(\theta)$ normalized to 0

1.4.4 Impulse Responses and Discussions

Figure 1.5 shows the IRFs for the CD-VU model with different degrees of capital resource tightness. Generally speaking, despite having a variable capital utilization, the dynamic properties of the CD-VU model are not too different from those of the standard CD-RBC model.

If the convexity of the capital utilization cost converges to infinity ($\xi \to \infty$), the CD-VU model converges to the CD-RBC model, in which case capital resources are quite tight, the real wage rate is strongly countercyclical, and the response of output to demand shocks is quite limited.

Table 1.3: Initial Responses and Relative Volatility of Investment to Consumption

	$\rho_c = 0$				$\rho_c = 0.99$			
	c	i	Y	σ_i/σ_c	c	i	Y	σ_i/σ_c
Basic CU	1.00%	0.00%	0.83%	0.00	1.00%	43.0%	8.13%	31.63
Leontief-RBC	0.84%	-4.12%	0.00%	4.57	-0.53%	2.58%	0.00%	4.36
CD-RBC	0.95%	-4.12%	0.09%	4.34	0.48%	2.58%	0.84%	5.07
CD-VU ($\xi = 1$)	0.96%	-4.12%	0.10%	4.31	0.63%	2.58%	0.97%	3.99
CD-VU ($\xi = 0.1$)	0.99%	-4.12%	0.12%	4.24	0.90%	2.58%	1.18%	2.86
The U.S. data				6.49				6.49

Note: Initial responses to a 1 percentage point increase in demand disturbance e_c are expressed as log deviations from steady state. σ_i/σ_c stands for the relative volatility of investment to consumption. The U.S. data is from the BEA. A path of 5,000 quarters is simulated to calculate the statistics for each calibrated model. Logarithms of the original series are Hodrick-Prescott (HP) filtered with a smoothing parameter of 1,600 to calculate the relative volatility.

If ξ is smaller, capital resources are less tight, the real wage rate is less countercyclical, and consumption is more responsive. However, regardless of the magnitude of capital resource tightness, the response of investment to consumption demand shocks is always as small

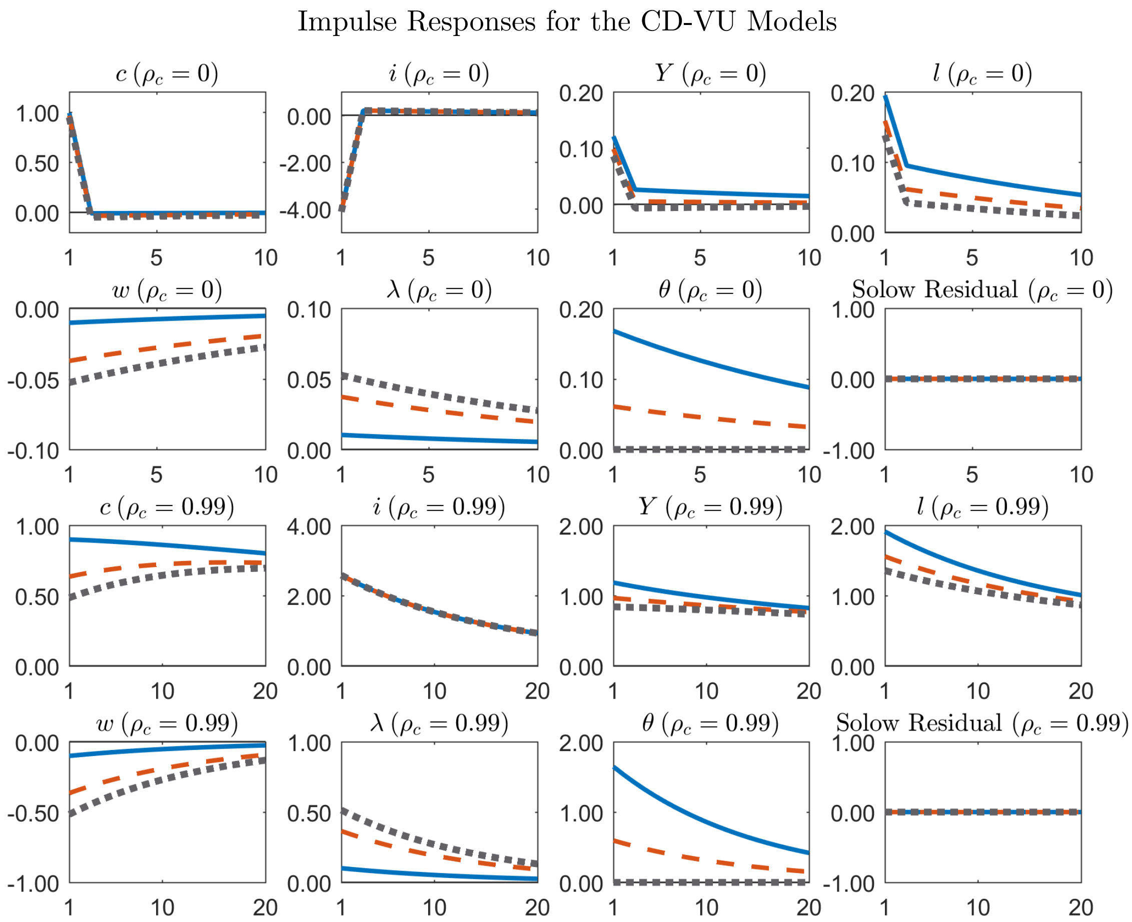

Figure 1.5: Responses to a 1 percentage point increase in demand disturbance e_c. The solid lines are for the CD-VU model with $\xi = 0.1$. The dashed lines are for the CD-VU model with $\xi = 1$. The dotted lines are for the CD-RBC model or the CD-VU model with $\xi \to \infty$. All variables are expressed as log deviations from steady state.

as in the CD-RBC model, which is dwarfed if compared to that in the basic CU model (see Table 1.3). Since consumption is more responsive but investment is not, the relative volatility of investment to consumption *decreases*. As a result, to achieve a large relative volatility of investment to consumption as in the U.S. data, capital resources must be tight enough. Table 1.3 shows that among the CD-VU models, the standard CD-RBC model ($\xi \to \infty$) actually has the highest relative volatility of investment to consumption.[18] In this case, however, the real wage rate is highly countercyclical. Hence, the standard VU model fails to generate a large relative volatility of investment to consumption without causing a strongly countercyclical real wage rate.[19]

Finally, the Solow residual is always acyclical, despite a pro-cyclical capital utilization rate. Although an increase in capital utilization increases output for a given amount of capital

[18] The relative volatility of investment to consumption in the CD-RBC model is still a bit smaller than that observed in the U.S. data.

[19] If investment adjustment cost is introduced to help obtain a hump-shaped impulse response of investment, it becomes even more difficult to achieve a large relative volatility of investment to consumption in the standard VU model under consumption demand shocks.

and labor, the cost of capital utilization also increases. These two have opposite effects on the Solow residual, and at a first order approximation, they cancel out. It turns out that the Solow residual is a good estimation of the true total factor productivity (TFP) when the economy is perfectly competitive.

To sum up, in the standard VU model, capacity is fully utilized in steady state and capital resources are tight. Unlike the basic CU model, the standard VU model fails to generate a large movement in investment, an acyclical or pro-cyclical real wage rate, and a strongly pro-cyclical Solow residual under consumption demand shocks.

1.5 Estimating the Capacity Underutilization Model

In this section, I extend the basic CU model to a full CU model. I estimate the full CU model using Bayesian estimation techniques. Based on the estimated results, I show that demand shocks are the main driving forces of business cycles. Particularly, consumption demand shocks alone can explain most of the business cycle fluctuations and generate the correct co-movement among consumption, investment, hours, and the Solow residual in the full CU model.

1.5.1 Model Extensions

The full CU model is obtained by extending the basic CU model with home production, indirect labor, capital and investment adjustment costs, and exogenous expenditure.

Home Production

I assume that households use their time to produce goods or services at home. The total number of hours available is normalized to one. The amount of home produced goods or services is given by $c_{h,t} = Z_{h,t}(1 - l_t)$, where $Z_{h,t}$ is the productivity of time spent at home. The one-period utility function of the representative household is now given by

$$u(z_{c,t}, c_t, c_{h,t}) = \begin{cases} \phi e^{z_{c,t}} \frac{\left(c_t/\left(\phi e^{z_{c,t}}\right)\right)^{1-\gamma}-1}{1-\gamma} + \bar{\omega} c_{h,t}, & \gamma \neq 1 \\ \phi e^{z_{c,t}} \ln\left(c_t/(\phi e^{z_{c,t}})\right) + \bar{\omega} c_{h,t}, & \gamma = 1 \end{cases} \tag{1.63}$$

which replaces the utility function (1.18) in the household's problem. The inclusion of home production enriches the model dynamics: an increase in the productivity at home increases the opportunity cost of supplying labor and thus reduces the labor supply.

Later in section 1.5.1, I will relate the productivity at home to the labor productivity of firms $z_{l,t}$: $Z_{h,t} = e^{\phi_l z_{l,t}}$. The parameter ϕ_l is usually set to zero in standard models. In this case, an increase in labor productivity does *not* increase the opportunity cost of supplying labor. Hence, output moves positively with labor productivity. However, if ϕ_l is positive, an increase in labor productivity also increases the productivity at home and reduces the labor supply. Particularly, if ϕ_l is equal to one, an increase in labor productivity will not cause an increase in output but will be fully absorbed by a decrease in working hours. Hence, the magnitude of ϕ_l affects the importance of labor productivity shocks in driving business cycles. I will estimate ϕ_l later in section 1.5.2.

Indirect Labor

I also assume that the Leontief production function of firm $j \in [0, 1]$ at time t is now given by

$$y_{j,t} = \min\left\{Ak_{j,t}, \frac{l_{f,j,t}}{\alpha_{f,t}}, \frac{l_{v,j,t}}{\alpha_{v,t}}\right\}, \tag{1.64}$$

where $l_{v,j,t}$ is the amount of direct labor, $l_{f,j,t}$ is the amount of indirect labor, and the inverse of $\alpha_{f,t}$ is the productivity of the indirect labor.

Direct labor produces goods or services directly. The hours of direct labor fluctuate with output and can be easily adjusted within a period. Examples of direct labor positions are machine operators, assembly line operators, and cleaners.

Indirect labor, by contrast, supports the production process of firms, but is not directly involved in the active conversion of materials into finished products. Like capital stock, indirect labor is predetermined. Examples of indirect labor positions are production supervisor, managerial and various administrative labor positions, such as accounting, marketing, and human resource positions.

By definition, the production capacity of firm j at time t is determined by the short run fixed factors, i.e., indirect labor and capital stock:

$$\bar{y}_{j,t} = \min\left\{Ak_{j,t}, \frac{l_{f,j,t}}{\alpha_{f,t}}\right\}, \tag{1.65}$$

where $\alpha_{f,t}$ can also be understood as the amount of indirect labor required to form a unit of capacity.

The introduction of indirect labor allows output to be more volatile than labor. Hence, labor productivity measured by output to labor ratio can be pro-cyclical under demand shocks.

Capital and Investment Adjustment Costs

To get a persistent and hump-shaped investment response, I introduce capital and investment adjustment costs. The capital stock of firm j is accumulated according to

$$k_{j,t+1} = k_{j,t}\left(1-\delta\right) + i_{j,t}\left(1 - S\left(i_{j,t}, i_{j,t-1}, k_{j,t}\right)\right), \tag{1.66}$$

where S is the adjustment cost function.

Following Hayashi (1982) and Christiano et al. (2005), I assume that the adjustment cost function is given by

$$S\left(i_{j,t}, i_{j,t-1}, k_{j,t}\right) = \frac{\phi_i}{2}\left(\frac{i_{j,t}}{i_{j,t-1}} - 1\right)^2 + \frac{\phi_k}{2}\left(\frac{i_{j,t}}{k_{j,t}} - \delta\right)^2 \frac{k_{j,t}}{i_{j,t}}, \tag{1.67}$$

where $\phi_k \geq 0$ and $\phi_i \geq 0$ are parameters that capture the curvature of the capital and investment adjustment costs respectively.

This functional form implies that to change investment or to deviate from the investment level that maintains the current level of capital is costly. Hence, the adjustment costs are zero in steady state, but the dynamics around the steady state will be influenced by the curvature of these two adjustment cost components.

Exogenous Expenditure

Following Smets and Wouters (2007), government spending and net exports are treated as exogenous expenditure. I abstract away from the crowding-out and (or) crowding-in effects of government spending on consumption and investment by assuming that the exogenous expenditure is produced by an independent sector that requires direct labor only. Let g_t be the exogenous expenditure and $l_{g,t}$ be the corresponding labor hired. We have $l_{g,t} = \alpha_{g,t} g_t$, where $\alpha_{g,t}$ is the direct labor required per unit of exogenous expenditure. The final product of the economy is $Y_t = c_t + i_t + g_t$ and the total amount of hours worked is $l_t = l_{f,t} + l_{v,t} + l_{g,t}$.

Exogenous Shocks

I introduce four types of exogenous shocks in total.

First, the exogenous expenditure is given by $g_t = ge^{z_{g,t}}$, where $z_{g,t}$ follows an AR(1) process with an i.i.d. Normal disturbance: $z_{g,t} = \rho_g z_{g,t-1} + e_{g,t}$. $e_{g,t}$ is a shock to the exogenous expenditure.

Second, the labor productivity $z_{l,t}$ determines the productivity of all types of labor. Specifically, $\alpha_{v,t} = \alpha_v e^{-z_{l,t}}$, $\alpha_{f,t} = \alpha_f e^{-z_{l,t}}$, and $\alpha_{g,t} = \alpha_g e^{-z_{l,t}}$. In addition, the productivity of the time spent at home is also affected by the labor productivity: $Z_{h,t} = e^{\phi_l z_{l,t}}$. $z_{l,t}$ is assumed to follow an AR(1) process with an i.i.d. Normal disturbance: $z_{l,t} = \rho_l z_{l,t-1} + e_{l,t}$. $e_{l,t}$ is a shock to the labor productivity.

Third, there is an investment demand $z_{i,t}$ that affects the subjective discount factor, i.e., the importance of the present relative to the future. The lifetime preference of the representative household is now given by:

$$\mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t e^{-z_{i,t}} u\left(z_{c,t}, c_t, c_{h,t}\right) \right). \tag{1.68}$$

I assume that $z_{i,t}$ follows an AR(2) processes with an i.i.d. Normal disturbance: $z_{i,t} = \rho_{i,1} z_{i,t-1} + \rho_{i,2}\left(z_{i,t-1} - z_{i,t-2}\right) + e_{i,t}$. The assumption is designed to capture both the low-frequency movement and the high-frequency zigzag pattern in investment. $e_{i,t}$ is a shock to the investment demand.

Finally, as in the basic CU model, there is a consumption demand $z_{c,t}$ that affects the marginal utility of consumption relative to the marginal dis-utility of labor. I assume that $z_{c,t}$ follows an AR(1) process with an i.i.d. Normal disturbance and is also affected by the disturbance to the investment demand as follows: $z_{c,t} = \rho_c z_{c,t-1} + e_{c,t} + \rho_{ci} e_{i,t}$. The parameter ρ_{ci} allows us to capture the divergence between consumption and capital in the U.S. from 1968 to 1993.[20] $e_{c,t}$ is a shock to the consumption demand.

All shocks, $e_{c,t}$, $e_{i,t}$, $e_{g,t}$, and $e_{l,t}$, are mutually uncorrelated.

[20] Appendix 4.4 gives some further discussions and shows the estimated consumption series under investment demand shocks $e_{i,t}$.

1.5.2 Estimation

I use Bayesian estimation techniques to estimate the full CU model.[21] Having four exogenous shocks, I am able to match four detrended U.S. macro time series: real consumption, real investment, real exogenous expenditure, and the amount of hours worked.[22]

To alleviate the burden on estimation, I choose nine calibration targets that must be matched throughout this estimation procedure. The elasticity of inter-temporal substitution (EIS) is calibrated to 0.5. In the literature, the estimated EIS ranges roughly from 0 to 2 with a mean about 0.5 (e.g., Hall, 1988, Gruber, 2013, Havránek, 2015, and Best et al., 2020). The exogenous expenditure to output ratio in steady state is matched to the average ratio of the exogenous expenditure to GDP from 1947 to 2016 in the BEA NIPA. The other seven calibration targets are the same as described in section 1.3.5.

Because of these targets, nine parameter values are *not* free to pick. Six parameters are fixed as their values follow directly from the nine calibration targets. Three parameters can be expressed as *functions* of other parameters and the nine calibration targets:

$$\beta = \left(\left(1-\bar{\omega}\alpha_v\right)Au - \bar{\omega}\alpha_f A + 1 - \delta\right)^{-1}, \tag{1.69}$$

$$\alpha_f = \left(l - \alpha_v\left(c+i\right) - \alpha_g g\right)\frac{\delta}{Ai}, \tag{1.70}$$

$$\Lambda = 1 - \bar{\omega}\alpha_v. \tag{1.71}$$

Table 1.4 summarizes these nine parameters and their mostly associated calibration targets.

The other parameters are estimated and their priors are assumed as follows. The standard errors of the innovations follow an inverse-gamma distribution with a mean of 0.03 and a standard deviation of infinity. The persistence parameters of the stochastic processes, ρ_c, ρ_g, $\rho_{i,1}$, and ρ_l, follow a uniform distribution which ranges from -1 to 1. The parameter $\rho_{i,2}$ that allows the investment demand to have a zigzag pattern and the parameter ρ_{ci} that allows the investment innovation to have an effect on the consumption demand both follow a Normal distribution with a mean of 0 and a standard deviation of 0.2. The parameter ϕ_l that controls the relationship between the productivity at home and the productivity of labor is uniformly distributed within the range from -1 to 1. The curvature of the capital adjustment cost ϕ_k is normally distributed around 2 with a standard deviation of 1. The

[21] The model is log-linearized around the steady state and the estimation procedure is done with the software platform Dynare.

[22] See Appendix 4.2 for a description of the data.

Table 1.4: Parameters Pinned down by the Calibration Targets – Full CU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	2.0000	Elasticity of inter-temporal substitution 0.5
A	0.1235	Investment to output ratio 0.17
g	0.2000	Exogenous expenditure to output ratio 0.2
$\bar{\omega}$	0.6798	Output normalized to 1
ϕ	0.6300	Price in utils normalized to 1
β	Eq. (1.69)	Capacity utilization rate 0.8
α_f	Eq. (1.70)	Labor underutilization rate 0.088
Λ	Eq. (1.71)	Labor share of income 0.62

curvature of the investment adjustment cost ϕ_i is normally distributed around 0.2 with a standard deviation of 0.1. The direct labor required per output in private sector α_v and the direct labor required per exogenous expenditure α_g are both normally distributed around 0.6 with a standard error of 0.3.

Table 1.5 summarizes the priors and shows the mode, the mean, and the 5th and 95th percentiles of the posterior distribution of the parameters obtained by the Metropolis-Hastings (MH) algorithm.[23]

Based on the posterior modes of the structural parameters and the nine calibration targets, the subjective discount factor β is 0.99, the indirect labor required per unit of capacity α_f is 0.20, and the unit cost of processing information Λ is 0.49, which implies that the price elasticity of demand $\varepsilon = \Lambda^{-1}$ is 2.06.

1.5.3 Variance Decomposition

What are the main driving forces of business cycles? Table 1.6 gives the forecast error variance decomposition of output, consumption, investment, hours, and the capacity utilization rate at a 10-year horizon based on the posterior modes of the parameters of the full CU model.

According to the estimated full CU model, business cycle movements are primarily driven by three types of demand shocks, i.e., the consumption demand, the investment demand, and the exogenous expenditure shocks. Labor productivity shocks account for only 2.23 percent of the variation in output and 13.42 percent of the variation in hours. Among these three types of demand shocks, consumption demand turns out to be the main driving

[23] Total number of MH draws is 100,000 and the acceptance ratio is about 23.5%.

Table 1.5: Bayesian Estimation – Full CU Model

	Prior Distribution			Posterior Distribution			
Parameter	Distribution	Mean	Std Dev	Mode	Mean	5th Percentile	95th Percentile
σ_c	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
σ_i	Invgamma	0.03	∞	0.03	0.03	0.02	0.05
σ_g	Invgamma	0.03	∞	0.02	0.02	0.02	0.03
σ_l	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
ρ_c	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
$\rho_{i,1}$	Uniform	0.00	0.58	0.98	0.98	0.97	0.99
$\rho_{i,2}$	Normal	0.00	0.20	-0.25	-0.28	-0.38	-0.17
ρ_g	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
ρ_l	Uniform	0.00	0.58	0.96	0.96	0.93	0.99
ρ_{ci}	Normal	0.00	0.20	-0.14	-0.14	-0.19	-0.07
ϕ_l	Uniform	0.00	0.58	0.64	0.64	0.39	0.92
ϕ_k	Normal	2.00	1.00	2.01	2.17	1.20	3.07
ϕ_i	Normal	0.20	0.10	0.18	0.19	0.13	0.26
α_v	Normal	0.60	0.30	0.76	0.75	0.63	0.86
α_g	Normal	0.60	0.30	0.55	0.55	0.42	0.67

Note: Std Dev stands for standard deviation (of the priors). The sample period is from the first quarter of 1948 to the first quarter of 2017.

Table 1.6: Forecast Error Variance Decomposition (%)

		Full CU Model				Full VU Model			
		e_c	e_i	e_g	e_l	e_c	e_i	e_g	e_l
Output	Y	56.53	15.00	26.25	2.23	33.40	4.30	22.09	40.21
Consumption	c	72.66	26.44	0.00	0.01	64.70	15.89	0.00	19.41
Investment	i	78.70	12.75	0.00	8.56	12.58	0.74	0.00	86.69
Hours	l	60.85	14.80	10.93	13.42	66.59	11.58	11.38	10.44
CU Rate	u	71.50	22.03	0.00	6.47	27.50	10.54	0.00	61.96

Note: The forecast error variance decomposition for the estimated models at a 10-year (40-quarter) horizon. CU Rate stands for capacity utilization rate.

force of business cycles as most of the variations in consumption, investment, hours, and the capacity utilization rate are explained by innovations to consumption demand.

For comparison, I extend the standard VU model described in section 1.4.1 by including home production, capital and investment adjustment costs, exogenous expenditure, and the four exogenous shocks in the same way as I did for the basic CU model. The extended VU model is called the full VU model. The calibration and estimation procedure of the full VU model is set to be as close as possible to that of the full CU model.[24] The variance decomposition results of the full VU model are listed in Table 1.6.

Consistent with the standard business cycle literature, a larger fraction of the variation in output and most of the variation in investment are driven by labor productivity shocks. However, labor productivity shocks cannot explain most of the variation in hours (see Smets and Wouters, 2007, for a related discussion). In addition, consumption is not volatile enough under labor productivity shocks (see Bai et al., 2012, for similar results found in a standard RBC model). Therefore, standard models often rely on multiple types of shocks to explain business cycles.

1.5.4 Impulse Responses and Discussions

To understand the variance decomposition results better, I plot the IRFs for both the full CU model and the full VU model in Figure 1.6 based on the parameter values evaluated at the posterior modes.

In the full CU model, a positive innovation to consumption demand leads to large increases in consumption, investment, output, and hours. Not surprisingly, innovations to consumption demand drive business cycles. A positive innovation to investment demand leads to an increase in investment but a small decrease in consumption because the parameter $\rho_{ci} = -0.14$ is estimated to be negative. A positive innovation to exogenous expenditure increases output and hours but has no effect on private consumption and investment because the exogenous expenditure is assumed to be produced independently of the private sector. Finally, since ϕ_l is estimated to be about 0.64, which is quite close to 1, a positive innovation to labor productivity has only a small effect on consumption and investment. As output does not increase that much, labor is displaced by the improved labor productivity. The estimation result shows that the full CU model prefers to shut down the effect of labor productivity shocks on output by setting ϕ_l close to 1 because

[24] See Appendix 4.3 for a detailed description of the calibration and estimation of the full VU model.

Estimated Impulse Responses for the Full CU Model and the Full VU Model

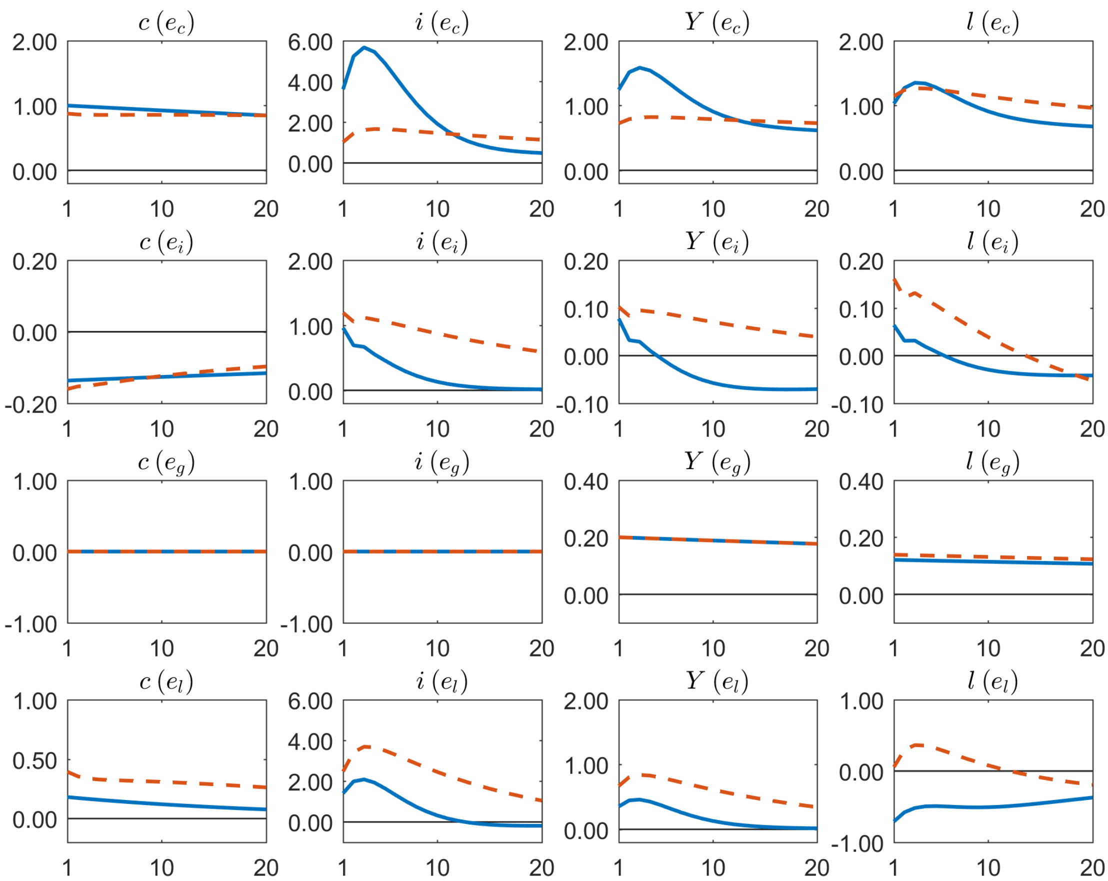

Figure 1.6: Responses to a 1 percentage point increase in disturbances e_c, e_i, e_g, and e_l. The solid lines are the impulse responses for the full CU model. The dashed lines are the impulse responses for the full VU model. All variables are expressed as log deviations from steady state.

the consumption demand channel works better than the labor productivity channel and is powerful enough to explain most of the business cycle fluctuations.

In the full VU model, a positive innovation to consumption demand generates sizable increases in consumption and hours but fails to generate a large increase in investment. A positive innovation to investment demand or to exogenous expenditure does not generate the desired business cycle co-movement for the same reason as in the full CU model. Finally, ϕ_l is estimated to be about -0.02 in the full VU model, which is very close to 0. Thus, a positive innovation to labor productivity can have a large effect on output. The estimation result shows that the full VU model requires the labor productivity channel to explain the business cycle fluctuations. However, the number of hours worked fails to increase much under a positive labor productivity shock because there is an income effect that dampens the response of hours to changes in the labor productivity. Hence, no *single* type of shock in the full VU model is able to drive most of the business cycles fluctuations.

1.5.5 Consumption Demand Drives Business Cycles

To highlight the role of consumption demand shocks in driving business cycles, I shut down all the other types of shocks and feed only the consumption demand shocks into the models. The structural shocks of the estimated models are backed out from the U.S. data using the Kalman smoother technique.

Model Fitted Values Under Consumption Demand Shocks

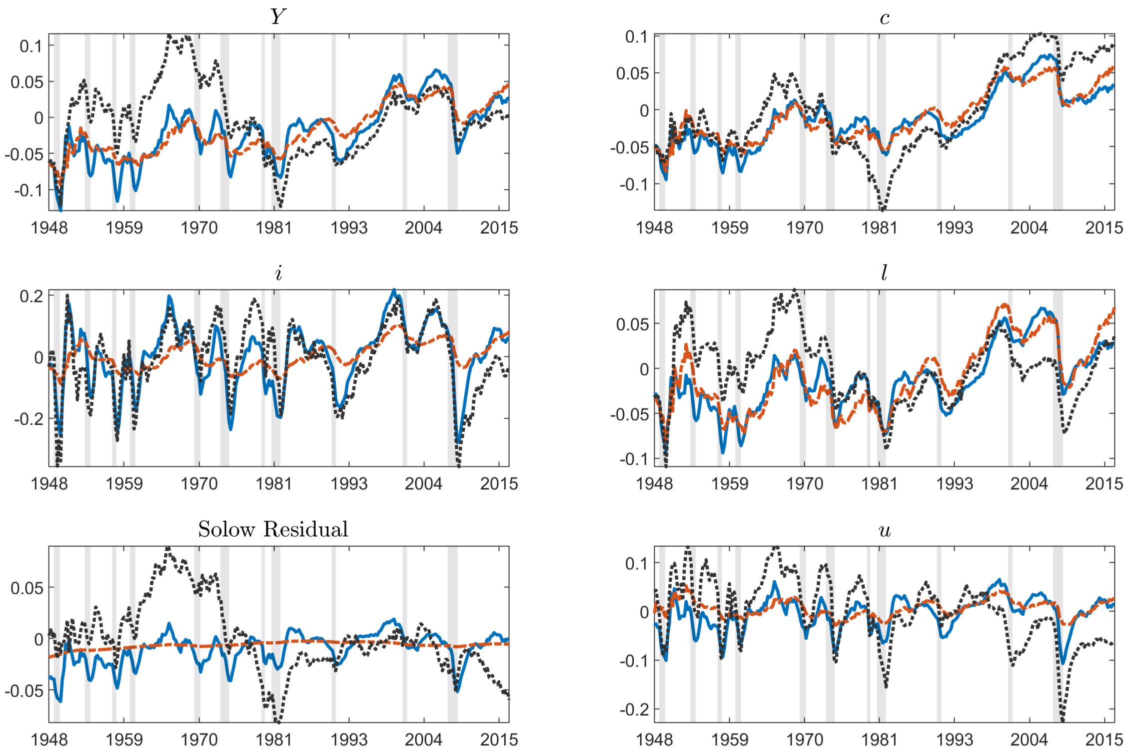

Figure 1.7: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Figure 1.7 shows the model fitted values under consumption demand shocks. The full CU model fits the U.S. data very well. In the full VU model, however, investment has little response to consumption demand shocks and the Solow residual almost does not fluctuate.

Table 1.7 compares the main business cycle statistics of the fitted values with those of the U.S. data. The business cycle properties of the full CU model under consumption demand shocks are quite close to those of the data. Consumption demand shocks not only generate sizable business cycle fluctuations but also generate the correct business cycle co-movement among the key aggregate variables. Particularly, investment is highly volatile, and consumption, investment, hours, and the Solow residual are as pro-cyclical as in the U.S. data. Because of capital resource slackness, the real wage rate in the full CU model is independent of demand and is fully determined by labor productivity. Hence, the real

Table 1.7: Business Cycle Statistics

	The U.S. Data		Full CU Model (e_c)		Full VU Model (e_c)		Full VU Model (e_l)	
	Std Dev	Cov w Y	Std Dev	Cov w Y	Std Dev	Cov w Y	Std Dev	Cov w Y
Output	1.62	1.00	1.76	1.00	0.79	1.00	1.58	1.00
Consumption	1.16	0.57	1.04	0.57	0.83	1.05	0.64	0.39
Investment	7.55	3.95	6.75	3.78	1.59	1.99	7.01	4.42
Hours	1.68	0.89	1.47	0.83	1.25	1.58	0.81	0.47
Solow Resid	1.00	0.39	0.89	0.48	0.03	0.02	1.16	0.72
Real Wage Rate	0.90	0.08	0.00	0.00	0.28	-0.35	1.25	0.77

Note: Std Dev stands for standard deviation. Cov w Y stands for covariance with output. Solow Resid stands for Solow residual. Covariance with output is reported relative to the variance of output. The U.S. data is from the BEA and the BLS. All variables are HP-filtered logarithms of the original series.

wage rate is naturally acyclical under consumption demand shocks.

I also calculate the main business cycle statistics for the fitted values generated by the full VU model under labor productivity shocks. The results are listed in the last three columns of Table 1.7. Although labor productivity shocks in the full VU model can generate large fluctuations in investment and the Solow residual, they fail to generate large movements in hours and consumption. Hence, the full CU model under consumption demand shocks not only outperforms the full VU model under consumption demand shocks but also outperforms the full VU model under labor productivity shocks.

To sum up, the full VU model primarily relies on labor productivity shocks to generate a volatile investment, but hours and consumption are not volatile enough under labor productivity shocks. By contrast, when viewed through the lens of the full CU model, consumption demand shocks can explain most of the business cycle fluctuations.

1.6 Conclusion

I build a macroeconomic model with chronic excess capacity. Buyers are not fully attentive to prices as they are subject to an information processing cost. Because of this friction, buyers search for capacity in a somewhat undirected way. Hence, firms with a larger capacity are more likely to be visited by buyers. For a given amount of total demand from buyers, if a firm expands its capacity relative to that of others, it steals demand from its competitors. Furthermore, each unit of demand stolen is profitable because firms can take advantage of the inattentive buyers to charge a positive net markup. Hence, from an individual firm's perspective, capacity expansion not only relaxes the capacity constraint but also steals profits from other firms. From an aggregate perspective, however, the only value of capacity is to relax the capacity constraint. This capacity competition externality encourages firms to over-invest in capacity and causes the economy to exhibit chronic excess capacity.

When capacity is in excess, capital resources are slack and output is highly responsive to demand shocks. I estimate a version of the capacity underutilization model. Quantitatively, consumption demand shocks can explain most of the business cycle fluctuations, generating large responses in consumption, investment, and hours, an acyclical real wage rate, and a pro-cyclical Solow residual, whereas labor productivity shocks explain only a small fraction of the variations in output and hours. Hence, when viewed through the lens of the capacity underutilization model, demand is the main driving force of business cycles.

Chapter 2

Difficulties of Demand Driving Business Cycles in Efficient Utilization Models

2.1 Introduction

In this chapter, I study what prevents demand shocks from driving business cycles in models with efficient capital or capacity utilization. Demand shocks can be divided into two broad categories: real demand shocks and monetary demand shocks. This chapter focuses on the role of real demand shocks in driving business cycles.

Investment demand shocks, such as news shocks, discount factor shocks, and investment specific technology shocks are widely considered in the business cycle literature (e.g., Greenwood et al., 2000, Jaimovich and Rebelo, 2009, and Justiniano et al., 2010). However, Barro and King (1984) show that if consumption and leisure are both normal goods, investment demand shocks fail to generate a positive co-movement among consumption, investment, and hours in standard neoclassical models. The intuition is simple. If an increase in investment causes an increase in hours, in the short run, the marginal product of labor falls and the marginal dis-utility of labor rises relative to the marginal utility of consumption. The former effect can be regarded as a result of capital resource tightness, while the latter effect can be interpreted as a result of labor resource tightness. As it becomes more difficult to obtain goods after an increase in investment, consumption is crowded out. This mechanism is known as the Barro-King curse in the literature.

Consumption demand shocks, such as taste shocks that change the marginal utility of consumption relative to the marginal dis-utility of labor are not subject to the Barro-King curse. If the desire for consumption increases, it is worthwhile for the economy to work more. With a Cobb-Douglas production technology, labor can substitute capital fairly easily. Thus, enough output can be squeezed out for consumption when more hours are worked. If the increase in consumption demand is persistent enough, firms would also have an incentive to invest for future, making consumption, investment, and hours to move in the same direction.

However, the lack of resource slackness still limits the role of consumption demand shocks in driving business cycles. Because of resource tightness, it is more difficult to obtain goods in boom and much easier to obtain goods in recession; thus, the response of output to demand shocks is dampened. Furthermore, because of capital resource tightness, the marginal product of labor is countercyclical and the real wage rate is countercyclical, while most of the evidence in the literature suggests that the real wage rate is either acyclical or pro-cyclical (e.g., Bils, 1985, Solon et al., 1994, and Brandolini, 1995). Finally, the labor productivity is countercyclical and the Solow residual is acyclical in the standard neoclassical models under demand shocks.

Facing these difficulties, it is reasonable to think that if capital or capacity underutilization can be introduced explicitly, we may be able to describe an economy with capital resource slackness. Hence, the problems associated with demand-driven business cycles can be alleviated or even removed.

Four different ways to model capital or capacity utilization existing in the literature are reviewed in this chapter. First, capital may not be fully utilized because of a convex utilization cost (e.g., Christiano et al., 2005 and Smets and Wouters, 2007). Second, if plants are subject to idiosyncratic productivity shocks, plants with a low productivity ex post may be left idle (e.g., Cooley et al., 1995). Third, if firms face idiosyncratic demand shocks, extra capacity is held as a precaution in case demand is high and firms that experience a low demand ex post can have their capacity underutilized (e.g., Fagnart et al., 1997 and Fagnart et al., 1999). Finally, if there are matching frictions in the goods market, capacity expansion can reduce the tightness of the goods market and lower the purchasing cost for buyers. Thus, firms have an incentive to supply extra capacity (e.g., Bai et al., 2012).

A common feature to these models is that capital, plant, or capacity is underutilized for a cost-saving reason. In the variable capital utilization model, capital is unused because of an increasing marginal cost of utilization. In the variable plant utilization model, plants are left idle because the productivities of those plants are too low to compensate the cost of labor. In the precautionary capacity model, extra capacity saves the expected cost incurred to satisfy a unit of aggregate demand. In the competitive goods market search model, extra capacity reduces market tightness and saves the purchasing cost for buyers. Hence, none of these underutilization phenomena would mean that too much capital is accumulated. In fact, I show that the capital stock chosen by firms minimizes the present value of aggregate costs. This is true even when there are monopolistic distortions. In the sense that the present value of aggregate costs is minimized, I say that capital accumulation is *partially efficient* and these models are called efficient utilization models.[1]

I prove that as long as capital accumulation is partially efficient, *aggregate capacity* must be fully utilized in the long run and capital resources must be tight locally around the steady state. Hence, the efficient utilization models fail to incorporate chronic excess capacity nor capital resource slackness. As a result, they tend to face the same issues as the standard real business cycles (RBC) model does. For example, the response of output to demand shocks is dampened and the real wage rate is countercyclical. Although the magnitude of capital resource tightness can be made arbitrarily small in efficient utilization models, making the real wage rate less countercyclical and consumption more responsive, the response of investment to consumption demand shocks is inherently small. Intuitively, precisely because capital resources are tight, there is a pro-cyclical upward pressure on the real marginal cost that firms want to adjust their capital stock to mitigate. Thus, if capital resources are not tight, firms will have no incentive to adjust capital. It turns out that to generate a large relative volatility of investment to consumption under consumption demand shocks, the efficient utilization models have to rely on capital resource tightness to dampen the response of consumption and to provide an incentive for firms to adjust capital. This, however, would cause the real wage rate to be too countercyclical.

Because of these difficulties, the existing literature has mainly turned to supply channels or use multiple types of shocks to generate stylized business cycle fluctuations. For example, Smets and Wouters (2007) emphasize on various types of supply shocks such as productivity shocks and markup shocks. Greenwood et al. (2000) and Jaimovich and Rebelo (2009) assume that if capital is used more intensively, it depreciates increasingly faster. This depreciation in use assumption allows investment specific technology shocks to work as labor productivity shocks.[2] In Khan and Tsoukalas (2011) and Ascari et al. (2019), nominal

[1] The name, "partial efficiency", is based on the idea that capital accumulation is efficient in a partial equilibrium, though not efficient in the general equilibrium.

[2] Johnson (1994) uses post-war data on the U.S. manufacturing sector to test the standard depreciation

rigidities and accommodating monetary policy allow a decrease in markup to boost aggregate supply after a positive investment demand shock.[3] Justiniano et al. (2010) rely on both productivity shocks and investment specific technology shocks to generate business cycles. In Bai et al. (2012), the matching technology and the production technology combined is of increasing returns to scale, and the model relies on both consumption demand shocks and investment demand shocks to generate business cycle fluctuations.

The capacity underutilization (CU) model developed in Chapter 1 is not subject to the difficulties that efficient utilization models have to face. In the CU model, firms are encouraged to over-invest in capital because capacity expansion can steal profitable demand from other firms. I show that because of this capacity competition externality, capital accumulation is not partially efficient, the economy exhibits chronic excess capacity, and capital resources can be slack. A calibration exercise is conducted to show that consumption demand shocks alone can generate realistic business cycle fluctuations in the CU model. The theoretical finding that a single type of demand shock can explain most of the business cycle fluctuations resonates with two recent empirical papers (e.g., Andrle et al., 2017 and Angeletos et al., 2018). The results suggest that the capacity competition externality is a powerful mechanism that can alleviate the dependence on supply channels to drive business cycles, reviving the original idea proposed by Keynes.

The rest of the chapter is organized as follows. Section 2.2 introduces a standard one-good economy in which capital accumulation is partially efficient, a framework that is often used in the business cycle literature. Section 2.3 reviews the four efficient utilization models and shows that they are special cases of the standard one-good economy. Section 2.4 studies the properties of the standard one-good economy under consumption demand shocks. First, I prove that when capital accumulation is partially efficient, aggregate capacity must be fully utilized in steady state and capital resources must be tight locally around the steady state. I then illustrate the difficulties of demand driving business cycles in the standard one-good economy. Particularly, either investment is not volatile enough or the real wage rate is too countercyclical under consumption demand shocks. In addition, the labor productivity is too countercyclical and the Solow residual is not pro-cyclical enough. Section 2.5 presents and calibrates a capacity underutilization model and shows

in use (DIU) model and finds that DIU is not the source of the observed variations in utilization. It is also questionable whether depreciation is an important *variable* cost in the real world. A convex variable cost of depreciation would imply a pro-cyclical share of depreciation cost in gross domestic product (GDP). The ratio of consumption of fixed capital (CFC) to GDP published by the Bureau of Economic Analysis, however, is strongly countercyclical. The usual business practice also considers depreciation as mainly a fixed cost rather than a variable cost.

[3]Whether markups are countercyclical or not, however, is controversial in the literature (e.g., Nekarda and Ramey, 2013).

that the inclusion of a capacity competition externality is a possible solution to the above difficulties. Section 2.6 concludes.

2.2 A Standard One-Good Economy

In this section, I present a standard one-good economy model in which monopolistic distortions and price stickiness are the only sources of inefficiency. However, capital accumulation decisions are not directly distorted. In this sense, I say that capital accumulation is *partially* efficient. The economy is then viewed from an aggregate demand-aggregate supply (AD-AS) perspective. Within this framework, monetary demand shocks, real demand shocks, and supply shocks are formally defined. The definitions are consistent with the common notions in the business-cycle literature. I show that with nominal rigidities, monetary demand shocks work via a supply channel because they affect real aggregate supply by affecting markups. Without nominal rigidities, the standard one-good economy is subject to the Barro-King curse. At the end of this section, a standard RBC model is calibrated and simulated as an example to illustrate the business-cycle properties of the standard one-good economy under different types of shocks.

2.2.1 Setup

Households

Consider a continuum of identical households indexed by $h \in [0, 1]$. Each provides a differentiated labor service. The labor demanded by firms is an aggregation of the differentiated labor services

$$l_t = \left(\int_0^1 l_{h,t}^{\frac{\varepsilon_{w,t}-1}{\varepsilon_{w,t}}} dh \right)^{\frac{\varepsilon_{w,t}}{\varepsilon_{w,t}-1}}, \tag{2.1}$$

where $l_{h,t}$ is the amount of labor supplied by household h, l_t is the amount of aggregate labor, and $\varepsilon_{w,t} > 1$ is the elasticity of substitution between any two types of differentiated labor services. Hence, the aggregation is of a constant elasticity of substitution (CES). The labor demand faced by household h is given by

$$l_{h,t} = l_t \left(\frac{w_{h,t}}{w_t} \right)^{-\varepsilon_{w,t}}, \tag{2.2}$$

where $w_{h,t}$ is the real wage rate charged by household h and $w_t = \left(\int_0^1 w_{h,t}^{1-\varepsilon_{w,t}} dh\right)^{\frac{1}{1-\varepsilon_{w,t}}}$ is the real wage rate of aggregate labor.

Each household aims to maximizes her expected lifetime utility

$$\max_{\{c_{h,t},l_{h,t},w_{h,t},b_{h,t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} \beta_{\tau-1}\right) \left(z_{c,t} u\left(\frac{c_{h,t}}{z_{c,t}}\right) + \bar{\omega}_t v\left(1-l_{h,t}\right)\right), \tag{2.3}$$

subject to her labor demand curve (2.2) and her budget constraint

$$c_{h,t} + b_{h,t+1} = w_{h,t} l_{h,t} + d_t + \frac{b_{h,t} P_{t-1}}{P_t} \mathcal{I}_t, \tag{2.4}$$

where $\beta_{t-1} > 0$ is the subjective discount factor that discounts the utility at time t to time $t-1$, $z_{c,t} > 0$ and $\bar{\omega}_t > 0$ are preference shocks, $c_{h,t}$ is the consumption of household h, $b_{h,t+1}$ is the real value of the risk free bonds that household h aims to carry to time $t+1$, d_t is the amount of dividends received from firms, P_t is the price of aggregate goods, and $\mathcal{I}_t$ is the gross nominal interest rate.

Let $\lambda_{h,t} \geq 0$ be the Lagrangian multiplier for the household's budget constraint. Since all households are identical, we have $\forall h$, $\lambda_{h,t} = \lambda_t$; and λ_t could be interpreted as the shadow price of aggregate goods measured in terms of utils, or simply, the price in utils.

Firms

There is a continuum of identical firms indexed by $j \in [0,1]$. Each firm produces a certain type of differentiated goods. The aggregate goods demanded by households or firms are a CES aggregation of the differentiated goods:

$$y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon_{p,t}-1}{\varepsilon_{p,t}}} dj\right)^{\frac{\varepsilon_{p,t}}{\varepsilon_{p,t}-1}}, \tag{2.5}$$

where $y_{j,t}$ is the amount of differentiated goods produced by firm j, y_t is the amount of aggregate goods, and $\varepsilon_{p,t} > 1$ is the elasticity of substitution between any two types of differentiated goods. The goods demand faced by firm j is given by

$$y_{j,t} = y_t \left(\frac{P_{j,t}}{P_t}\right)^{-\varepsilon_{p,t}}, \tag{2.6}$$

where $P_{j,t}$ is the price charged by firm j and $P_t = \left(\int_0^1 P_{j,t}^{1-\varepsilon_{p,t}} dj\right)^{\frac{1}{1-\varepsilon_{p,t}}}$ is the price of aggregate goods.

The amount of labor $l_{j,t}$ that needs to be hired by firm j to produce $y_{j,t}$ units of differentiated goods depends on the production technology used and the capital stock $k_{j,t}$ held:

$$l_{j,t} = z_{l,t}^{-1} L\left(y_{j,t}, k_{j,t}\right), \tag{2.7}$$

where $z_{l,t} > 0$ is the labor productivity and L is the effective labor required as a function of output and capital. Suppose that the labor cost is the only variable cost for the firm. The variable cost function of the firm can be written as:

$$C\left(y_{j,t}, k_{j,t}; z_{l,t}, w_t\right) = \frac{w_t}{z_{l,t}} L\left(y_{j,t}, k_{j,t}\right). \tag{2.8}$$

I assume that the production technology is such that the variable cost function C satisfies the following regularity conditions:

1. (*Standard properties*) The variable cost function C is non-negative, continuous, convex, increasing in terms of output, and decreasing in terms of capital.
2. (*Necessity of capital*) The domain of the variable cost function $\mathbf{dom}\left(C\right)$ is either given by $\{(y,k) \mid y \geq 0, k \geq \chi y\}$, where $\chi > 0$, or given by $\{(y,k) \mid y \geq 0, k > \chi y\} \cup \{(0,0)\}$, where $\chi \geq 0$. In the latter case, if $y > 0$ and $k \to \chi y$, the variable cost function converges to infinity: $C\left(y, k; z_l, w\right) \to \infty$.
3. (*Constant returns to scale*) The variable cost function C is homogeneous of degree one.
4. (*Smoothly diminishing marginal returns to capital*) If for a given y, there exists an interval $[a, b]$ on which the variable cost function is strictly decreasing in capital, then C is twice differentiable in capital k on the interval (a, b) such that $C_k\left(y, k; z_l, w\right) < 0$ and $C_{kk}\left(y, k; z_l, w\right) > 0$.

Note that if $\chi > 0$, there is a production limit which is the smallest upper bound on output: $y \leq \chi^{-1} k$. The Leontief production technology, the Cobb-Douglas production technology, and the production technology with a constant elasticity of substitution (CES) between capital and labor smaller than one, are all special cases of the technology described above.

Each firm aims to maximize the present value of dividend flows $d_{j,t}$:

$$\max_{\{y_{j,t}, P_{j,t}, d_{j,t}, i_{j,t} k_{j,t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} Q_\tau\right) d_{j,t}, \tag{2.9}$$

subject to the demand curve (2.6), the resource constraint

$$d_{j,t} + i_{j,t} = \frac{P_{j,t}}{P_t} y_{j,t} - C\left(y_{j,t}, k_{j,t}; z_{l,t}, w_t\right) - \frac{1}{2}\phi_p y_t \left(\frac{P_{j,t}}{P_{j,t-1}} - 1\right)^2, \tag{2.10}$$

and the law of motion for capital

$$k_{j,t+1} = k_{j,t}\left(1 - \delta\right) + z_{i,t} i_{j,t}, \tag{2.11}$$

where $Q_{t+1} = \beta_t {}^{\lambda_{t+1}}/_{\lambda_t}$ is the stochastic discount factor that discounts the value at time $t+1$ to time t, $i_{j,t}$ is the investment made by firm j, $\phi_p \geq 0$ is a parameter that controls the price stickiness, $\delta \in (0, 1]$ is the depreciation rate, and $z_{i,t} > 0$ is an investment specific technology shock.

Monetary Policy, Aggregation, and Market Clearing

The central bank sets the gross nominal interest rate according to a simple Taylor rule:

$$\mathcal{I}_t = z_{m,t} \Pi_t^{\phi_\pi} \left(\frac{y_t}{y}\right)^{\phi_y}, \tag{2.12}$$

where $z_{m,t}$ is a monetary policy shock, $\Pi_t \equiv \frac{P_t}{P_{t-1}}$ is the gross inflation rate, $\phi_\pi > 1$ and $\phi_y \geq 0$ are parameters, and y is the aggregate output in steady state.

The aggregate demand should be equal to the aggregate supply:

$$c_t + i_t + \frac{1}{2}\phi_p y_t \int_0^1 \left(\Pi_{j,t} - 1\right)^2 dj = y_t, \tag{2.13}$$

where $c_t = \int_0^1 c_{h,t} dh$ is the aggregate consumption, $i_t = \int_0^1 i_{j,t} dj$ is the aggregate investment, and $\Pi_{j,t} \equiv \frac{P_{j,t}}{P_{j,t-1}}$ is the gross inflation rate of firm j.

The aggregate labor demand should be equal to the aggregate labor supply: $\int_0^1 l_{j,t} dj = l_t$.

The risk free bonds are of zero net supply: $\int_0^1 b_{h,t} dh = 0$.

The total amount of dividends received by households is equal to the total amount of dividends distributed out by firms: $d_t = \int_0^1 d_{j,t} dj$.

Finally, the eight exogenous shocks, $z_{c,t}$, $\bar{\omega}_t$, $\varepsilon_{p,t}$, $\varepsilon_{w,t}$, $z_{l,t}$, $z_{i,t}$, β_t, and $z_{m,t}$, are mutually independent and each follows a log-AR(1) process.

2.2.2 Log-linearized Symmetric Equilibrium

Consider a symmetric equilibrium. If a variable is of the form $x_{j,t}$, we have $x_{i,t} = x_{j,t}$ for all $i, j \in [0,1]$. Hence, after obtaining the equilibrium conditions, we can omit the subscripts that index a variable to a particular household or firm. Furthermore, if a variable is of the form x_t, its value in steady state is denoted by x. Let $\hat{x}_t \equiv \ln x_t - \ln x$ denote the log-deviation of x_t from its steady state x. The log-linearized symmetric equilibrium is a stable stochastic process of nine variables ($\hat{c}_t$, $\hat{\lambda}_t$, $\hat{i}_t$, $\hat{w}_t$, $\hat{l}_t$, $\hat{k}_{t+1}$, $\hat{y}_t$, $\hat{\mathcal{I}}_t$, and $\hat{\Pi}_t$) that satisfies the household's consumption condition

$$\gamma\left(\hat{z}_{c,t} - \hat{c}_t\right) = \hat{\lambda}_t, \tag{2.14}$$

where $\gamma \equiv -\frac{u''(c/z_c)c}{u'(c/z_c)z_c} > 0$ is the inverse of the elasticity of inter-temporal substitution, the household's labor supply condition

$$\hat{\lambda}_t + \hat{w}_t = \hat{\bar{\omega}}_t + \eta\hat{l}_t + \hat{\mu}_{w,t}, \tag{2.15}$$

where $\eta \equiv \frac{v''(1-l)(-l)}{v'(1-l)} > 0$ is the inverse of the Frisch elasticity of labor supply and

$$\hat{\mu}_{w,t} = -\frac{1}{\varepsilon_w - 1}\hat{\varepsilon}_{w,t}, \tag{2.16}$$

is the log deviation of the wage markup, the firm's pricing condition

$$0 = \hat{w}_t - \hat{z}_{l,t} + \zeta\left(\hat{y}_t - \hat{k}_t\right) + \hat{\mu}_{p,t}, \tag{2.17}$$

where $\zeta \equiv \frac{L_{yy}(y,k)y}{L_y(y,k)} > 0$ measures the convexity of the variable cost function (2.8) in steady state and

$$\hat{\mu}_{p,t} = -\frac{1}{\varepsilon_p - 1}\left(\hat{\varepsilon}_{p,t} + \phi_p\left(\hat{\Pi}_t - \beta\mathbb{E}_t\left(\hat{\Pi}_{t+1}\right)\right)\right) \tag{2.18}$$

is the log deviation of the price-cost markup, the firm's investment condition

$$\begin{aligned}\hat{\lambda}_t - \hat{z}_{i,t} = \hat{\beta}_t + \mathbb{E}_t\left(\hat{\lambda}_{t+1}\right) + \beta\frac{wl}{z_i^{-1}k}\zeta\mathbb{E}_t\left(\hat{y}_{t+1} - \hat{k}_{t+1}\right) \\ - \beta\left(r+\delta\right)\mathbb{E}_t\left(\hat{\mu}_{p,t+1}\right) - \beta\left(1-\delta\right)\mathbb{E}_t\left(\hat{z}_{i,t+1}\right),\end{aligned} \tag{2.19}$$

where $r \equiv \beta^{-1} - 1 > 0$ is the real interest rate in steady state, the amount of labor required for production

$$\hat{l}_t = -\hat{z}_{l,t} + \nu\left(\hat{y}_t - \hat{k}_t\right) + \hat{k}_t, \tag{2.20}$$

where $\nu \equiv \frac{L_y(y,k)y}{L(y,k)} > 0$ is the output elasticity of labor demand in steady state, the law of motion for capital

$$\hat{k}_{t+1} = \hat{k}_t (1-\delta) + \delta \left(\hat{\imath}_t + \hat{z}_{i,t} \right), \tag{2.21}$$

the aggregate resource constraint

$$\frac{c}{y}\hat{c}_t + \frac{i}{y}\hat{\imath}_t = \hat{y}_t. \tag{2.22}$$

the Fisher's equation

$$\hat{\lambda}_t = \hat{\beta}_t + \mathbb{E}_t \left(\hat{\lambda}_{t+1} \right) + \hat{\mathcal{I}}_t - \mathbb{E}_t \left(\hat{\Pi}_{t+1} \right), \tag{2.23}$$

and the monetary policy

$$\hat{\mathcal{I}}_t = \phi_\pi \hat{\Pi}_t + \phi_y \hat{y}_t + \hat{z}_{m,t}. \tag{2.24}$$

2.2.3 Sources of Inefficiency

Monopolistic power and price stickiness are the only sources of inefficiency. If markets are perfectly competitive and prices are perfectly flexible, there is no resource loss due to price adjustment costs and both the price-cost markup $\mu_{p,t}$ and the wage markup $\mu_{w,t}$ will always be one. In this perfectly competitive case, one can verify that the real allocations are efficient by comparing them with those implied by the corresponding social planner's problem:

$$\max_{\{c_t, y_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} \beta_{\tau-1} \right) \left(z_{c,t} u \left(\frac{c_t}{z_{c,t}} \right) + \bar{\omega}_t v \left(1 - z_{l,t}^{-1} \mathcal{L} \left(y_t, k_t \right) \right) \right), \tag{2.25}$$

subject to $c_t + z_{i,t}^{-1} k_{t+1} = y_t + z_{i,t}^{-1} k_t (1-\delta)$, where $\mathcal{L}$ is the effective labor required by the social planner as a function of aggregate output and capital:

$$\mathcal{L}(y,k) \equiv \min_{\{y_j, k_j\}_{j\in[0,1]}} \int_0^1 L(y_j, k_j)\, dj, \tag{2.26}$$

subject to the CES aggregation (2.5) and the capital stock available $\int_0^1 k_j dj \leq k$. Note that the above minimization problem is symmetric and convex. Thus, a symmetric allocation solves the minimization problem and the required labor function of the social planner is the same as the required labor function of firms: $\mathcal{L}(y,k) = L(y,k)$.

2.2.4 Partially Efficient Capital Accumulation

Despite monopolistic power and price stickiness, the capital accumulation decisions made by firms are not directly distorted. In this sense, capital accumulation is partially efficient.

Definition 2.1. (*Partially efficient capital accumulation*) Capital accumulation is said to be partially efficient if and only if the capital stock chosen by firms in a decentralized way minimizes the present value of aggregate costs:

$$\min_{\{i_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} Q_\tau \right) \left(\mathcal{C}\left(y_t, k_t; z_{l,t}, w_t\right) + i_t \right), \tag{2.27}$$

subject to the law of motion for aggregate capital $k_{t+1} = k_t (1 - \delta) + i_t$, where $\mathcal{C}$ is the aggregate variable cost as a function of aggregate demand and capital.

The aggregate variable cost function of the standard one-good economy is given by

$$\mathcal{C}(y, k; z_l, w) = \min_{\{y_j, k_j\}_{j \in [0,1]}} \int_0^1 \frac{w}{z_l} L(y_j, k_j)\, dj, \tag{2.28}$$

subject to the CES aggregation (2.5) and the capital stock available $\int_0^1 k_j dj \leq k$.

Since the problem is symmetric and convex, a symmetric allocation solves the minimization problem, and the aggregate variable cost function is the same as the variable cost function at the firm-level: $\mathcal{C}(y, k; z_l, w) = C(y, k; z_l, w)$. Therefore, the capital stock chosen by firms in a decentralized way also minimizes the present value of aggregate costs.

The partial efficiency of capital accumulation does not mean that the aggregate capital stock in equilibrium will be at the efficient level because monopolistic power and price stickiness can still distort capital accumulation *indirectly* by affecting the stochastic discount factor, the real wage rate, and the aggregate demand. However, the fact that the capital accumulation decisions are not distorted directly has important consequences. It will be shown later in section 2.4 that the partial efficiency of capital accumulation, together with regularity conditions 1-4, rules out chronic excess capacity and capital resource slackness.

2.2.5 AD-AS System

The equilibrium system can be viewed from an aggregate demand-aggregate supply (AD-AS) perspective, which is useful to clarify the concepts of demand and supply.

Real Aggregate Demand

The household's consumption condition (2.14) characterizes the consumption demand, which describes a negative contemporaneous relationship between the price in utils $\hat{\lambda}_t$ and consumption $\hat{c}_t$.

Substitute the capital law of motion (2.21) into the firm's investment condition (2.19). We get an equation that characterizes the investment demand, which describes a negative contemporaneous relationship between the price in utils $\hat{\lambda}_t$ and investment $\hat{i}_t$:

$$\hat{\lambda}_t = -z_i \beta \frac{wl}{k} \zeta \left(\delta \hat{i}_t + (1-\delta) \hat{k}_t \right) + \mathbb{E}_t \left(\Theta_{t+1} \right) + \Psi_t, \tag{2.29}$$

where Θ_{t+1} is a combination of future prices and output and Ψ_t is a combination of various exogenous shocks.[4]

Combining the consumption demand condition (2.14), the investment demand condition (2.29), and the aggregate resource constraint (2.22) gives us the real aggregate demand (AD) condition

$$\begin{aligned} \left(1 + z_i \beta \frac{wl}{k} \zeta \frac{\delta}{\gamma} \frac{c}{i}\right) \hat{\lambda}_t = & -z_i \beta \frac{wl}{k} \zeta \left(\delta \frac{y}{i} \hat{y}_t + (1-\delta) \hat{k}_t \right) \\ & + \mathbb{E}_t \left(\Theta_{t+1} \right) + \Psi_t + z_i \beta \frac{wl}{k} \zeta \delta \frac{c}{i} \hat{z}_{c,t}, \end{aligned} \tag{2.30}$$

which describes a negative contemporaneous relationship between the price in utils $\hat{\lambda}_t$ and the aggregate demand $\hat{y}_t$.

Monetary Aggregate Demand

Combining the Fisher's equation (2.23) and the monetary policy (2.24) gives us the monetary AD condition

$$\phi_\pi \hat{\Pi}_t = -\phi_y \hat{y}_t + \hat{\lambda}_t + \mathbb{E}_t \left(\hat{\Pi}_{t+1} - \hat{\lambda}_{t+1} \right) - \hat{\beta}_t - \hat{z}_{m,t}, \tag{2.31}$$

which describes a negative contemporaneous relationship between the monetary price $\hat{\Pi}_t$ and the aggregate demand $\hat{y}_t$.

[4] $\Theta_{t+1} \equiv \hat{\lambda}_{t+1} + z_i \beta \frac{wl}{k} \zeta \hat{y}_{t+1} + \frac{\beta(r+\delta)}{\varepsilon_p - 1} \phi_p \left(\hat{\Pi}_{t+1} - \beta \hat{\Pi}_{t+2} \right)$ and $\Psi_t \equiv \hat{\beta}_t + \hat{z}_{i,t} \left(1 - z_i \beta \frac{wl}{k} \zeta \delta \right) - \beta (1-\delta) \mathbb{E}_t \left(\hat{z}_{i,t+1} \right) + \frac{\beta(r+\delta)}{\varepsilon_p - 1} \mathbb{E}_t \left(\hat{\varepsilon}_{p,t+1} \right)$.

Aggregate Supply

The aggregate supply (AS) condition can be obtained by combining the firm's pricing condition (2.17), the required labor for production (2.20), and the household's labor supply condition (2.15):

$$\hat{\lambda}_t + \frac{\phi_p}{\varepsilon_p - 1}\hat{\Pi}_t = (\zeta + \eta\nu)\left(\hat{y}_t - \hat{k}_t\right) + \eta\hat{k}_t + \beta\mathbb{E}_t\left(\frac{\phi_p}{\varepsilon_p - 1}\hat{\Pi}_{t+1}\right) - \frac{1}{\varepsilon_p - 1}\hat{\varepsilon}_{p,t} - (1+\eta)\,\hat{z}_{l,t} - \frac{1}{\varepsilon_w - 1}\hat{\varepsilon}_{w,t} + \hat{\bar{\omega}}_t, \tag{2.32}$$

which shows a positive contemporaneous relationship between the price in utils $\hat{\lambda}_t$ and the aggregate supply $\hat{y}_t$. If prices are sticky ($\phi_p > 0$), the AS condition (2.32) also shows a positive contemporaneous relationship between the monetary price $\hat{\Pi}_t$ and the aggregate supply $\hat{y}_t$. Hence, the AS condition is also known as the New Keynesian Philips Curve (NKPC) in the business cycle literature (e.g., Woodford, 2005).

Reduced Equilibrium System

The equilibrium system can now be reduced to a stable stochastic process of four variables ($\hat{y}_t$, $\hat{\lambda}_t$, $\hat{\Pi}_t$, and $\hat{k}_{t+1}$) that satisfies the real AD condition (2.30), the monetary AD condition (2.31), the AS condition (2.32), and the law of motion for capital (2.21).

For a given set of state variables, the short run equilibrium is characterized by $\hat{y}_t$, $\hat{\lambda}_t$, and $\hat{\Pi}_t$, and is pinned down by the real AD (2.30), the monetary AD (2.31), and the AS (2.32) conditions, which are surfaces in the output-price in utils-monetary price space. For example, Figure 2.1 shows these three surfaces in a standard one-good economy. The upper right panel of Figure 2.1 shows that a positive shock to the nominal interest rate moves the monetary AD surface inward, causing a lower inflation rate. The intersection of the real AD surface and the monetary AD surface projected on the output-price in utils plane gives us a downward sloping real AD curve. Since the real AD surface is orthogonal to the output-price in utils plane, a shift in the monetary AD surface has no effect on the real AD curve. The intersection curve of the AS surface and the monetary AD surface projected on the output-price in utils plane gives us an upward sloping real AS curve. When there are nominal rigidities, moving the monetary AD surface inward would cause a lower output and move the real AS curve to the left.

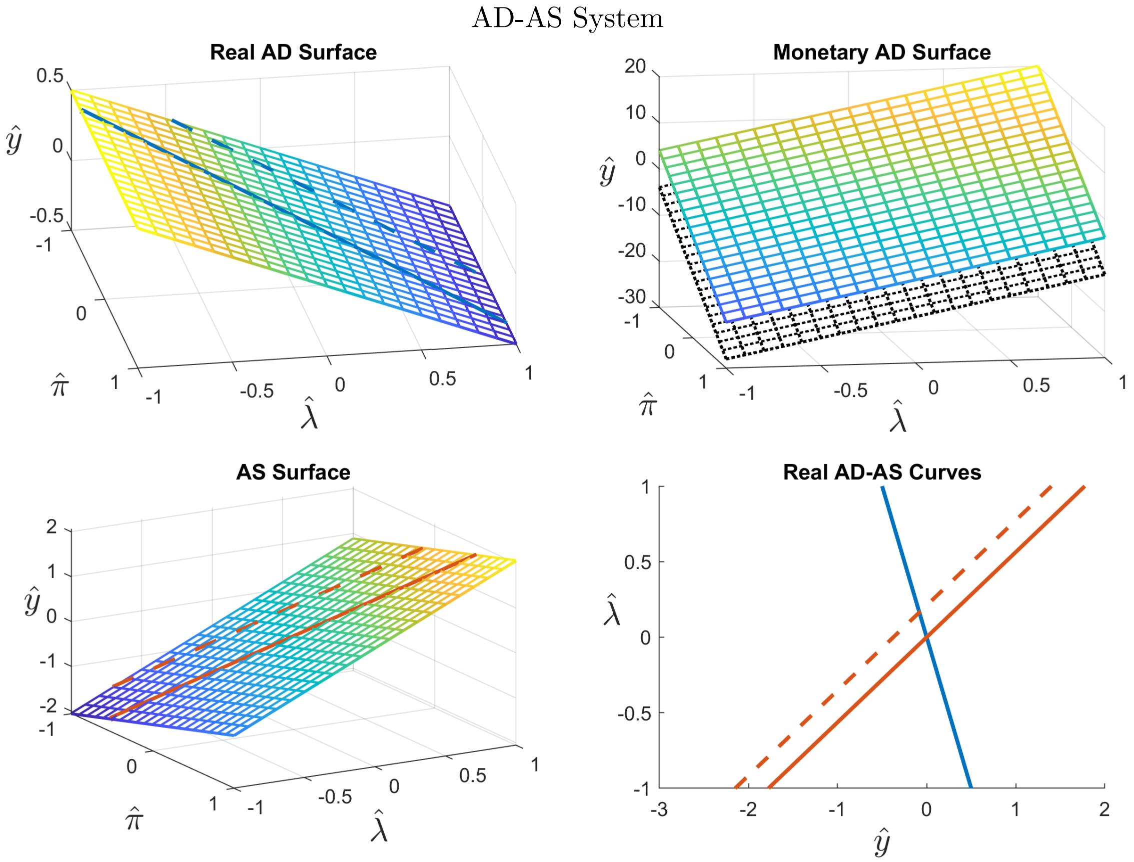

Figure 2.1: The solid surfaces and the solid lines represent a standard one-good economy in steady state. The dotted surface and the dashed lines represent the same one-good economy with a positive shock to the nominal interest rate that shifts the monetary AD surface inward. Each line on a surface is the intersection line of the surface and the monetary AD surface.

Demand and Supply

Definition 2.2. (*Demand shocks and supply shocks*) A real demand shock is a shock that disturbs the real AD condition (2.30). A monetary demand shock is a shock that disturbs the monetary AD condition (2.31). A supply shock is a shock that disturbs the AS condition (2.32).

I now classify the eight exogenous shocks based on the above definition.

An increase in the marginal utility of consumption $\hat{z}_{c,t}$ is a positive real demand shock as it increases the consumption demand.

An increase in the dis-utility of labor $\hat{\bar{\omega}}_t$ is a negative supply shock as it reduces the labor supply.

An increase in patience $\hat{\beta}_t$ is both a positive real demand shock as it increases investment demand and a negative monetary demand shock as it creates a deflationary pressure by making money hoarding more attractive.

An increase in the labor demand elasticity $\hat{\varepsilon}_{w,t}$, which is similar to a decrease in the labor dis-utility, is a positive supply shock as it increases the supply of labor by lowering the wage markup.

An increase in the goods demand elasticity $\hat{\varepsilon}_{p,t}$ is a positive supply shock as it lowers the price-cost markup. In addition, if the increase in the goods demand elasticity $\hat{\varepsilon}_{p,t}$ is persistent, the expected increase of $\hat{\varepsilon}_{p,t+1}$ is also a positive real demand shock as it increases the investment return by reducing the price-cost markup in the future.

An improvement in the investment specific technology $\hat{z}_{i,t}$ could be a positive or a negative real demand shock depending on the parameter values. On one hand, the improved investment specific technology lowers the cost of investment and boosts investment demand. On the other hand, the improved investment specific technology makes the investment more efficient; thus, less investment is required to achieve the desired capital level in the future. For example, if the convexity of the variable cost function in steady state ζ is low enough, the first effect dominates and the improved investment specific technology works as a positive real demand shock. However, if ζ is sufficiently high, the second effect dominates and the improved investment specific technology becomes a negative real demand shock.

An increase in the labor productivity $\hat{z}_{l,t}$ is clearly a positive supply shock.

Finally, an increase in the nominal interest rate $\hat{z}_{m,t}$ is a negative monetary demand shock as it creates a deflationary pressure.

2.2.6 New Keynesian Mechanism

The AD-AS framework illustrates how monetary demand shocks affect the real economy. Consider a positive shock to the nominal interest rate, which shifts inward the monetary AD surface and creates a deflationary pressure. The intersection curve of the monetary AD surface and the AS surface moves inward along the AS surface. When there are nominal rigidities ($\phi_p > 0$), the intersection curve falls and corresponds to a lower level of output (see the lower left panel of Figure 2.1). Hence, the real AS curve, which is the projection of the intersection curve on the output-price in utils plane, moves to the left (see the lower right panel of Figure 2.1). In this sense, a negative monetary demand shock works like a negative shock on the real AS, causing a fall in output and a rise in price in utils. Intuitively, as it is costly to adjust prices, the deflationary pressure would cause an increase in the price-cost markup, exacerbating the monopolistic distortion that depresses the real AS.

2.2.7 Barro-King Curse

Without nominal rigidities, the standard one-good economy is subject to the Barro-King curse. Consider, for example, a positive investment demand shock, which shifts the real AD curve outward and moves the equilibrium along the upward sloping real AS curve. As we can see from the AS condition (2.32), the real AS curve is upward sloping because the variable cost function is convex ($\zeta > 0$) and because the utility of leisure is concave ($\eta \geq 0$). Hence, both output and the price in utils increase. According to the household's consumption condition (2.14), consumption must fall when the price in utils increases. But investment and hours must rise because output increases. Therefore, an increase in investment demand crowds out consumption, causing consumption, investment, and hours to move in different directions.

2.2.8 Standard RBC Model: An Example

To end this section, I use a standard RBC model as an example to illustrate the business cycle properties of the one-good economy under different types of shocks.

Suppose that prices are perfectly flexible ($\phi_p = 0$) and all markets are perfectly competitive. Shocks to the price elasticity of goods demand and the wage elasticity of labor demand are no longer considered as both converge to infinity: $\varepsilon_p \to \infty$ and $\varepsilon_w \to \infty$. The production function for the standard RBC model is assumed to be Cobb-Douglas (CD): $y_t = (z_{l,t} l_t)^{\alpha} k_t^{1-\alpha}$, where $\alpha \in (0,1)$ is the CD share of labor. The required effective labor is given by $L(y,k) = y^{\frac{1}{\alpha}} k^{1-\frac{1}{\alpha}}$ and the convexity of the variable cost function in steady state is $\zeta = \frac{1-\alpha}{\alpha} > 0$.

Calibration

There are seven parameters remained to be calibrated. The depreciation rate δ is calibrated to match the average ratio of gross private domestic investment to private fixed assets from 1947 to 2016 in the National Income and Product Accounts (NIPA) published by the U.S. Bureau of Economic Analysis (BEA). The curvature of the consumption utility

Table 2.1: Parameters and Calibration Targets – Standard RBC

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Elasticity of inter-temporal substitution 1
η	0.0000	Frisch elasticity of labor supply ∞
α	0.6200	Labor share of income 0.62
β	0.9747	Investment to output ratio 0.17

γ is calibrated to give an elasticity of inter-temporal substitution of 1, which corresponds to a log utility of consumption. The curvature of the dis-utility of labor η is chosen to imply an infinite Frisch elasticity of labor supply as in the indivisible labor setup (e.g., Hansen, 1985 and Rogerson, 1988). This choice of η helps the RBC model to generate a sizable labor volatility under productivity shocks (e.g., King and Rebelo, 1999). The CD share of labor α is calibrated to match the average labor share of income estimated by the Bureau of Labor Statistics (BLS) from 1946 to 2016. The subjective discount factor in steady state β is calibrated to be such that the investment to output ratio in steady state is equal to the average of the gross private domestic investment to the gross domestic product (GDP) ratio from 1947 to 2016 in the NIPA.

Table 2.1 summarizes the calibrated parameter values and their mostly associated targets.

Quantitative Results and Discussions

Tables 2.2 compares the business cycle statistics in the calibrated RBC model under different shocks with those in the data. There are five stylized business-cycle facts that we aim to match simultaneously using a single type of shock. First, there is a co-movement between consumption, investment, and hours. Second, the volatility of investment is much larger than the volatility of consumption. Third, the real wage rate is not too countercyclical.[5] Fourth, the labor productivity measured as the output to labor ratio is not too countercyclical. Fifth, the Solow residual measured based on the average labor share of income is strongly pro-cyclical.

The investment specific technology shock $z_{i,t}$ and the patience shock β_t are both shocks to

[5]The cyclicality of the real wage rate shown in Table 2.2 is for the average real wage rate of the business sector deflated with the GDP deflator. For robustness, Appendix 4.6 shows the cyclicality of the average real wage rates of different sectors including the business sector, the no-farm sector, and the non-financial corporations. The average nominal wage rates are deflated using different deflators, e.g., the GDP deflator, the consumer price index (CPI), and the own sector deflator. The results show that the real wage rates on average are roughly acyclical. It is also well known in the literature that cyclicality of the average real wage rate tends to underestimate the pro-cyclicality of the real wage rate at the individual level because of a composition bias (e.g., Bils, 1985).

investment demand. Both are subject to the Barro-King curse and thus fail to generate the co-movement between consumption, investment, and hours. A positive investment demand shock, though induces more hours, inevitably crowds out consumption.

The consumption demand shock $z_{c,t}$ is not subject to the Barro-King curse because the marginal rate of substitution between consumption and leisure is directly affected. Hence, consumption and hours can move in the same direction. If the shock is persistent, the desire to invest could be strong enough so that there is a co-movement among consumption, investment, and hours. However, the relative volatility of investment to consumption is too small. Intuitively, capital resources are tight in the RBC model: a short run increase in output would cause an increase in the real marginal cost of production and a decrease in the marginal product of labor. Thus, it is not wise to concentrate investment in a single period. The tendency to smooth investment under consumption demand shocks causes a small volatility of investment relative to that of consumption.

The labor dis-utility shock $\bar{\omega}_t$ is able to generate a co-movement between consumption, investment, and hours and a large relative volatility of investment to consumption. However, like all other shocks that do not affect productivity, e.g., $z_{c,t}$, $z_{i,t}$, β_t, the labor dis-utility shock $\bar{\omega}_t$ generates a strongly countercyclical real wage rate, a strongly countercyclical labor productivity, and an acyclical Solow residual.

The labor productivity shock $z_{l,t}$ turns out to be the only type of shock that can generate all the stylized business-cycle facts. Therefore, the business cycle literature tends to consider productivity shocks as the primary driving forces of business cycles since RBC theory has been introduced by Kydland and Prescott (1982). This view, however, contradicts the Keynes tradition which emphasizes on real demand shocks as the main driving forces of business cycles. The new Keynesian (NK) literature developed based on RBC theory tries to re-introduce the importance of demand by including nominal rigidities. However, the standard NK literature emphasizes more on the effectiveness of monetary demand than on the importance of real demand. For shocks that directly affect the real economy, supply shocks, including productivity shocks and markup shocks, remain to be the most important (e.g., Smets and Wouters, 2007).

Table 2.2: Business Cycle Statistics – Standard RBC

	The U.S. Data	$z_{c,t}$		$z_{i,t}$		β_t		$\bar{\omega}_t$		$z_{l,t}$	
Persistence ρ		0.00	0.99	0.00	0.99	0.00	0.99	0.00	0.99	0.00	0.99
$\rho(c,h)$	0.58 $[0.42, 0.72]$	0.88	0.96	-0.99	-0.91	-0.99	-0.93	0.42	0.90	0.39	0.83
$\rho(c,i)$	0.68 $[0.57, 0.78]$	-1.00	0.93	-0.99	-0.90	-0.99	-0.92	0.40	0.86	0.40	0.86
σ_i/σ_c	6.49 $[5.55, 7.42]$	4.34	5.08	13.4	12.3	13.5	12.9	58.6	8.44	58.6	8.44
$\mathbf{Cov}(w,y)/\sigma_y^2$	0.08 $[-0.03, 0.18]$	-0.30	-0.60	-0.66	-0.54	-0.66	-0.55	-0.64	-0.59	0.05	0.43
$\mathbf{Cov}(LP,y)/\sigma_y^2$	0.11 $[0.00, 0.21]$	-0.30	-0.60	-0.66	-0.54	-0.66	-0.55	-0.64	-0.59	0.05	0.43
$\mathbf{Cov}(SR,y)/\sigma_y^2$	0.39 $[0.30, 0.49]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.63

Note: $\rho(x_1, x_2)$ stands for the correlation coefficient between x_1 and x_2. σ_i/σ_c is the relative volatility of investment to consumption. $\mathbf{Cov}(w,y)/\sigma_y^2$ stands for the covariance between x and output relative to the variance of output. The brackets are the 95% confidence intervals calculated using a parametric bootstrapping method (see Appendix 4.5 for details). The U.S. data is from the BEA and the BLS. A path of 5,000 quarters is simulated to calculate the statistics for each calibrated model. All variables are Hodrick-Prescott (HP) filtered logarithms of the original series.

2.3 Efficient Utilization Models: Special Cases of the Standard One-Good Economy

As consumption demand shocks are not subject to the Barro-King curse, they seem promising in reviving the idea of Keynes.[6] However, under consumption demand shocks, the relative volatility of investment to consumption tends to be too small because of capital resource tightness. If there is some unused capital or capacity, we might be able to live in an economy with capital resource slackness, and thus, the issues related to consumption demand shocks might be resolved. This possibility will be discussed in the rest of the chapter. From now on, I will focus on the role of consumption demand shocks and restrict myself to a real economy with perfectly flexible prices.

In this section, I introduce four different methods in the literature that model capital or capacity utilization explicitly. First, capital may not be fully utilized because of a convex utilization cost (e.g., Christiano et al., 2005 and Smets and Wouters, 2007). Second, if plants are subject to idiosyncratic productivity shocks, plants with a low productivity ex post may be left idle (e.g., Cooley et al., 1995). Third, if firms face idiosyncratic demand shocks, extra capacity is held as a precaution in case demand is high and firms that experience a low demand ex post can have their capacity underutilized (e.g., Fagnart

[6]There are models in the literature that can overcome the Barro-King curse (e.g., Jaimovich and Rebelo, 2009, Khan and Tsoukalas, 2011, and Ascari et al., 2019). Three methods are often used. First, the Jaimovich-Rebelo preferences are used to reduce the short run wealth effect on the supply of labor. However, as long as consumption and leisure are still normal goods, the Jaimovich-Rebelo preferences alone will not be enough to break the Barro-King curse. Second, depreciation in use is assumed so that the investment specific technology shock can partly work as a labor productivity shock. Third, production networks and trend growth rate are introduced to flatten the NKPC. From the perspective of the one-good economy developed here, the second methods adds a labor productivity shock, while the third strengthens the role of nominal rigidities that endogenously move markups. Hence, both work via a supply channel: an increase in real aggregate demand is simultaneously accompanied with an increase in real aggregate supply.

et al., 1997 and Fagnart et al., 1999). Finally, if there are matching frictions in the goods market, capacity expansion can reduce the tightness of the goods market and lower the purchasing cost for buyers. Thus, firms have an incentive to supply extra capacity (e.g., Bai et al., 2012).

However, despite unused capital, plants, or capacity at the firm-level, I show that the economy described by these methods has no fundamental difference from the standard one-good economy developed in section 2.2, the capital accumulation decisions made by firms are still partially efficient, and the whole economy can even be fully efficient if markets are perfectly competitive. Thus, the efficient utilization models would have to face the same issues as the standard one-good economy does.

2.3.1 Variable Capital Utilization

Technology and Cost Minimization

The production function of a representative firm is given by $F(\theta_t k_t, l_t)$, where θ_t is the capital utilization rate. The production function F is assumed to be twice differentiable, concave, and homogeneous of degree one. The derivatives of F satisfy: $F_K > 0$, $F_L > 0$, and $F_{KK} < 0$. Both the capital and the labor are assumed to be necessary such that $\forall x \geq 0$, $F(0, x) = F(x, 0) = 0$.

The cost of utilizing capital is assumed to be paid in the firm's own product and is given by $a(\theta_t) k_t$, which is non-negative $a \geq 0$, increasing $a' > 0$, and convex $a'' > 0$ in terms of the capital utilization rate. The representative firm acts as a price taker and chooses its labor and capital utilization rate to minimize its variable cost:

$$C(y_t, k_t; w_t) = \min_{\theta_t \geq 0, l_t \geq 0} w_t l_t, \tag{2.33}$$

subject to $y_t \leq F(\theta_t k_t, l_t) - a(\theta_t) k_t$. Suppose that there is an upper limit for the capital utilization rate $\bar{\theta}$. Because of the convex utilization cost, it could be optimal for firms to choose a utilization rate smaller than $\bar{\theta}$, generating a notion of capital underutilization.

Following the convention in the literature, I assume that the upper limit $\bar{\theta}$ is infinite, $a'(0) = 0$, and if $\theta \to \infty$, $a'(\theta) \to \infty$. These assumptions ensure the existence of an interior solution. One can show that the labor and the capital utilization rate that solve the above minimization problem exist uniquely. Both will be positive if $y_t > 0$. By the envelope theorem and the implicit function theorem, C is non-negative, twice differentiable, and homogeneous of degree one; and $C_y > 0$, $C_k < 0$, and $C_{kk} > 0$. Finally, the necessity of capital can be proved by contradiction.[7]

Partially Efficient Capital Accumulation

The aggregate variable cost incurred to satisfy the aggregate demand y_t is the same as the variable cost function of the representative firm: $\mathcal{C}(y_t, k_t; w_t) = C(y_t, k_t; w_t)$. Hence, the capital stock determined by firms in a decentralized way also minimizes the present value of aggregate costs, that is, the capital accumulation is partially efficient. With a standard household's problem as described in section 2.2, the variable capital utilization model developed here is a special case of the standard one-good economy.

Perfectly Competitive Markets and Efficiency

Assume further that all markets are perfectly competitive. The economy is efficient and can be summarized as a social planner's problem:

$$\max_{\{c_t, y_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(z_{c,t} u \left(\frac{c_t}{z_{c,t}} \right) + \bar{\omega} v \left(1 - \mathcal{L}(y_t, k_t)\right) \right), \tag{2.34}$$

subject to $c_t + k_{t+1} = y_t + k_t (1 - \delta)$, where the labor required by the social planner $\mathcal{L}$ is given by

$$\mathcal{L}(Y_t, k_t) = \min_{l_t \geq 0, \theta_t \geq 0} l_t \tag{2.35}$$

subject to $y_t \leq F(\theta_t k_t, l_t) - a(\theta_t) k_t$. Hence, the efficient variable utilization model works just like the RBC model except that the required labor function $\mathcal{L}$ is different.

2.3.2 Variable Plant Utilization

Technology and Cost Minimization

Consider a representative firm who owns a continuum of identical plants with measure m_t. A plant can either be operated or not. If a plant $j \in [0, m_t]$ is operated, it requires k_0 units of capital and h_0 units of hours as inputs and the output of the plant is $1 + s_{j,t}$,

[7] For a given positive aggregate demand $y > 0$, if there is an upper bound for $\mathcal{C}$ as $k \to 0$, l^* that solves the cost minimization problem has an upper bound: $l^* \leq \bar{l}$. Therefore, $y \leq F(\theta k, \bar{l}) - a(\theta) k \leq F(\theta^* k, \bar{l}) - a(\theta^*) k \leq F(\theta^* k, \bar{l})$, where $\theta^* > 0$ maximizes the net output: $F_K(\theta^* k, \bar{l}) = a'(\theta^*)$. By the implicit function theorem, θ^* decreases with k. If θ^* has an upper bound as $k \to 0$, then $\theta^* k \to 0$. If θ^* increases without limit as $k \to 0$, $a'(\theta^*)$ increases without limit. Since $F_K(\theta^* k, \bar{l}) = a'(\theta^*)$, we have again that $\theta^* k \to 0$. To sum up, $y \leq F(\theta^* k, \bar{l}) \to 0$, a contradiction to that $y > 0$.

where $s_{j,t}$ is a plant specific productivity shock distributed independently and identically distributed across time and plants. Following Cooley et al. (1995), $s_{j,t}$ is assumed to be uniformly distributed on the interval $[-\sigma, \sigma]$, where $\sigma \in (0,1)$ is a parameter that captures the volatility of the plant specific shock.

The capital stock of the plants have to be installed one period before the realizations of the plant specific shocks. Suppose that the representative firm carries k_t units of capital stock at the beginning of the period t. The measure of plants that the firm owns is given by $m_t = {}^{k_t}/_{k_0}$.[8] The representative firm chooses the fraction of plants operated $n_t \in [0,1]$ to minimize the variable cost. High productivity plants shall be operated first. Hence, for a given n_t, there is a threshold given by $s^*(n_t) = \sigma - 2\sigma n_t$: plants with a productivity larger than $s^*(n_t)$ will be operated and plants with a productivity smaller than $s^*(n_t)$ will be left idle. The cost minimization problem of the firm is summarized below:

$$C(y_t, k_t; w_t) \equiv \min_{n_t \in [0,1], l_t \geq 0} w_t l_t, \tag{2.36}$$

subject to the production constraint $y_t \leq k_t k_0^{-1} \int_{\sigma - 2\sigma n_t}^{\sigma} (1+s) \frac{1}{2\sigma} ds$ and the required labor for production $n_t k_t k_0^{-1} h_0 \leq l_t$. Because of the heterogeneous plant productivities, the higher the plant utilization rate is, the smaller the productivity of the marginal plant. Therefore, the firm may find it optimal to choose a utilization rate smaller than one, generating a notion of plant underutilization.

The labor and the plant utilization rate that solve the above minimization problem are uniquely pinned down by the two constraints:$y_t = k_t k_0^{-1} \left((1+\sigma) n_t - \sigma n_t^2 \right)$ and $l_t = n_t k_t k_0^{-1} h_0$. The former shows that the output is a strictly increasing and concave function of the plant utilization rate n_t. By the implicit function theorem, one can show that C is non-negative, twice differentiable, and homogeneous of degree one; and $C_y > 0$, $C_k < 0$, and $C_{kk} > 0$. In addition, there exists a constant $\chi = k_0 > 0$ that sets a production limit on output: $y_t \leq \chi^{-1} k_t = k_0^{-1} k_t$, which can be achieved if and only if all plants are utilized ($n_t = 1$).

[8] The original fixed-plant economy set up by Cooley et al. (1995) assumes that the measure of plants m_t is fixed and the capital stock k_t is used to increase the productivity of plants. However, since the measure of plants is fixed, the aggregate production function will be of decreasing returns to scale in terms of capital and labor. I restore constant returns to scale by assuming that the measure of plants is fixed in the short run but the capital stock is used to expand the measure of plants in the long run.

Partially Efficient Capital Accumulation

The aggregate variable cost incurred to satisfy the aggregate demand y_t is the same as the variable cost function of the representative firm: $\mathcal{C}(y_t, k_t; w_t) = C(y_t, k_t; w_t)$. Hence, the capital stock chosen by firms in a decentralized way also minimizes the present value of aggregate costs, that is, the capital accumulation is partially efficient. With a standard household's problem as described in section 2.2, the plant utilization model is a special case of the standard one-good economy.

Perfectly Competitive Markets and Efficiency

If all markets are perfectly competitive, the economy is efficient and can be summarized as a social planner's problem:

$$\max_{\{c_t, y_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(z_{c,t} u \left(\frac{c_t}{z_{c,t}} \right) + \bar{\omega} v \left(1 - \mathcal{L}(y_t, k_t) \right) \right), \tag{2.37}$$

subject to $c_t + k_{t+1} = y_t + k_t(1-\delta)$, where the labor required by the social planner $\mathcal{L}$ is given by

$$\mathcal{L}(y_t, k_t) = \min_{n_t \in [0,1], l_t \geq 0} l_t, \tag{2.38}$$

subject to the production constraint $y_t \leq k_t k_0^{-1} \left((1+\sigma) n_t - \sigma n_t^2 \right)$ and the required labor for production $n_t k_t k_0^{-1} h_0 \leq l_t$. Hence, the efficient plant utilization model works just like the RBC model except that the required labor function $\mathcal{L}$ is different.

2.3.3 Demand Uncertainty and Precautionary Capacity

Technology and Cost Minimization

There is a continuum of firms with measure one. Each is indexed by $j \in [0,1]$ and produces a certain type of differentiated goods. The aggregated goods demanded by households and firms is a CES aggregation of the differentiated goods:

$$y_t = \left(\int_0^1 v_{j,t}^{\frac{1}{\varepsilon}} y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon-1}{\varepsilon}}, \tag{2.39}$$

where $y_{j,t}$ is the amount of differentiated goods produced by firm j, $v_{j,t}$ is the firm specific quality shock that affects the demand faced by firm j, y_t is the amount of aggregate goods, and $\varepsilon > 1$ is the elasticity of substitution among the differentiated goods. The demand faced by firm j is given by

$$y_{j,t} = v_{j,t} p_{j,t}^{-\varepsilon} y_t, \tag{2.40}$$

where $p_{j,t} = {}^{P_{j,t}}/_{P_t}$ is the real price charged by firm j and $P_t = \left(\int_0^1 v_{j,t} P_{j,t}^{1-\varepsilon} dj\right)^{\frac{1}{1-\varepsilon}}$ is the aggregate price. Equation (2.40) shows that $v_{j,t}$ can also be interpreted as a demand shock for the firm. Following Fagnart et al. (1997), $v_{j,t}$ is assumed to be independently and identically distributed (i.i.d.) across time and firms, and is drawn from a log-normal distribution with mean of one and standard deviation of $\sigma > 0$.

Each firm operates with a Leontief production technology. The cost minimization problem of an individual firm is given by:

$$C\left(y_{j,t}, k_{j,t}; w_t\right) = \min_{l_{j,t} \geq 0} w_t l_{j,t}, \tag{2.41}$$

subject to the Leontief production function $y_{j,t} \leq \min\left\{{}^{l_{j,t}}/_{\alpha_v}, A k_{j,t}\right\}$, where $\alpha_v > 0$ is the required labor per output and $A > 0$ is the productivity of capital. The labor that solves the above minimization problem is uniquely pinned down by the required labor for production: $l_{j,t} = \alpha_v y_{j,t}$. Since capacity is installed one period before the realization of the firm specific quality shocks, if the realized shock to demand $v_{j,t}$ is low enough, some capacity will be left idle, generating a notion of capacity underutilization.

Aggregate Variable Cost

The aggregate variable cost incurred to satisfy the aggregate demand is not as simple as the variable cost function for an individual firm. For a given aggregate demand y_t and an aggregate capital stock k_t, the aggregate variable cost function takes all firms and all possible realizations of the quality shocks into account:

$$\mathcal{C}\left(y_t, k_t; w_t\right) = \min_{\left\{y_{j,t}(\mathbf{v}) \geq 0, k_{j,t} \geq 0\right\}_{j \in [0,1], \mathbf{v} \geq 0}} w_t \int_{\mathbf{v}} \int_0^1 \alpha_v y_{j,t}\left(\mathbf{v}\right) dj d\mathcal{F}\left(\mathbf{v}\right), \tag{2.42}$$

subject to a CES aggregation constraint

$$y_t \leq \int_{\mathbf{v}} \left(\int_0^1 \mathbf{v}\left(j\right)^{\frac{1}{\epsilon}} y_{j,t}\left(\mathbf{v}\right)^{\frac{\epsilon-1}{\epsilon}} dj\right)^{\frac{\epsilon}{\epsilon-1}} d\mathcal{F}\left(\mathbf{v}\right), \tag{2.43}$$

the capacity constraint $y_{j,t}\left(\mathbf{v}\right) \leq A k_{j,t}$, and the capital stock available $\int_0^1 k_{j,t} dj \leq k_t$, where $\mathbf{v}$ is a possible realization of the quality shocks, $\mathbf{v}\left(j\right)$ gives the realized quality of firm j, $y_{j,t}\left(\mathbf{v}\right)$ denotes the output allocated to firm j if the realized quality shocks are

$\mathbf{v}$, and $\mathcal{F}$ is the distribution of the quality shocks. Note that the capital allocation has to be made before the realization of the firm specific quality shocks. Hence, $k_{j,t}$ is not contingent on $\mathbf{v}$.

Because the above minimization problem is strictly convex and symmetric in j, an allocation that minimizes the aggregate variable cost must be symmetric in j such that $\forall j$, $y_{j,t}(\mathbf{v}) = y_t(\mathbf{v}(j))$, and $k_{j,t} = k_t$. Hence, by the law of large numbers, the expected labor cost (2.42) is reduced to $w_t \int_{\mathbf{V}} \int_0^1 \alpha_v y_t(\mathbf{v}(j))\, dj d\mathcal{F}(\mathbf{v}) = w_t \int_0^\infty \alpha_v y_t(v) f(v)\, dv$, where v is a possible realization of the quality shock of a firm and f is the probability density function of the quality shock. By the same argument, the CES aggregation constraint (2.43) is reduced to

$$y_t \leq \left(\int_0^\infty v^{\frac{1}{\epsilon}} y_t(v)^{\frac{\epsilon-1}{\epsilon}} f(v)\, dv \right)^{\frac{\epsilon}{\epsilon-1}}. \tag{2.44}$$

The cost minimization problem can be rewritten as

$$\mathcal{C}(y_t, k_t; w_t) = \min_{\{y_t(v)\geq 0\}_{v\geq 0}} w_t \int_0^\infty \alpha_v y_t(v) f(v)\, dv, \tag{2.45}$$

subject to the simplified CES aggregation constraint (2.44) and the capacity constraint $y_t(v) \leq A k_t$.

To solve the simplified minimization problem (2.45), note that there is a threshold $v_t^* \geq 0$ that determines whether the capacity constraint is binding or not. If a firm experiences a quality shock v that is larger than v_t^*, the output allocated to the firm will be equal to its capacity. If the firm experiences a quality shock v that is smaller than v_t^*, the output allocated to the firm will be smaller than its capacity and fall proportionally with v because it is now less attractive to use the goods from the firm to satisfy the aggregate demand. To sum up, we have

$$y_t(v) = \begin{cases} \frac{v}{v_t^*} A k_t, & v \leq v_t^*, \\ A k_t, & v > v_t^*. \end{cases} \tag{2.46}$$

A smaller v_t^* corresponds to a larger fraction of firms having their capacity fully utilized, and thus, a higher labor cost:

$$\mathcal{C}(y_t, k_t; w_t) = w_t \alpha_v \left(\frac{1}{v_t^*} \int_0^{v_t^*} v f(v)\, dv + \int_{v_t^*}^\infty f(v)\, dv \right) A k_t. \tag{2.47}$$

Therefore, v_t^* in optimum should be as large as possible, making the CES aggregation constraint (2.44) binding:

$$y_t = Ak_t \left((v_t^*)^{-\frac{\epsilon-1}{\epsilon}} \int_0^{v_t^*} vf(v)\,dv + \int_{v_t^*}^{\infty} v^{\frac{1}{\epsilon}} f(v)\,dv \right)^{\frac{\epsilon}{\epsilon-1}}. \tag{2.48}$$

By the implicit function theorem, one can show that $\mathcal{C}$ is non-negative, twice differentiable, and homogeneous of degree one; and $\mathcal{C}_y > 0$, $\mathcal{C}_k < 0$, and $\mathcal{C}_{kk} > 0$. In addition, there exists a constant $\chi > 0$ that sets a production limit: $y_t \leq \chi^{-1} k_t = \left(\int_0^{\infty} v^{\frac{1}{\epsilon}} f(v)\,dv \right)^{\frac{\epsilon}{\epsilon-1}} Ak_t$, which can be achieved if and only if all capacity is fully utilized ($v_t^* \to \infty$).

Partially Efficient Capital Accumulation

When each firm makes its own investment decision, it has to face some demand uncertainty as the firm specific quality shocks have not yet been realized. Hence, it is the gross profit in expectation that matters for capital accumulation. The gross profit of firm j is a function the firm's capital stock and the firm's realized quality shock:

$$\pi(k_{j,t}, v_{j,t}; w_t, y_t) = \max_{p_{j,t}\geq 0, y_{j,t}\geq 0} p_{j,t} y_{j,t} - C(y_{j,t}, k_{j,t}; w_t), \tag{2.49}$$

subject to the demand function (2.40) and the capacity constraint $y_{j,t} \leq Ak_{j,t}$. The expected gross profit is given by

$$\int_0^{\infty} \pi(k_{j,t}, v; w_t, y_t) f(v)\,dv, \tag{2.50}$$

which is strictly concave in capital. Although capacity may be underutilized in low demand cases, capacity expansion can still relax the capacity constraint if demand turns out to be high, justifying the value of extra capacity as a precaution. The same logic justifies the value of extra capacity in the aggregate variable cost minimization problem (2.42). Since all firms are identical ex ante, one can verify that in equilibrium, the return of capital viewed from a representative firm's perspective is the same as that viewed from an aggregate perspective:

$$\int_0^{\infty} \pi_k(k_t, v; w_t, y_t) f(v)\,dv = -\mathcal{C}_k(y_t, k_t; w_t). \tag{2.51}$$

Thus, the capital stock chosen by firms in a decentralized way also minimizes the present value of aggregate costs, that is, the capital accumulation is partially efficient. With a standard household's problem as described in section 2.2, the precautionary capacity model developed by Fagnart et al. (1997) is a special case of the standard one-good economy.

Perfectly Competitive Markets and Efficiency

If there are infinitely many firms that can produce the same differentiated goods j, the goods markets become perfectly competitive. If the labor market is also perfectly competitive, the economy is efficient and can be summarized as a social planner's problem:

$$\max_{\{c_t, y_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(z_{c,t} u \left(\frac{c_t}{z_{c,t}} \right) + \bar{\omega} v \left(1 - \mathcal{L}\left(y_t, k_t\right)\right) \right), \tag{2.52}$$

subject to $c_t + k_{t+1} = y_t + k_t (1 - \delta)$, where the labor required by the social planner $\mathcal{L}$ is given by

$$\mathcal{L}(y_t, k_t) = \min_{\{y_{j,t}(\mathbf{v}) \geq 0, k_{j,t} \geq 0\}_{j \in [0,1], \mathbf{v} \geq 0}} \int_{\mathbf{v}} \int_0^1 \alpha_v y_{j,t}(\mathbf{v})\, dj d\mathcal{F}(\mathbf{v}), \tag{2.53}$$

subject to the CES aggregation constraint (2.43), the capacity constraint $y_{j,t}(\mathbf{v}) \leq A k_{j,t}$, and the capital stock available $\int_0^1 k_{j,t} dj \leq k_t$. Hence, the efficient precautionary capacity model works just like the RBC model except that the required labor function $\mathcal{L}$ is different.

2.3.4 Competitive Search in the Goods Market

Technology and Cost Minimization

Suppose that demand has to be matched with supply before a unit of goods can be produced and sold. Let D_t denote the total amount of goods requested by buyers, or simply, the demand from buyers. Let S_t denote the total amount of goods that could potentially be produced by firms, or the total capacity supplied. The matching technology, which determines the amount of goods sold and produced, is given by the following matching function

$$y_t = M(D_t, S_t) \equiv \left(D_t^{\frac{\epsilon-1}{\epsilon}} + S_t^{\frac{\epsilon-1}{\epsilon}} \right)^{\frac{\epsilon}{\epsilon-1}}, \tag{2.54}$$

where $\epsilon < 1$ is a parameter that controls the matching efficiency. For $\epsilon \in (0,1)$, we have $y_t < \min\{D_t, S_t\}$ for any $D_t, S_t > 0$. Hence, in general, there will be some matching frictions that prevent the demand requested by buyers from being fully fulfilled and the capacity supplied by firms from being fully utilized.

The buyer, who could be a household that needs consumption or a firm that needs investment, has to hire some labor for purchasing. For each unit of demand requested, $\phi > 0$ units of labor are required.

The firm, who wants to supply some capacity, also has to hire some labor and install some capital. Suppose that there is a continuum of identical firms with measure one. Each is indexed by $j \in [0,1]$. The capacity that could be supplied by firm j is given by a production function, $S_{j,t} \leq F(k_{j,t}, l_{j,t})$, which is assumed to be twice differentiable, concave, and homogeneous of degree one. In addition, $F_K > 0$, $F_L > 0$, $F_{KK} < 0$, and $\forall x \geq 0$, $F(0,x) = F(x,0) = 0$, which means that capital and labor are both necessary for production.[9]

The capital stock $k_{j,t}$ is inherited from the previous period. The labor $l_{j,t}$ is hired and paid at the beginning of this period but before the matching takes place. Once the matching has taken place, the labor payment is sunk and the firm can produce with no additional cost until the output reaches the capacity $S_{j,t}$. Hence, the variable cost of the firm is not affected by the actual output level but is simply an increasing function of the capacity supplied:

$$C(S_{j,t}, k_{j,t}; w_t) = \min_{l_{j,t} \geq 0} w_t l_{j,t}, \tag{2.55}$$

subject to the production technology $S_{j,t} \leq F(k_{j,t}, l_{j,t})$.

Competitive Search Protocol

Following Moen (1997) and Bai et al. (2012), a competitive search protocol is assumed. Each firm sets up its own market characterized by both the price and the market tightness. The latter is defined as the ratio of demand to supply: $x_{j,t} \equiv D_{j,t}/S_{j,t}$, where $D_{j,t}$ is the demand that goes to firm j. Each buyer is infinitely small and takes both the price and the market tightness as given. If a buyer purchases from firm j, the unit cost of acquiring goods is given by

$$\varsigma(p_{j,t}, x_{j,t}; w_t) \equiv p_{j,t} + \frac{x_{j,t}}{M(x_{j,t}, 1)} \phi w_t, \tag{2.56}$$

where $p_{j,t}$ is the real price charged by firm j and $x_{j,t}/M(x_{j,t},1)$ is the amount of demand that has to be requested to yield a successful matching. The buyer chooses the best trade off between the price and the market tightness to minimize the unit cost of purchasing. The firm chooses the supply of capacity and a combination of the price and the market

[9] In Bai et al. (2012), capital and labor can increase the amount of goods purchased *after* a matching has taken place, and what is searched by buyers is a fixed measure of firms $T = 1$. Thus, the sales or output is given by $y = M(D,T)S = M(D,1)F(k,l)$, which is of increasing returns to scale (IRS). To remove the effects that IRS brings, I restore constant returns to scale by assuming that the matching is between the goods demanded and the capacity supplied and that capital and labor can increase the supply of capacity: $y = M(D,S) = M(D, F(k,l))$.

tightness to maximize its gross profit:

$$\max_{p_{j,t},x_{j,t},S_{j,t}\geq 0} p_{j,t}M\left(x_{j,t},1\right)S_{j,t}-C\left(S_{j,t},k_{j,t};w_t\right) \tag{2.57}$$

subject to the participation constraint $\varsigma\left(p_{j,t},x_{j,t};w_t\right)\leq\varsigma_t^*$, where $M\left(x_{j,t},1\right)<1$ gives the probability of selling or the capacity utilization rate and ς_t^* is the lowest unit cost offered by other firms or the outside option of the buyers. In equilibrium, all firms would offer the same price and the same market tightness, and the supply of capacity is proportional to the firm's capital: $\forall j$, $p_{j,t}=1$, $x_{j,t}=x_t$, and $S_{j,t}\propto k_{j,t}$. Although the capacity supplied is not fully utilized because of the matching frictions, capacity expansion is still profitable as it attracts demand. Consequently, a marginal increase in capital is valuable because it saves the cost of supplying a given level of capacity, or equivalently, because it expands capacity that attracts profitable demand:

$$-C_k\left(S_{j,t},k_{j,t};w_t\right)=M\left(x_{j,t},1\right)\frac{S_{j,t}}{k_{j,t}}-C\left(\frac{S_{j,t}}{k_{j,t}},1;w_t\right). \tag{2.58}$$

Partially Efficient Capital Accumulation

The aggregate variable cost incurred to satisfy the aggregate demand y_t takes into account all the variable costs paid by firms *and* buyers:

$$\mathcal{C}\left(y_t,k_t;w_t\right)=\min_{\left\{D_{j,t},S_{j,t},l_{j,t},k_{j,t}\geq 0\right\}_{j\in[0,1]}} w_t\int_0^1\left(\phi D_{j,t}+l_{j,t}\right)dj, \tag{2.59}$$

subject to the matching technology $y_t\leq\int_0^1 M\left(D_{j,t},S_{j,t}\right)dj$, the production technology $S_{j,t}\leq F\left(k_{j,t},l_{j,t}\right)$, and the capital stock available $\int_0^1 k_{j,t}dj\leq k_t$. Because the above minimization problem is convex and symmetric, there exists a symmetric allocation that minimizes the aggregate variable cost. By the envelope theorem and the implicit function theorem, $\mathcal{C}$ is non-negative, twice differentiable, and homogeneous of degree one; and $\mathcal{C}_y>0$, $\mathcal{C}_k<0$, and $\mathcal{C}_{kk}>0$. In addition, one can prove by contradiction that capital is necessary for production.[10]

Thanks to the competitive search protocol, the extra capacity supplied by firms to compete for demand is justified by the variable cost saved for buyers. One can verify that, in a symmetric equilibrium, the marginal value of capital viewed from a representative firm's perspective (see equation (2.58)) is the same as the marginal cost saved by capital in

[10] If for a given positive aggregate demand $y>0$, $\mathcal{C}$ is upper bounded as the aggregate capital converges to zero ($k\to 0$), the total labor hired to solve the cost minimization problem must also be upper bounded: $\int_0^1 l_j dj\leq\bar{l}$. Hence, $y\leq\int_0^1 S_j dj\leq F\left(k,\bar{l}\right)\to 0$, as $k\to 0$; a contradiction to that $y>0$.

aggregate:

$$-C_k\left(S_t, k_t; w_t\right) = M\left(x_t, 1\right) \frac{S_t}{k_t} - C\left(\frac{S_t}{k_t}, 1; w_t\right) = -\mathcal{C}_k\left(y_t, k_t; w_t\right). \tag{2.60}$$

Thus, the capital accumulation decisions made by firms in a decentralized way also minimizes the present value of aggregate costs, that is, the capital accumulation is partially efficient. The rest of the household's problem is standard as described in section 2.2. Thus, the competitive goods market search model is a special case of the standard one-good economy.

Perfectly Competitive Markets and Efficiency

Suppose that the labor market is perfectly competitive. Because of the competitive search protocol in the goods market, the economy is efficient and can be summarized as a social planner's problem:

$$\max_{\{c_t, y_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(z_{c,t} u\left(\frac{c_t}{z_{c,t}}\right) + \bar{\omega} v\left(1 - \mathcal{L}\left(y_t, k_t\right)\right)\right), \tag{2.61}$$

subject to $c_t + k_{t+1} = y_t + k_t(1-\delta)$, where the labor required by the social planner $\mathcal{L}$ is given by

$$\mathcal{L}\left(y_t, k_t; w_t\right) \equiv \min_{\{D_{j,t}, S_{j,t}, l_{j,t}, k_{j,t} \geq 0\}_{j \in [0,1]}} \int_0^1 \left(\phi D_{j,t} + l_{j,t}\right) dj, \tag{2.62}$$

subject to the matching technology $y_t \leq \int_0^1 M\left(D_{j,t}, S_{j,t}\right) dj$, the production technology $S_{j,t} \leq F\left(k_{j,t}, l_{j,t}\right)$, and the capital stock available $\int_0^1 k_{j,t} dj \leq k_t$. Hence, the competitive goods market search model works just like the RBC model except that the required labor function $\mathcal{L}$ is different.

2.4 Properties of the Standard One-Good Economy

In this section, I study the properties of the standard one-good economy to evaluate the efficient utilization models described in section 2.3. I show that when capital accumulation is partially efficient, aggregate capacity must be fully utilized in steady state. Hence, none of these models could feature chronic excess capacity. Furthermore, I show that when aggregate capacity is fully utilized in steady state, capital resources must be tight locally around the steady state. Hence, none of these models describe an economy with capital resource slackness. As in the standard RBC model, capital resource tightness implies an upward sloping real marginal cost curve and a countercyclical marginal product of labor. Therefore, it is not surprising to see that the efficient utilization models, or other models that could be encompassed by the standard one-good economy, tend to face the same issues as the standard RBC model. Particularly, I show that either the relative volatility of investment to consumption is too small or the real wage rate is too countercyclical under consumption demand shocks. In addition, the labor productivity is too countercyclical and the Solow residual is not pro-cyclical enough.

2.4.1 No Chronic Excess Capacity

In section 2.3, capacity refers to the production limit within which goods can be produced with a constant marginal cost. The notion of capacity can be extended to the case where there is no production limit or the marginal cost of production is increasing in output. Following Morrison (1985), theoretical capacity is defined as the output level at which the short run average total cost (SRAC) curve is tangent to the long run average total cost (LRAC) curve. Along the SRAC curve, one minimizes the average total cost of production by adjusting only the variable inputs, such as labor. Along the LRAC curve, one minimizes the same average total cost by adjusting both the variable inputs and the short run fixed capital. Hence, at the point of tangency, one has no incentive to adjust the capital stock.

Definition 2.3. (*Capacity*) Capacity is the unique output level $\bar{y}$ at which the short run average total cost (SRAC) curve is tangent to the long run average total cost (LRAC) curve if the capital level is positive ($k > 0$). If the capital level is zero ($k = 0$), capacity is zero.

Definition 2.4. (*Excess capacity*) Capacity is in excess if and only if aggregate capacity is underutilized.

For a positive aggregate capital level $k_t > 0$, the aggregate capacity is the output level at which the aggregate SRAC curve is tangent to the aggregate LRAC curve. The aggregate SRAC is given by

$$\mathcal{SRAC}\left(y_t, k_t; w_t\right) \equiv \frac{\mathcal{C}\left(y_t, k_t; w_t\right) + \left(r + \delta\right) k_t}{y_t}, \tag{2.63}$$

where $r \equiv \beta^{-1} - 1 > 0$ is the real interest rate in steady state; and the aggregate LRAC is the minimum of the aggregate SRAC when aggregate capital stock can be adjusted:

$$\mathcal{LRAC}\left(y_t; w_t\right) \equiv \min_{k \in \{x | (y_t, x) \in \mathbf{dom}(\mathcal{C})\}} \mathcal{SRAC}\left(y_t, k; w_t\right). \tag{2.64}$$

We now show that under regularity conditions 1-4, the aggregate LRAC exists and there exists a unique point at which the aggregate SRAC curve is tangent to the aggregate LRAC. Thus, aggregate capacity is properly defined.

Claim 2.5. For a given positive aggregate output level $y_t > 0$, there exists a unique positive aggregate capital level that minimizes the aggregate SRAC:

$$k^* (y_t, w_t) \equiv \arg \min_{k \in \{x | (y_t, x) \in \mathbf{dom}(\mathcal{C})\}} \mathcal{SRAC} (y_t, k; w_t) . \tag{2.65}$$

Proof. If there exists a feasible production limit, the $\mathcal{SRAC}(y_t, k; w_t)$ is continuous and convex in k on the interval $[\chi y_t, \infty)$; and if $k \to \infty$, we have that $\mathcal{SRAC}(y_t, k; w_t) \geq \frac{(r+\delta)k}{y_t} \to \infty$. Hence, there exists $k^* > 0$ that minimizes the aggregate SRAC. If there is no production limit or the production limit is infeasible, the $\mathcal{SRAC}(y_t, k; w_t)$ is continuous and convex in k on the interval $(\chi y_t, \infty)$; and if $k \to \infty$ or $k \to \chi y_t$, we have that $\mathcal{SRAC}(y_t, k; w_t) \geq \min \left\{ \frac{\mathcal{C}(y_t,k;w_t)}{y_t}, \frac{(r+\delta)k}{y_t} \right\} \to \infty$. Hence, there exists $k^* > 0$ that minimizes the aggregate SRAC.

If there exists $k' \neq k^*$ that also minimizes the aggregate SRAC, then any $k \in [a, b]$ can minimize the aggregate SRAC, where $a = \min\{k', k^*\} < b = \max\{k', k^*\}$. For any $k_1, k_2 \in [a, b]$ and $k_1 > k_2$, we have $\mathcal{C}(y_t, k_1; w_t) = \mathcal{C}(y_t, k_2; w_t) + (r + \delta)(k_2 - k_1) < \mathcal{C}(y_t, k_2; w_t)$. Hence, $\mathcal{C}$ is strictly decreasing in k on $[a, b]$. By the assumed regularity condition 4, $\mathcal{C}$ is twice differentiable in terms of capital on the interval (a, b) such that $\mathcal{C}_{kk}(y_t, k; w_t) > 0$. As a result, $\mathcal{SRAC}(y, k; w_t)$ is strictly convex in capital on (a, b), a contradiction to the previous conclusion that any $k \in [a, b]$ minimizes the SRAC. Hence, k^* must be unique. □

Claim 2.6. The aggregate LRAC curve exists and is flat.

Proof. By definition, the long run average total cost curve can be expressed as:

$$\begin{aligned}
\mathcal{LRAC}(y_t; w_t) &= \min_{k \in \{x | (y_t, x) \in \mathbf{dom}(\mathcal{C})\}} \left\{ \mathcal{C}\left(1, \frac{k}{y_t}; w_t\right) + (r + \delta) \frac{k}{y_t} \right\} \\
&= \min_{\tilde{k} \in \{x | (1, x) \in \mathbf{dom}(\mathcal{C})\}} \left\{ \mathcal{C}\left(1, \tilde{k}; w_t\right) + (r + \delta) \tilde{k} \right\} \\
&= \mathcal{C}(1, k^*(1, w_t); w_t) + (r + \delta) k^*(1, w_t),
\end{aligned}$$

where $k^*(1, w_t)$ exists according to Claim 2.5. It is clear that the aggregate LRAC is independent of the aggregate output level $y_t > 0$. Hence, the aggregate LRAC curve is flat. □

Claim 2.7. Aggregate capacity is properly defined by Definition 2.3. Particularly, for a given aggregate capital level $k_t \geq 0$ and a real wage rate $w_t > 0$, the aggregate capacity is given by

$$\bar{y}_t = \frac{k_t}{k^*(1, w_t)}, \tag{2.66}$$

where k^* is the function that gives the capital level that minimizes the corresponding aggregate SRAC (see equation (2.65)).

Proof. Consider the case where the aggregate capital stock is positive ($k_t > 0$). Since the aggregate LRAC curve is flat, the tangent point $(\bar{y}_t, k_t)$, if there is any, must be a minimum point on the aggregate SRAC:

$$\min_{y \in \{x>0 | (x,k_t) \in \mathbf{dom}(\mathcal{C})\}} \left\{ \mathcal{C}\left(1, \frac{k_t}{y}; w_t\right) + (r+\delta)\frac{k_t}{y} \right\}.$$

This is equivalent to finding the capital to output ratio that minimizes the aggregate SRAC:

$$\min_{\tilde{k} \in \{x | (1,x) \in \mathbf{dom}(\mathcal{C})\}} \left\{ \mathcal{C}\left(1, \tilde{k}; w_t\right) + (r+\delta)\tilde{k} \right\}.$$

According to Claim 2.5, $k^*(1, w_t)$ is the unique capital to output ratio that minimizes the aggregate SRAC. Therefore, the tangent point exists uniquely and is given by $\left({}^{k_t}/_{k^*(1,w_t)}, k_t\right)$. It is proper to define aggregate capacity as the output level at the unique tangent point: $\bar{y}_t = {}^{k_t}/_{k^*(1,w_t)}$. If $k_t = 0$, aggregate capacity is also properly defined as zero. □

Proposition 2.8. *If capital accumulation is partially efficient, under regularity conditions 1-4, aggregate capacity must be fully utilized in steady state.*

Proof. Since capital accumulation is partially efficient, the capital accumulation decisions made by firms in equilibrium minimize the present value of aggregate costs:

$$\min_{\{k_{t+1} \geq 0\}_{t \geq 0}} \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} Q_\tau \right) \left(\mathcal{C}(y_t, k_t; w_t) + (k_{t+1} - k_t(1-\delta)) \right) \right).$$

In a steady state where $Q_t = \beta$, $w_t = w > 0$, and $y_t = y > 0$, the above minimization problem implies that the aggregate capital level k must minimize the aggregate total cost: $\mathcal{C}(y, k; w) + (r+\delta)k$. Hence, (y, k) is the point at which the aggregate SRAC curve is tangent to the aggregate LRAC curve: $\mathcal{SRAC}(y, k; w) = \mathcal{LRAC}(y; w)$. By Claim 2.7, the point of tangency that gives the aggregate capacity $\bar{y}$ is unique for each aggregate capital level. Thus, $y = \bar{y}$, which means that aggregate capacity is fully utilized in steady state. □

Intuitively, if capital accumulation is partially efficient, the only reason for firms to expand capital is to lower the aggregate total cost. However, if the aggregate capacity level is larger than the aggregate output level in steady state, it is capital reduction rather than expansion that lowers the cost. Hence, firms will reduce their capital until aggregate capacity becomes fully utilized in steady state. By Definition 2.4, the economy does not exhibit chronic excess capacity.[11]

Proposition 2.8 shows that partial efficiency of capital accumulation rules out the existence of chronic excess capacity in the standard one-good economy.

Despite the existence of capital, plant, or capacity underutilization in those efficient utilization models, aggregate capacity is fully utilized. The capital, plant, or capacity is not utilized for a cost-saving reason. In the variable capital utilization model, some capital is left unused because of an increasing marginal cost of utilization. In the variable plant utilization model, some plants are left idle because the productivities of those plants are too low to compensate for the labor cost. In the precautionary capacity model, some firms have their capacity underutilized because the extra capacity is a necessary precaution that saves the expected cost incurred to satisfy a unit of aggregate demand. In the competitive goods market search model, firms have their capacity underutilized because the extra capacity reduces market tightness and thus lowers the search cost for buyers. Hence, none of these underutilization phenomena would mean that too much capital is accumulated or too little capacity is utilized in aggregate.

2.4.2 Capital Resource Tightness

Capital resources are said to be *tight* if a marginal decrease in capital makes the current output level infeasible or leads to an increase in the real marginal cost (MC), and capital resources are said to be *slack* if a marginal decrease in capital has no effect on the real MC.

Definition 2.9. (*Capital resource tightness*) Let $\partial_y \mathcal{C}(y, k; w)$ denote the sub-differential of the aggregate variable cost function with respect to aggregate output.[12] Consider a point $(y, k) \in \mathbf{dom}(\mathcal{C})$ and $y > 0$. If there exists $\varepsilon > 0$ such that for all $\Delta \in (0, \varepsilon)$, $(y, k - \Delta)$ is either not feasible or $\inf \partial_y \mathcal{C}(y, k - \Delta; w) > \inf \partial_y \mathcal{C}(y, k; w)$, we say that capital resources

[11] The idea of Proposition 2.8 is not new. For instance, Hall (1986) concludes that there is chronic excess capacity in the U.S. based on his finding that the estimated marginal cost saved by capital, $-\mathcal{C}_k(y, k; w)$, is smaller than the rental cost of capital, $r + \delta$, in various U.S. industries.

[12] The sub-differential of a convex function $f : D \to \mathbf{R}$ at an interior point x_0 is the set of all $c \in \mathbf{R}$ such that $f(x) - f(x_0) \geq c(x - x_0)$ for all $x \in D$. One can show that the sub-differential of a convex function always exists and is a non-empty closed interval.

are *tight* locally around (y, k). If there exists $\varepsilon > 0$ such that for all $\Delta \in (0, \varepsilon)$, $(y, k - \Delta)$ is feasible and $\inf \partial_y \mathcal{C}(y, k - \Delta; w) = \inf \partial_y \mathcal{C}(y, k; w)$, we say that capital resources are *slack* locally around (y, k).

By Definition 2.9, the shape of the aggregate variable cost function $\mathcal{C}$ determines when capital resources must be tight and when they could be slack. We know that $\mathcal{C}(1, k; w)$ is convex and decreasing in terms of capital to output ratio. Hence, if there exists $k_1 < k_2$ such that $\mathcal{C}(1, k_2; w) = \mathcal{C}(1, k_1; w)$, then for all $k \geq k_1$, we have $\mathcal{C}(1, k; w) = \mathcal{C}(1, k_1; w)$. Let $\mathbb{S}(w)$ be the set of all such k_1 from which the aggregate variable cost function no longer decreases with the capital to output ratio:

$$\mathbb{S}(w) \equiv \left\{ \tilde{k} \mid \left(1, \tilde{k}\right) \in \mathbf{dom}(\mathcal{C}), \forall k \geq \tilde{k}, \mathcal{C}(1, k; w) = \mathcal{C}\left(1, \tilde{k}; w\right) \right\}. \tag{2.67}$$

$\mathbb{S}(w)$ could be interpreted as a set of *satiation* in which the capital to output ratio is so large that additional capital no longer saves the aggregate variable cost. If $\mathbb{S}(w) \neq \emptyset$, let $\kappa(w) \equiv \inf \mathbb{S}(w)$ be the minimum capital to output ratio that satiates the appetite for capital as a production factor.

The shape of the aggregate variable cost function $\mathcal{C}$ is described by one of the following cases:

1. If $\mathbb{S}(w) = \emptyset$, $\mathcal{C}$ is strictly decreasing in capital on its entire domain.
2. If $\mathbb{S}(w) \neq \emptyset$ and $\kappa(w) > \chi$, $\mathcal{C}$ is strictly decreasing in capital on $(\chi y, \kappa(w) y)$ and flattens out when $k \in [\kappa(w) y, \infty)$.
3. If $\mathbb{S}(w) \neq \emptyset$ and $\kappa(w) = \chi$, $\mathcal{C}$ is flat in capital on its entire domain and the production technology is Leontief.

Figure 2.2 illustrates the shape of an aggregate variable cost function described by the second case.[13] The thick solid line highlights the part where capital resources are slack. Intuitively, capital resources are slack if and only if the capital stock is so large that the aggregate variable cost function flattens out.

Lemma 2.10. *Under regularity conditions 1-4, for a given positive output level $y > 0$, capital resources are slack if and only if the set of satiation is non empty $\mathbb{S}(w) \neq \emptyset$ and the capital to output ratio is large enough such that it belongs to the interior of the satiation set: $k/y > \kappa(w) \equiv \inf \mathbb{S}(w)$.*

[13] The variable cost function is drawn based on a production technology given by $y = \min\left\{\left(0.38k^{-1} + 0.62l^{-1}\right)^{-1}, l\right\}$.

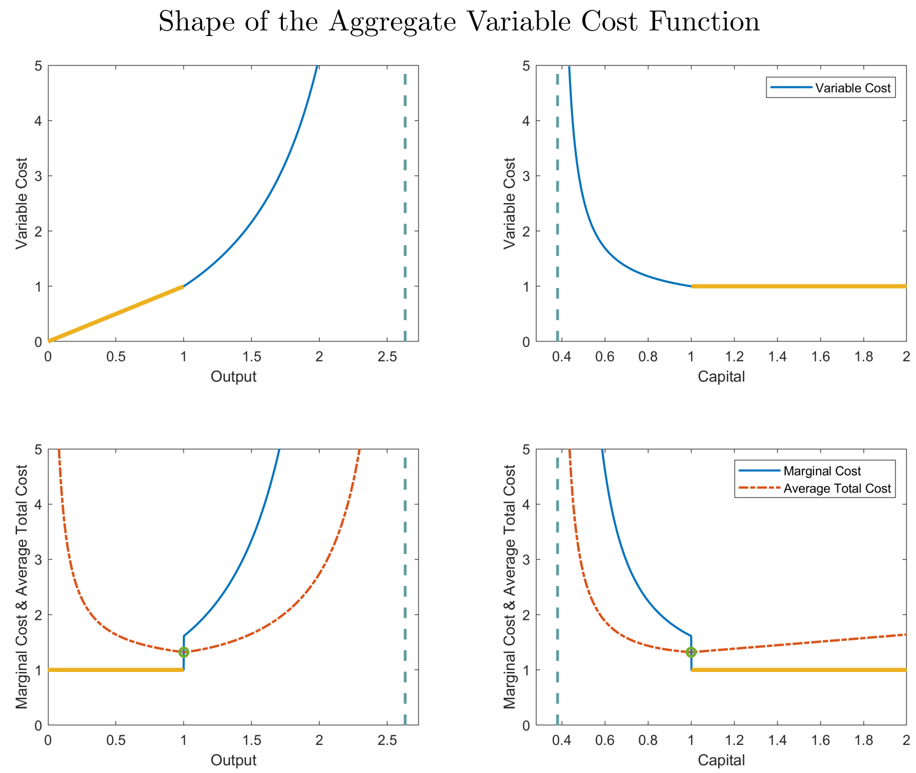

Figure 2.2: The thick solid line highlights the part where capital resources are slack. The vertical dashed line indicates a non-feasible production limit. The circle marker is the minimum of the average total cost and is also the point where capacity is fully utilized.

Proof. First note that capital resources are tight locally around the production limit $k = \chi y$ because a marginal decrease in capital makes the current output level infeasible. Second, if $\mathbb{S}(w) = \emptyset$, $\mathcal{C}$ is strictly decreasing in capital on its entire domain. Since the marginal returns to capital diminishes smoothly (see regularity condition 4), $\mathcal{C}$ is twice differentiable for all $k > \chi y$ and $\mathcal{C}_{yk}(y, k; w) = -\mathcal{C}_{kk}(y, k; w) \frac{k}{y} < 0$. Hence, capital resources are tight for all $k \in (\chi y, \infty)$. Third, if $\mathbb{S}(w) \neq \emptyset$ and $\kappa(w) > \chi$, $\mathcal{C}$ is strictly decreasing in capital on $(\chi y, \kappa(w) y)$. By the same argument, capital resources are tight for $k \in (\chi y, \kappa(w) y)$. In addition, if $k = \kappa(w) y$, there exists $\varepsilon = (\kappa(w) - \chi) y > 0$ such that for all $\Delta \in (0, \varepsilon)$, $\mathcal{C}(y, k - \Delta; w)$ is twice differentiable and $\mathcal{C}_y(y, k - \Delta; w)$ is strictly increasing in Δ. Thus, $\mathcal{C}_y(y, k - \Delta; w) > \lim_{\Delta \downarrow 0} \mathcal{C}_y(y, k - \Delta; w) \geq \inf \partial_y C(y, k; w)$, meaning that capital resources are tight for all $k \in (\chi y, \kappa(w) y]$. Fourth, if $\mathbb{S}(w) \neq \emptyset$ and $\kappa(w) = \chi$, capital resources are tight locally around the production limit ($k = \chi y = \kappa(w) y$). Finally, if and only if $\mathbb{S}(w) \neq \emptyset$ and $k > \kappa(w) y$, we have that capital resources are slack because $\mathcal{C}$ is flat and twice differentiable on $(\kappa(w) y, \infty)$ and $\mathcal{C}_{yk}(y, k; w) = -\mathcal{C}_{kk}(y, k; w) \frac{k}{y} = 0$. $\square$

We are now ready to show that if aggregate capacity is fully or overly utilized, capital resources must be tight. Intuitively, if capital resources are slack, the level of capital must be so high that a small reduction in capital would have no negative effect on production;

but holding capital is costly. Hence, a small reduction in capital can *lower* the aggregate SRAC. However, if aggregate capacity is fully or overly utilized, any reduction in capital should cause an *increase* in the aggregate SRAC. This contradiction implies that capital resources must be tight.

Proposition 2.11. *If aggregate capacity is fully or overly utilized, under regularity conditions 1-4, capital resources must be tight.*[14]

Proof. Let $y > 0$ be the aggregate output, $w > 0$ be the real wage rate, and $\bar{y}$ be the aggregate capacity. By Definition 2.3, the aggregate SRAC curve is tangent to the aggregate LRAC curve at $(\bar{y}, k)$. By Claims 2.5-2.6, the point of tangency $(\bar{y}, k)$ is the unique minimum of the aggregate SRAC curve. Since the aggregate SRAC curve is convex, for any $y' > y \geq \bar{y}$, we have $\mathcal{SRAC}\left(y', k; w\right) > \mathcal{SRAC}\left(y, k; w\right)$. Note that $\mathcal{SRAC}\left(\cdot, \cdot; w\right)$ is homogeneous of degree zero in output and capital. Thus, the above inequality also implies that for any $k' < k$, $\mathcal{SRAC}\left(y, k'; w\right) > \mathcal{SRAC}\left(y, k; w\right)$. If the capital resources are slack, by Lemma 2.10, we know that the set of satiation $\mathbb{S}\left(w\right)$ is non-empty and the capital stock k is larger than $\kappa\left(w\right) y$. Thus, there exists a small decrease in capital, $\Delta \in \left(0, k - \kappa\left(w\right) y\right)$, that can reduce the aggregate SRAC further:

$$\frac{\mathcal{C}\left(y, k-\Delta; w\right) + \left(r+\delta\right)\left(k-\Delta\right)}{y} < \frac{\mathcal{C}\left(y, k; w\right) + \left(r+\delta\right) k}{y}.$$

contradicting the statement that any reduction in capital must increase the aggregate SRAC. □

Corollary 2.12. *If capital accumulation is partially efficient, under regularity conditions 1-4, capital resources must be tight locally around the steady state.*

Proof. The corollary can be inferred from Proposition 2.8 and Proposition 2.11. □

Hence, the partial efficiency of capital accumulation not only rules out chronic excess capacity but also rules out capital resource slackness. Suppose that the aggregate variable cost function $\mathcal{C}$ is twice differentiable locally around the steady state.[15] Capital resource tightness can be measured by the negative capital elasticity of the real MC. Since the

[14]Smoothly diminishing marginal return to capital (regularity condition 4) is not necessary for Propositions 2.8 and 2.11 to hold. One can assume instead that the capital level that minimizes the aggregate SRAC is *unique*. If one wants to ensure the uniqueness of the steady state, this alternative assumption is necessary. The smoothly diminishing marginal return to capital is simply a stronger property to imply the uniqueness (see Claim 2.5).

[15]$\mathcal{C}$ is twice differentiable almost everywhere. The only interior points where $\mathcal{C}$ may not be twice differentiable are those where the capital to output ratio is just large enough to satiate the appetite for capital as a production factor: $k = \kappa\left(w\right) y$. Locally around this point, capital resources are infinitely tight.

production technology is of constant returns to scale, the measure of capital resource tightness is also equal to the steepness of the real MC curve or the convexity of the aggregate variable cost curve:

$$\zeta = -\frac{\partial \ln \mathcal{C}_y \left(y, k; w\right)}{\partial \ln k} = \frac{\partial \ln \mathcal{C}_y \left(y, k; w\right)}{\partial \ln y} = \frac{\mathcal{C}_{yy} \left(y, k; w\right) y}{\mathcal{C}_y} > 0, \tag{2.68}$$

which must be positive according to Corollary 2.12.

The rest of this section shows that capital resource tightness has important consequences on the dynamic properties of the model.

2.4.3 Countercyclical Real Wage Rate

Since capital resources are tight, an increase in output would cause a decrease in the marginal product of labor, creating a pressure for the real MC to increase. With a constant markup, this pressure will be transmitted to a decrease in the real wage rate according to the firm's pricing condition (2.17). With perfectly flexible prices and consumption demand shocks only, equation (2.17) is reduced to

$$\hat{w}_t = -\zeta \left(\hat{y}_t - \hat{k}_t\right), \tag{2.69}$$

which links the cyclicality of the real wage rate directly to the magnitude of capital resource tightness.

One could use the covariance between the real wage rate and the aggregate output normalized by the variance of the aggregate output to measure the cyclicality of the real wage rate. The business cycle fluctuations of the capital stock $\hat{k}_t$ can largely be ignored as they are typically very small compared to others. Thus, we have in approximation

$$\beta_y^w \equiv \frac{\mathbf{Cov}\left(\hat{w}_t, \hat{y}_t\right)}{\mathbf{Var}\left(\hat{y}_y\right)} \approx -\zeta < 0, \tag{2.70}$$

which says that the tighter the capital resources are, the more countercyclical the real wage rate.

2.4.4 Upward Sloping AS Curve and its Dampening Effect on Output

With perfectly flexible prices, the AS surface (2.32) is orthogonal to the output-price in utils plane. The projection of the AS surface on the output-price in utils plane gives us the real AS curve.

There are two forces that cause the real AS curve to be upward sloping. First, the real MC curve is upward sloping and marginal product of labor is decreasing with output because of capital resource tightness. Second, the marginal dis-utility of supplying labor increases with hours. The second effect can be understood as a result of labor resource tightness measured by the inverse of the Frisch elasticity of labor supply $\eta \geq 0$. The following equation illustrates how the capital resource tightness measured by ζ and the labor resource tightness measured by η affect the cyclicality of the price in utils $\hat{\lambda}_t$:

$$\hat{\lambda}_t = \zeta \left(\hat{y}_t - \hat{k}_t\right) + \eta \hat{l}_t. \tag{2.71}$$

Equation (2.71) is obtained by combining the household's labor supply condition (2.15) and the firm's pricing condition (2.17).

Substituting the log-linearied required labor function (2.20) for $\hat{l}_t$ in equation (2.71) gives us the real AS curve:

$$\hat{\lambda}_t = (\zeta + \eta\nu) \left(\hat{y}_t - \hat{k}_t\right) + \eta \hat{k}_t. \tag{2.72}$$

Equation (2.72) shows that the price in utils must be pro-cyclical as real demand shocks move the equilibrium output along the upward sloping real AS curve. A pro-cyclical price in utils due to the steep real AS curve, however, will reduce the response of output to demand shocks by making goods more expensive in boom and less expensive in recession.

2.4.5 Lack of Investment Volatility

If labor resources are slack $\eta = 0$ and if the magnitude of capital resource tightness converges to zero ($\zeta \to 0$), the real AS curve (2.72) becomes almost flat. One can see from the household's optimal consumption condition

$$\zeta \left(\hat{y}_t - \hat{k}_t\right) = \hat{\lambda}_t = \gamma \left(z_{c,t} - \hat{c}_t\right), \tag{2.73}$$

that the price in utils is almost acylical and the response of consumption to demand is almost not dampened.

However, even if the real AS curve is almost flat, the response of investment to consumption demand shocks is still much limited. To see this, combine the real AS curve (2.72) and the firm's investment condition (2.19) and set $\eta = 0$ to ensure the slackness of labor resources.

We have:

$$\zeta\left(\hat{y}_t - \hat{k}_t\right) = \zeta\left(1 + \beta\frac{wl}{k}\right)\mathbb{E}_t\left(\hat{y}_{t+1} - \hat{k}_{t+1}\right). \tag{2.74}$$

It is clear that the measure of capital resource tightness ζ appears on both sides and is canceled out. Hence, even if $\zeta \to 0$, the response of investment to consumption demand shocks is still much limited.

Intuitively, when capital resources become less tight, two forces work in opposite directions. On one hand, the price in utils $\hat{\lambda}_t$ is less pro-cyclical; thus, the dampening effect on investment is weaker (see the left hand side of equation (2.74)). On the other hand, the return of capital also becomes less pro-cyclical; thus, firms would have less incentive to adjust capital (see the right hand side of equation (2.74)). In fact, precisely because capital resources are tight, there is a pro-cyclical upward pressure on the real MC that firms can adjust their capital stock to mitigate. Thus, removing the curb on capital adjustment would also remove the impetus for capital adjustment.

In the quantitative exercises below, I show that as capital resources become less tight, the volatility of investment does not increase, but the volatility of consumption does. The relative volatility of investment to consumption declines. To achieve a large relative volatility of investment to consumption, capital resources must be so tight that the real wage rate becomes too countercyclical.

Calibration

I start with the case where labor resources are slack ($\eta = 0$) and the households are the most patient ($\beta \to 1$) to maximize the volatility of investment. Intuitively, the real AS curve becomes flatter when the labor resources are less tight; thus, it is easier for firms to adjust capital. In addition, if the future becomes more important, firms also have a stronger incentive to adjust capital.

Prices are assumed to be perfectly flexible: $\phi_p = 0$. The depreciation rate δ, the curvature of the consumption utility γ, the goods demand elasticity ε_p, and the output elasticity of labor demand ν in steady state are calibrated to match the same targets as described in section 2.2.8. Finally, the measure of capital resource tightness ζ and the persistence parameter of consumption demand shocks ρ_c would be set to different values to explore how the relative volatility of investment to consumption is affected.

Table 2.3 summarizes the calibrated parameter values and their mostly associated targets.

Table 2.3: Parameters and Calibration Targets – Standard One-Good Economy

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	Elasticity of inter-temporal substitution 1
ν	1.2742	Labor share of income 0.62
ε_p	4.7623	Investment to output ratio 0.17

Quantitative Results

Before we look at the relative volatility of investment to consumption, we should first ensure that the model generates a positive correlation between consumption and investment. If an increase in consumption demand is too transitory, households would rather reduce investment to save resources for current consumption. Furthermore, if capital resources are too tight, the real AS curve would be so steep that any desired investment would have to be achieved by sacrificing current consumption. Figure 2.3 shows the results of a numerical experiment: if consumption demand shocks are persistent enough ($\rho_c \geq 0.935$) and capital resources are not too tight ($\zeta \leq 1.34$), a positive correlation can be achieved.

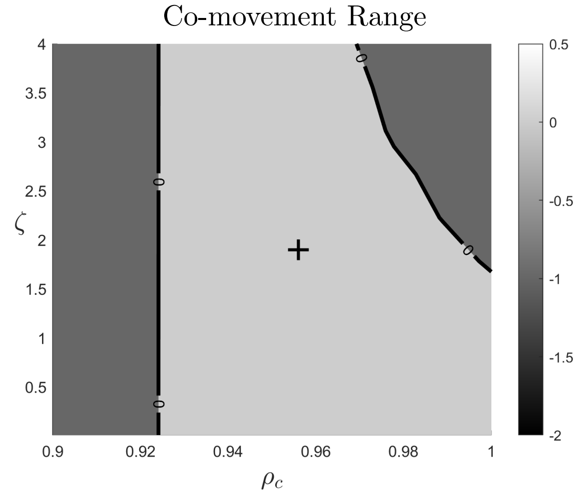

Figure 2.3: The light shaded area indicates the range for parameters values where consumption, investment, and hours co-move under consumption demand shocks in a standard one-good economy.

Different values of $\zeta \in (0, 1.34]$ and $\rho_c \in [0.935, 1)$ are then explored to exam their effects on the relative volatility of investment to consumption. Each contour line drawn in the left panel of Figure 2.4 shows the values of ζ and ρ_c that lead to the same level of relative volatility of investment to consumption. The results show that the relative volatility of investment to consumption falls as capital resources become less tight.

We also see that the relative volatility of investment to consumption falls as the consumption demand shocks becomes less persistent. Consider a positive consumption demand shock for example. If the increase in consumption demand is concentrated in a short period of time, it is not wise to make a large investment. By contrast, if the increase is quite persistent, households prefer a smooth increase in consumption for a long period of time; thus, a large investment is justified.

Lack of Investment Volatility or Countercyclical Real Wage Rate

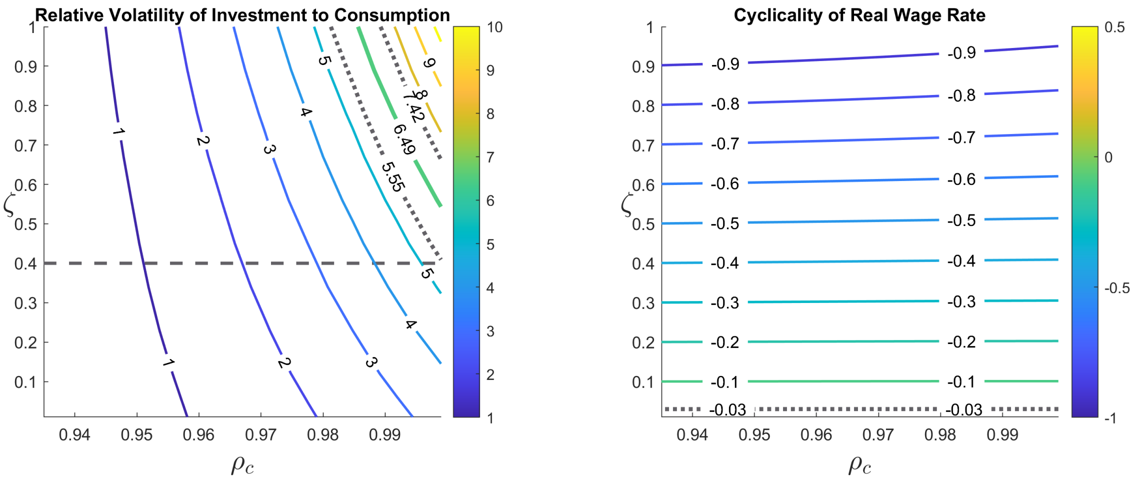

Figure 2.4: Each solid line is a contour line. The dotted lines are the contour lines corresponding to the boundaries of the 95% confidence intervals. The 95% confidence interval of the relative volatility is $[5.55, 7.42]$. The 95% confidence interval of the cyclicality of the real wage is $[-0.03, 0.18]$. The data is from the BEA and the 95% confidence intervals are calculated using a parametric bootstrapping method.

The dashed line in the left panel of Figure 2.4 shows the minimum value of ζ needed to obtain the observed relative volatility of investment to consumption. The value is about 0.4, which implies a real wage rate that is too countercyclical as shown in the right panel of Figure 2.4. Therefore, either the relative volatility of investment to consumption is too small or the real wage rate is too countercyclical under consumption demand shocks.

Robustness of the Quantitative Results

The above results are robust to different values of η. If labor resources become tighter $\eta > 0$, the real AS curve becomes steeper. However, unlike capital resource tightness, labor resource tightness does not affect the return of capital and thus the incentive of firms to adjust capital stock. Therefore, as the real AS curve becomes steeper, it simply becomes more costly for firms to move resources across time. In order to have less resource movement across time, the response of investment is reduced more than the response of consumption. Hence, the relative volatility of investment to consumption declines. The upper left panel of Figure 2.5 confirms this analysis.

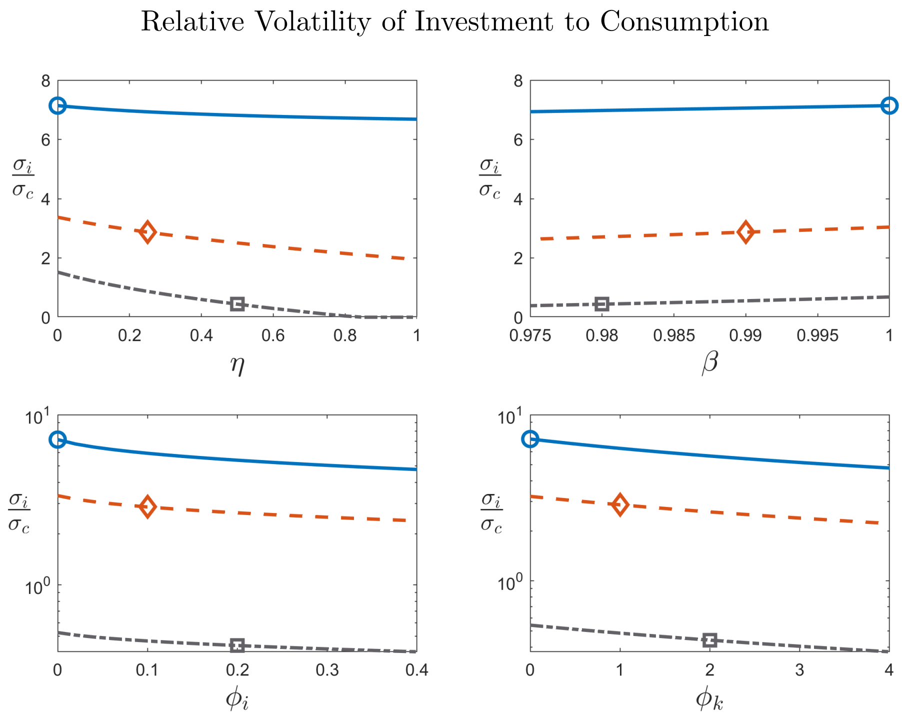

Figure 2.5: The circle marker indicates the case where $\zeta = 0.6$, $\rho_c \to 1$, $\beta \to 1$, and $\phi_i = \phi_k = \eta = 0$. The diamond marker indicates the case where $\zeta = 0.4$, $\rho_c = 0.995$, $\beta = 0.99$, $\phi_i = 0.1$, $\phi_k = 1$, and $\eta = 0.25$. The square marker indicates the case where $\zeta = 0.2$, $\rho_c = 0.99$, $\beta = 0.98$, $\phi_i = 0.2$, $\phi_k = 2$, and $\eta = 0.5$. Each line shows how the relative volatility of investment to consumption changes with the value of the parameter specified on the x-axis while the values of other parameters are the same as in the case indicated by the marker on the line.

The results are also robust to different values of β. If households are less patient $\beta < 1$, the future becomes less important. The firms would have less incentive to change capital but the incentive to consume will not be negatively affected. Hence, the relative volatility of investment to consumption declines. The upper right panel of Figure 2.5 confirms this analysis.

We could also extend the model to include capital and investment adjustment costs. Both tend to reduce the relative volatility of investment to consumption. Suppose that the capital stock of firm j is accumulated according to

$$k_{j,t+1} = k_{j,t}\left(1-\delta\right) + i_{j,t}\left(1 - S\left(i_{j,t}, i_{j,t-1}, k_{j,t}\right)\right), \tag{2.75}$$

where S is the adjustment cost function. Following Hayashi (1982) and Christiano et al. (2005), I assume that

$$S\left(i_{j,t}, i_{j,t-1}, k_{j,t}\right) = \frac{\phi_i}{2}\left(\frac{i_{j,t}}{i_{j,t-1}} - 1\right)^2 + \frac{\phi_k}{2}\left(\frac{i_{j,t}}{k_{j,t}} - \delta\right)^2 \frac{k_{j,t}}{i_{j,t}}, \tag{2.76}$$

where $\phi_k \geq 0$ and $\phi_i \geq 0$ are parameters that capture the curvature of the capital and investment adjustment costs respectively. The existence of adjustment costs reduces the incentive to change capital but does not affect the incentive to consume. Hence, the relative volatility of investment to consumption falls. The lower panels of Figure 2.5 confirm this analysis.

So far we have chosen $\gamma = 1$ to yield a log utility. In the literature, the elasticity of inter-temporal substitution is typically estimated to be smaller than one; thus, the value of γ is typically larger than one (see Havránek, 2015 for a meta-analysis). Our results are also robust to $\gamma > 1$. One can show that if γ, ζ, η, ϕ_i, and ϕ_k are all scaled up by the same proportion, the local dynamics of consumption, investment, and hours remains unchanged. Hence, if γ becomes larger, a larger value of ζ would be needed to match the relative volatility of investment to consumption observed in the data, making the real wage rate more countercyclical.

2.4.6 Countercyclical Labor Productivity and Acyclical Solow Residual

Without labor productivity shocks, the standard one-good economy tends to generate a labor productivity that is too countercyclical and a Solow residual that is not pro-cyclical enough. The labor productivity is measured as the output to labor ratio and the Solow residual is measured based on the average labor share of income.

Recall that the log-linearized labor cost function is given by

$$\hat{l}_t = \nu\left(\hat{y}_t - \hat{k}_t\right) + \hat{k}_t, \tag{2.77}$$

where $\nu = \frac{\varepsilon_p - 1}{\varepsilon_p}\frac{y}{wl} = \frac{wl + (r+\delta)k}{wl} > 1$ is the output elasticity of labor demand in steady state. Ignore the business-cycle fluctuations of the capital stock $\hat{k}_t$ that are typically very small compared to others. In approximation, we have that the cyclicality of the labor productivity is given by

$$\beta_y^{LP} \equiv \frac{\mathbf{Cov}\left(\hat{y}_t - \hat{l}_t, \hat{y}_t\right)}{\mathbf{Var}\left(\hat{y}_y\right)} \approx 1 - \nu = -\frac{(r+\delta)\,k}{wl} < 0, \tag{2.78}$$

and that the cyclicality of the Solow residual is given by

$$\begin{aligned} \beta_y^{SR} &\equiv \frac{\mathbf{Cov}\left(\hat{y}_t - \frac{wl}{y}\hat{l}_t - \left(1 - \frac{wl}{y}\right)\hat{k}_t, \hat{y}_t\right)}{\mathbf{Var}\left(\hat{y}_y\right)} \\ &\approx \frac{1}{\varepsilon_p} = 1 - \frac{wl + (r+\delta)\,k}{y} < 1 - \frac{wl + \delta k}{y}. \end{aligned} \tag{2.79}$$

Table 2.4: Cyclicality of Labor Productivity and Solow Residual

	β_y^{LP}			β_y^{SR}		
	$\rho_c = 0.93$	$\rho_c \to 1$	The U.S. Data	$\rho_c = 0.93$	$\rho_c \to 1$	The U.S. Data
$\beta = 0.97$	-0.62	-0.60		0.00	0.00	
$\beta = 0.99$	-0.41	-0.39	$[0.00, 0.21]$	0.13	0.12	$[0.30, 0.49]$
$\beta \to 1$	-0.28	-0.27		0.21	0.20	

Note: The U.S. data is from the BEA and the BLS and the brackets are the 95% confidence intervals calculated using a parametric bootstrapping method. A path of 5,000 quarters is simulated to calculate the statistics for each calibrated model. All variables are Hodrick-Prescott (HP) filtered logarithms of the original series.

Hence, the labor productivity is always countercyclical but the Solow residual is slightly pro-cyclical when firms have some market power: $\varepsilon_p < \infty$. When there exists a price-cost markup $\mu_p = \varepsilon_p/(\varepsilon_p - 1) > 1$, the Solow residual is a biased estimation of the acyclical total factor productivity (TFP) because the labor share of income underestimates the share of labor in total cost of production. Equation (2.79) shows that the larger the price-cost markup, the more pro-cyclical the Solow residual.

However, the price-cost markup in steady state cannot be indefinitely large. The larger the price-cost markup is, the smaller the total cost of production relative to output. The latter is lower bounded by the sum of the observed labor share of income and the share of depreciation cost in output as shown by the last inequality in expression (2.79). Therefore, the highest price-cost markup and the most pro-cyclical Solow residual can be achieved when the real interest rate in steady state converges to zero: $r \to 0$ or $\beta \to 1$.

Table 2.4 shows that even when $\beta \to 1$, the Solow residual is still not pro-cyclical enough compared to what is observed in the data.

Perhaps surprisingly, when the model is extended to incorporate some hoarded labor, the above results do not change. Consider an aggregate production function that is of constant returns to scale: $y = F(l_v, l_f, k)$, where l_v is the variable labor and l_f is the quasi-fixed labor that does not fluctuate much at the business cycle frequency. Hence, the business cycle fluctuations of labor is mainly given by the fluctuations of the variable labor: $\hat{l}_t \approx (l_v/l)\, \nu \hat{y}_t$, where $\nu = (wl_v)^{-1}(wl + (r+\delta)k)$ is the inverse of the share of variable labor in the total production cost. A simple calculation reveals that the cyclicality of the labor productivity and the cyclicality of the Solow residual are still upper bounded by the same ratios as in expressions (2.78) and (2.79) respectively.

2.5 A Possible Solution: Capacity Competition Externality

This section shows that the capacity underutilization (CU) model developed in Chapter 1 provides a possible solution to the issues discussed above. In the CU model, there is a capacity competition externality. If a firm expands its capacity while others do not, it steals profitable demand from others. Hence, capacity is underutilized at the firm-level. However, from an aggregate perspective, the extra capacity invested due to the capacity competition externality has no extra value. Hence, capital accumulation is not partially efficient in the CU model. I show that when the capacity competition externality is strong enough, the economy will exhibit chronic excess capacity. With a Leontief production technology, capital resources are slack locally around the steady state where capacity is in excess. The CU model is then calibrated to match the statistics observed in the U.S. data. The quantitative results show that the model can generate a positive co-movement between consumption, investment, and hours, a large relative volatility of investment to consumption, an acyclical real wage rate, a slightly pro-cyclical labor productivity, and a strongly pro-cyclical Solow residual under consumption demand shocks.

2.5.1 Setup of the CU Model

Households

There is a continuum of identical households as described in section 2.2. The labor market is assumed to be perfectly competitive ($\varepsilon_{w,t} \to \infty$ for all $t \geq 0$). The one-period utility of a representative household is assumed to be of the following functional form:

$$z_{c,t} u\left(\frac{c_t}{z_{c,t}}\right) + \bar{\omega} v\left(1 - l_t\right) = \begin{cases} \phi z_{c,t} \frac{\left(c_t/(\phi z_{c,t})\right)^{1-\gamma} - 1}{1-\gamma} + \bar{\omega} - \bar{\omega} l_t, & \gamma \neq 1 \\ \phi z_{c,t} \ln\left(c_t/z_{c,t}\right) + \bar{\omega} - \bar{\omega} l_t, & \gamma = 1 \end{cases} \tag{2.80}$$

where the utility of consumption is of constant relative risk aversion (CRRA) and the utility of leisure is linear as in the indivisible labor theory.

Technology and Capacity

There is a continuum of identical firms with measure one. Each is indexed by $j \in [0, 1]$ and operates with a Leontief production technology given by the following production function:

$$y_{j,t} = \min\left\{Ak_{j,t}, \frac{l_{f,j,t}}{\alpha_f}, \frac{l_{v,j,t}}{\alpha_v}\right\}, \tag{2.81}$$

where $l_{v,j,t}$ is the direct labor hired by firm j, $l_{f,j,t}$ is the indirect labor, and $\alpha_f^{-1} > 0$ gives the productivity of the indirect labor. Direct labor is variable. Indirect labor, however, is predetermined like capital stock. By Definition 2.3, the capacity of firm j is given by the output level at which the SRAC curve is tangent to the LRAC curve:

$$\bar{y}_{j,t} = \min\left\{Ak_{j,t}, \frac{l_{f,j,t}}{\alpha_f}\right\}. \tag{2.82}$$

The SRAC of firm j is given by

$$\mathbf{SRAC}\left(y_{j,t}, k_{j,t}, l_{f,j,t}; w_t\right) = \frac{C\left(y_{j,t}, k_{j,t}, l_{f,j,t}; w_t\right) + (r+\delta)\, k_{j,t} + w_t l_{f,j,t}}{y_{j,t}}, \tag{2.83}$$

where the indirect labor and the capital stock are both short run fixed factors and the variable cost function C is given by $C\left(y, k, l_f; w_t\right) = w_t \alpha_v y$, for all $y \leq \min\left\{Ak, {}^{l_f}/_{\alpha_f}\right\}$. The LRAC is the minimum of the SRAC when both the indirect labor and the capital stock are allowed to be adjusted. Equation (2.82) shows that the parameter α_f can be interpreted as the indirect labor required per unit of capacity supplied.

The goods produced by firms are perfect substitutes and can be used either as consumption or as investment.

Buyers and the Goods Market

Households or firms decide how many goods shall be purchased based on the aggregate price P_t and send out their buyers to purchase the goods for them. All buyers are the same. If a buyer chooses to purchase the goods from firm j the payoff is given by $v_t(j) = -\ln\left({}^{P_{j,t}}/_{P_t}\right)$, which is strictly decreasing in the real price charged by the firm.

A buyer has to process some price information in order to direct her actions towards the desired.

If no price information is processed, the matching between the demand from buyers and the capacity supplied by firms is assumed to be *completely random*. In this case, the likelihood of purchasing from firm j is given by $n_t^*(j) = {}^{\bar{y}_{j,t}}/_{\bar{y}_t}$, where $\bar{y}_t \equiv \int_0^1 \bar{y}_{j,t} dj$ is the total capacity supplied by firms. n_t^* is the most inattentive behavior of buyers and is called the *default* purchasing behavior.

To deviate from the default, buyers have to process some price information and incur some information processing cost. Let n_t be the purchasing behavior eventually obtained by a buyer. $n_t(j)$ gives the likelihood that the buyer purchases from firm j. The amount of information that needs to be processed is given by the relative entropy of the behavior chosen by the buyer n_t with respect to the default n_t^*:

$$D_{KL}(n_t||n_t^*) = \int_0^1 n_t(j) \ln\left(\frac{n_t(j)}{n_t^*(j)}\right) dj. \tag{2.84}$$

A representative buyer's problem can now be characterized as:

$$\max_{n_t \geq 0} - \int_0^1 \ln\left(\frac{P_{j,t}}{P_t}\right) n_t(j)\, dj - \Lambda D_{KL}(n_t||n_t^*), \tag{2.85}$$

subject to $\int_0^1 n_t(j)\, dj = 1$, where $\Lambda \in (0,1)$ is the unit cost of processing information.

As demonstrated in section 1.2.2 of Chapter 1, the behavior of the buyers yields the following demand system:

$$\forall j : \, y_{j,t} = n_t(j)\, y_t = \frac{\bar{y}_{j,t} P_{j,t}^{-\frac{1}{\Lambda}}}{\int_0^1 \bar{y}_{j,t} P_{j,t}^{-\frac{1}{\Lambda}} dj} y_t, \tag{2.86}$$

where y_t is the total amount goods demanded. The aggregate price P_t is simply the average price of the goods purchased by buyers: $P_t = \int_0^1 P_{j,t} n_t(j)\, dj$.

Equation 2.86 shows that the demand for the goods produced by firm j is proportional to the capacity held by the firm and is negatively related to the price charged by the firm. The higher the unit cost of processing information Λ is, the more inattentive to prices buyers are, and the lower the price elasticity of demand Λ^{-1}.

In addition to the usual price competition, there is a capacity competition among firms. To avoid thinking about prices, buyers tend to behave in an undirected way. Thus, firms with a larger capacity are more likely to be visited by buyers. For a given amount of total demand from buyers, the demand gained due to capacity expansion in one firm must be "stolen" from others.

Firms

Each firm maximizes the present value of dividend flows $d_{j,t}$:

$$\max_{\{P_{j,t},d_{j,t},y_{j,t},\bar{y}_{j,t},i_{j,t},k_{j,t+1},l_{f,j,t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} d_{j,t} \right), \tag{2.87}$$

subject to the demand curve (2.86), the resource constraint

$$d_{j,t} + i_{j,t} = \left(\frac{P_{j,t}}{P_t} - w_t \alpha_v \right) y_{j,t} - w_t l_{f,j,t}, \tag{2.88}$$

the capacity supplied $\bar{y}_{j,t} = \min \{ A k_{j,t}, {l_{f,j,t}}/{\alpha_f} \}$, the capacity constraint $y_{j,t} \leq \bar{y}_{j,t}$, and the law of motion for capital

$$k_{j,t+1} = k_{j,t} (1 - \delta) + i_{j,t} (1 - S (i_{j,t}, i_{j,t-1}, k_{j,t})), \tag{2.89}$$

where S is the adjustment cost function specified as in equation (2.76).

2.5.2 Properties of the CU Model

Capacity Competition Externality and Inefficient Capital Accumulation

Consider a symmetric equilibrium. Let q_t denotes the value of capital. A representative firm's investment condition is given by

$$q_t = \mathbb{E}_t \left(\beta \frac{\lambda_{t+1}}{\lambda_t} \left(A (\Lambda u_{t+1} + \mu_{t+1} - w_{t+1} \alpha_f) + q_{t+1} \left(1 - \delta - i_{t+1} \frac{\partial S_{t+1}}{\partial k_{t+1}} \right) \right) \right), \tag{2.90}$$

where $u_{t+1} \equiv {y_{t+1}}/{\bar{y}_{t+1}}$ is the capacity utilization rate and $\mu_{t+1} \geq 0$ is the Lagrangian multiplier of the capacity constraint. Equation (2.90) shows that the return of capital comes from two sources. First, the capacity constraint can be relaxed by capacity expansion: μ_{t+1}. Second, profitable demand can be stolen by capacity expansion: Λu_{t+1}.

Note that firms are able to charge a positive net markup because buyers are not fully attentive to prices. Hence, if a firm expands its capacity while other firms do not, it steals profits from others by stealing demand. Because of this capacity competition externality, capital has an extra value in addition to the value of relaxing the capacity constraint.

If the capital accumulation decisions made by firms are partially efficient, the present value of aggregate costs should be minimized:

$$\min_{\{i_t, l_{f,t+1}, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{\lambda_t}{\lambda_0} \left(\mathcal{C}\left(y_t, k_t, l_{f,t}; w_t\right) + w_t l_{f,t} + i_t \right), \tag{2.91}$$

subject to the law of motion for capital (2.89), where the aggregate variable cost function $\mathcal{C}$ is the same as the variable cost function at the firm-level: $\mathcal{C}\left(y, k, l_f; w_t\right) = w_t \alpha_v y$, for all $y \leq \min\left\{Ak, {}^{l_f}/_{\alpha_f}\right\}$.

However, the optimal investment condition obtained from the minimization problem above is given by

$$q_t = \mathbb{E}_t \left(\beta \frac{\lambda_{t+1}}{\lambda_t} \left(A\left(\mu_{t+1} - w_{t+1}\alpha_f\right) + q_{t+1} \left(1 - \delta - i_{t+1} \frac{\partial S_{t+1}}{\partial k_{t+1}} \right) \right) \right), \tag{2.92}$$

which shows that the only value of capital in aggregate is to relax the capacity constraint.

Hence, the capital accumulation decisions made by firms are not partially efficient.

Chronic Excess Capacity and Capital Resource Slackness

If the unit cost of processing information is sufficiently large, the capacity competition externality will be strong enough to cause chronic excess capacity.

Particularly, if and only if $\Lambda \in \left(\frac{r+\delta}{A} + w\alpha_f, 1 \right)$, capacity at the firm-level will be underutilized in steady state. Since the aggregate variable cost function $\mathcal{C}$ is the same as the variable cost function C at the firm-level, aggregate capacity is also underutilized. Thus, the economy exhibits chronic excess capacity.

Finally, in the steady state where capacity is in excess, a marginal decrease in capital will not cause an increase in the real MC as the real MC curve is flat: $\mathcal{C}_{yk} = \mathcal{C}_{yy} \frac{y}{k} = 0$. Thus, capital resources are slack locally around the steady state.

2.5.3 Calibration

To exam the model dynamics quantitatively, the model parameters are calibrated to match the statistics observed in the U.S. data. An advantage of the calibration exercise conducted here over the Bayesian estimation conducted in section 1.5 of Chapter 1 is that the parameters values are determined more transparently.

We have twelve parameters to calibrate. The depreciation rate δ is calibrated to match the average ratio of gross private domestic investment to private fixed assets from 1947 to 2016 in the NIPA published by the BEA.

The curvature of the consumption utility function γ is chosen to give an elasticity of intertemporal substitution of 0.5. In fact, because of a locally flat real AS curve, the value of γ does not matter for the local dynamics of the CU model and consumption is directly determined by consumption demand: $\hat{c}_t = \hat{z}_{c,t}$.

The persistence of consumption demand shocks ρ_c can be calibrated to match the autoregressive coefficient of the consumption series published by the BEA from the first quarter of 1948 to the first quarter of 2017.

The rest of the nine parameters are jointly calibrated to achieve the following targets.

The price in utils in steady state λ is normalized to 1. This target is mostly associated with the scaling parameter ϕ in the representative household's utility function.

The size of output in steady state y is normalized to 1. This target is mostly associated with the dis-utility of labor $\bar{\omega}$, which affects the size of the economy through the supply of labor.

The capacity utilization rate in steady state u is matched to the average of the total industry capacity utilization rate from 1967 to 2016 reported by the Federal Reserve Board (FRB). This target is mostly associated with the subjective discount factor β, which affects the opportunity cost of holding capacity.

The investment to output ratio in steady state $^{i}/_{y}$ is matched to the average ratio of gross private domestic investment to GDP from 1947 to 2016 in the NIPA. This target is mostly associated with the productivity of capital A, which affects the capital to output ratio.

The labor underutilization rate in steady state is defined as one minus the ratio of the labor hours actually utilized to the total labor hours that the representative household can potentially supply: $1 - l$, where the total hours that the representative household could supply is normalized to one. I choose the average of U-5 and U-6 from 1994 to 2016 published by the BLS as a target for the labor underutilization rate.This target is mostly associated with the direct labor required per unit of output α_v, which affects the demand for labor.

The labor share of income in steady state $^{wl}/_{y}$ is matched to the average labor share of income estimated by the BLS from 1946 to 2016. This target is mostly associated with the unit cost of processing information Λ, which affects the size of the monopolistic profit and thus the labor share of income.

The cyclicality of the labor productivity generated by the model is matched to the same statistics calculated based on the output and hours data published by the BEA and the BLS from the first quarter of 1948 to the first quarter of 2017. This target is mostly associated with the indirect labor required per unit of capacity α_f. Note that locally around the steady state, the fluctuation of hours is given by

$$\hat{l}_t = \left(1 - \frac{\alpha_f A k}{l}\right) \hat{y}_t + \frac{\alpha_f A k}{l} \hat{k}_t. \tag{2.93}$$

Ignoring the cyclical fluctuations of capital, we have that the cyclicality of the labor productivity is closely linked to the share of indirect labor hours in total hours worked:

$$\beta_y^{LP} \equiv \frac{\mathbf{Cov}\left(\hat{y}_t - \hat{l}_t, \hat{y}_t\right)}{\mathbf{Var}\left(\hat{y}_y\right)} \approx \frac{\alpha_f A k}{l} \geq 0. \tag{2.94}$$

The relative volatility of investment to consumption generated by the model is matched to the same statistics calculated based on the consumption and investment series published by the BEA from the first quarter of 1948 to the first quarter of 2017. This target is mostly associated with the curvature of the capital adjustment cost ϕ_k, which dampens the response of investment to consumption demand shocks.

The auto-correlation of investment generated by the model is matched to the same statistics calculated based on the investment series published by the BEA from the first quarter of 1948 to the first quarter of 2017. This target is mostly associated with the curvature of the investment adjustment cost ϕ_i, which helps generate a hump-shaped response of investment to consumption demand shocks.

Table 2.5 summarizes the calibrated parameter values and their mostly associated targets.

2.5.4 Quantitative Results

Table 2.6 compares the business cycle statistics of the model generated series with those in the data. The results show that the CU model is able to match all the five stylized business-cycle facts using consumption demand shocks. Particularly, the model generates a positive correlation between consumption, investment, and hours, a highly volatile investment, an

Table 2.5: Parameters and Calibration Targets – The CU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	2.0000	Elasticity of inter-temporal substitution 0.5
ϕ	0.8300	Price in utils normalized to 1
$\bar{\omega}$	0.6798	Output normalized to 1
β	0.9747	Capacity utilization rate 0.8
A	0.1544	Investment to output ratio 0.17
α_v	0.8117	Labor underutilization rate 0.088
α_f	0.0803	Cyclicality of the labor productivity 0.11
Λ	0.4482	Labor share of income 0.62
ϕ_k	2.6870	Relative volatility of investment to consumption 6.49
ϕ_i	0.0460	Auto-correlation of investment 0.80
ρ_c	0.9918	Auto-regressive coefficient of consumption 0.9918

Table 2.6: Business Cycle Statistics – The CU Model

	$\rho(c,h)$	$\rho(c,i)$	σ_i/σ_c	β_y^w	β_y^{LP}	β_y^{SR}
$z_{c,t}$	0.98	0.92	6.49	0.00	0.11	0.44
The U.S. Data	0.58	0.68	6.49	0.08	0.11	0.39
	$[0.42, 0.72]$	$[0.57, 0.78]$	$[5.55, 7.42]$	$[-0.03, 0.18]$	$[0.00, 0.21]$	$[0.30, 0.49]$

Note: $\rho(x_1, x_2)$ stands for the correlation coefficient between x_1 and x_2. σ_i/σ_c is the relative volatility of investment to consumption. The brackets are the 95% confidence intervals calculated using a parametric bootstrapping method. The U.S. data is from the BEA and the BLS. A path of 5,000 quarters is simulated to calculate the statistics for the calibrated CU model. All variables are Hodrick-Prescott (HP) filtered logarithms of the original series.

acyclical real wage rate, a slightly pro-cyclical labor productivity, and a strongly pro-cyclical Solow residual.

Figure 2.6 shows the fitted values investment and hours when consumption demand shocks are chosen to match the consumption data. The results are quite close to the investment and hours series observed in the data and shows that the CU model is able to capture most of the business cycle co-movement among consumption, investment, and hours under consumption demand shocks.

Consumption, Investment, and Hours Fitted by the CU Model

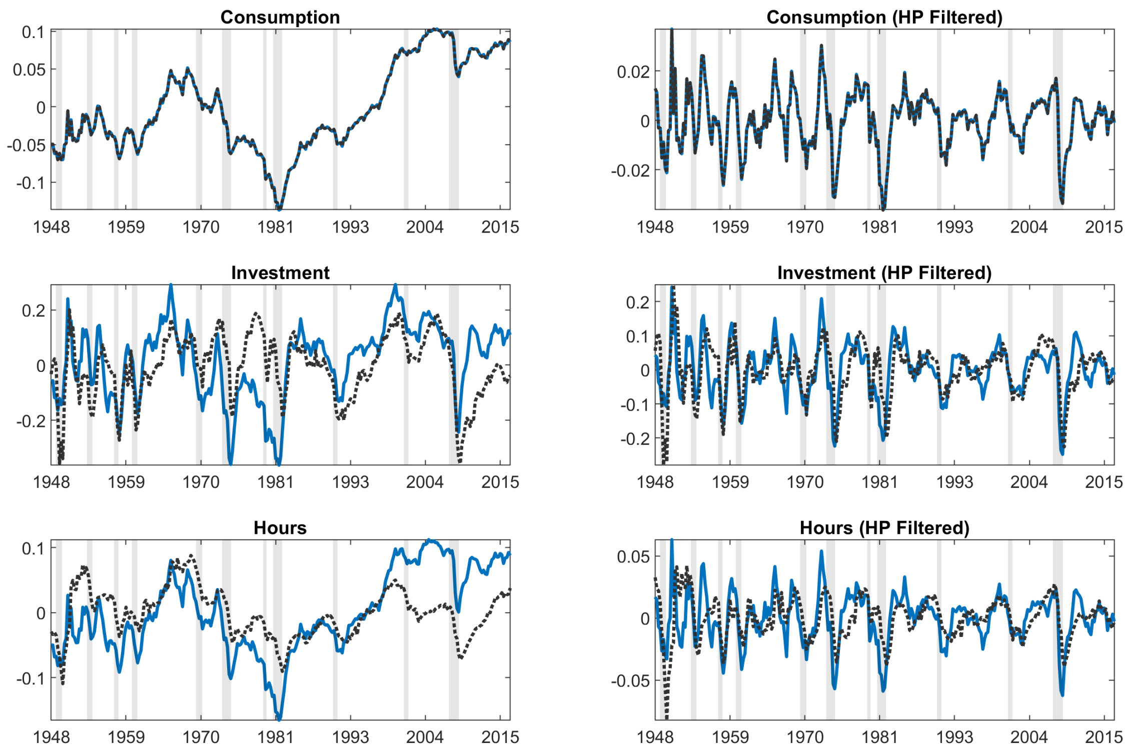

Figure 2.6: The solid lines are the values generated by the CU model under consumption demand shocks. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA and the BLS. All variables are logarithms of the original series.

2.6 Conclusion

This chapter studies a standard one-good economy to illustrate the difficulties of demand driving business cycles in a wide class of models in the existing literature. The variable capital utilization model (e.g., Christiano et al., 2005 and Smets and Wouters, 2007), the variable plant utilization model (e.g., Cooley et al., 1995), the precautionary capacity model (e.g., Fagnart et al., 1997 and Fagnart et al., 1999), and the competitive search in the goods market model (e.g., Bai et al., 2012) can all be regarded as special cases of the standard one-good economy. The following properties of the standard one-good economy are obtained.

First, the standard new Keynesian mechanism relies on countercyclical markups to move real aggregate supply pro-cyclically. Second, when prices are perfectly flexible, the standard one-good economy is subject to the Barro-King curse under investment demand shocks. Third, capital accumulation is partially efficient. Hence, aggregate capacity must be fully utilized in steady state and capital resources must be tight locally around the steady state. Fourth, when prices are perfectly flexible and capital resources are tight, either the relative volatility of investment to consumption is too small or the real wage rate is too countercyclical under consumption demand shocks. Fifth, under both types of real demand shocks, the labor productivity is too countercyclical and the Solow residual is not pro-cyclical enough.

The capacity underutilization (CU) model developed in Chapter 1 provides a possible solution to the issues discussed above. Firms are encouraged to over-invest in capital because capacity expansion can steal profitable demand from other firms but the extra capacity invested has no extra benefit in aggregate. Because of this capacity competition externality, capital accumulation is not partially efficient, the economy exhibits chronic excess capacity, and capital resources can be slack locally around the steady state. Quantitative results indicate that consumption demand shocks alone can generate realistic business cycle fluctuations in the CU model. The results suggest that the capacity competition externality is a powerful mechanism that can alleviate the dependence on supply channels to drive business cycles, reviving the original idea proposed by Keynes.

Chapter 3

Capacity Competition Externality and Chronic Excess Capacity

3.1 Introduction

Chapter 2 shows that if capital accumulation is too efficient, aggregate capacity will be fully utilized in steady state and capital resources must be tight locally around the steady state. By contrast, if there is a capacity competition externality, the return of capital at the firm-level can be larger than the return of capital in aggregate. In this case, too much capital will be inefficiently accumulated and capacity will be chronically in excess.

As illustrated in both Chapter 1 and Chapter 2, the existence of chronic excess capacity can have important consequences on the model dynamics. Real demand shocks have difficulties driving business cycles in models with efficient capital or capacity utilization but can drive business cycles quite easily in an economy with excess capacity.

In general, what kind of goods market structure could feature the capacity competition externality that causes chronic excess capacity? To answer this question, I start with a basic setup in which the goods market is populated with infinitely many small firms. The setup incorporates the standard Dixit-Stiglitz monopolistic competition framework as a special case. I then proceed to narrow down the class of market structures by imposing three intuitive assumptions.

First, each firm is able to scale up (or down) its market share proportionally by scaling up (or down) its relative capacity. Therefore, in addition to the usual price competition, there is a capacity competition among firms. For a given amount of total demand requested, if a firm expands its capacity while other firms do not, the firm can "steal" demand from others. With a production technology of constant returns to scale, the scalability assumption implies that the gross profit earned by a firm is proportional to the firm's capital stock. Hence, a firm can replicate its success in the market by expanding its capital or capacity, an implication that is quite reasonable for small firms. In this case, firms have no incentive to merge nor to split up. In fact, I show that when output is demand determined, firms have no incentive to merge nor to split up if and only if the demand allocated to each firm varies proportionally with the firm's capital stock. Hence, like constant returns to scale, the scalability assumption can be regarded as a good benchmark.

Second, each firm is able to charge a sufficiently high price, so if all goods produced at full capacity can be sold, the firm can make a positive net profit. The profitability assumption ensures that firms have an incentive to participate in the capacity competition described above because the demand stolen due to the capacity expanded is profitable enough. The scalability assumption and the profitability assumption imply that there is a negative capacity competition externality: if a firm expands its capacity while other firms do not, it steals profits from others by stealing demand. Because of this negative externality, firms tends to hold too much capacity. I show that adding this profitability assumption to the scalability assumption is necessary and sufficient to see long-term capacity underutilization at the firm-level.

Third, the invested capital has no positive externality that could potentially offset the capacity competition externality. In general, the invested capital may have a positive externality, such as increasing the efficiency of how goods produced by firms are aggregated into a final good or reducing the cost that has to be incurred by buyers to purchase the goods from firms. The no positive externality assumption ensures that the negative capacity competition externality dominates. In this case, any extra capacity held by firms is in excess.

I prove that the economy will exhibit chronic excess capacity as long as the three assumptions above are true.

There are several different ways that could micro-found this kind of goods market structure. Three examples are provided in this chapter.

First, the basic capacity underutilization (CU) developed in Chapter 1 is a possible microfoundation. Buyers are inattentive to prices and tend to search for capacity blindly and randomly. As a result, firms with a larger capacity are more likely to be visited and the scalability assumption is satisfied. Since buyers are not fully attentive to prices, firms are able to charge a positive net markup. Hence, the profitability assumption can be satisfied. Though buyers in aggregate act *as if* goods are imperfect substitutes, all goods are perfect substitutes. Hence, the capital invested has no positive externality. As expected, the economy exhibits chronic excess capacity.

Second, following Michaillat and Saez (2015), I incorporate the standard labor market matching framework into the goods market. The random matching takes place between the demand from buyers and the capacity supplied by firms. Prices are determined via a Nash bargaining process after matching takes place. I show that if the bargaining power of firms is sufficiently strong, the profitability assumption is satisfied. Furthermore, the scalability assumption holds because the random matching process allows firms with a larger capacity to sell more. Finally, since all goods are perfect substitutes as in the basic CU model, the no positive externality assumption is satisfied. With the same Leontief production technology and the same household's problem, I show that the main properties of the random matching model are the same as those of the basic CU model.

Third, I present a model in which investment in capital not only expands the production potential but also expands the variety of the goods supplied in the market. With a standard Dixit-Stiglitz setup, an increase in variety allows firms to steal demand from others. Hence, the scalability assumption holds. Since the goods produced by different firms are imperfect substitutes, firms can charge a positive net markup and the profitability assumption can be satisfied. Finally, I show that the value of expanding variety can be internalized by firms so that the capital invested has no positive externality in equilibrium. Again, with the same Leontief production technology and the same household's problem, I find that the properties of this variety expansion model are exactly the same as those of the basic CU model.

To sum up, the scalability, the profitability, and the no positive externality assumptions together describe a goods market structure in which the capacity competition externality dominates, and ensure that the economy exhibits chronic excess capacity. These three assumptions capture the key mechanism that drives the results of the capacity underutilization models studied in Chapter 1 and Chapter 2. In addition, there are several different ways to micro-found this goods market structure. Different models correspond to different interpretations but the main properties of the models remain unchanged.

The rest of this chapter is organized as follows. Section 3.2 introduces the three key assumptions. Sections 3.3-3.5 provide three examples. Particularly, section 3.3 reviews and generalizes the basic capacity underutilization (CU) model developed in Chapter 1. Section 3.4 incorporates the standard random matching and Nash bargaining framework into the goods market. Section 3.5 presents a model in which invested capital expands variety. Section 3.6 concludes.

3.2 Capacity Competition Externality and Chronic Excess Capacity

This section presents a partial equilibrium model in which prices and demand are determined exogenously. I start with a basic setup, which incorporates the standard Dixit-Stiglitz (DS) monopolistic competition framework as a special case. Three key assumptions are then introduced to describe the goods market structure that features a capacity competition externality and chronic excess capacity. First, each firm is able to expand its market share by expanding its relative capacity in the market. Hence, in addition to the usual price competition, there is a capacity competition among firms. Second, each firm is able to charge a sufficiently high price to make a positive net profit. Hence, there is a negative capacity competition externality: if a firm expands its capacity while other firms do not, it steals profits from others by stealing demand. Third, the invested capital has no positive externality. Hence, any capacity underutilized at the firm-level is in excess. I show that these three assumptions can cause chronic excess capacity in the economy.

3.2.1 Basic Setup

Production Technology

There is a continuum of firms indexed by $j \in [0,1]$. Firms use capital and hire labor to produce goods. They are price takers in the labor market. The production technology of firm j at time $t \geq 0$ can be represented by a variable cost function $C\left(y_{j,t}, k_{j,t}; w_t\right)$, where $y_{j,t}$ is the output of the firm, $k_{j,t}$ is the capital stock, and $w_t > 0$ is the real wage rate. The variable cost function is assumed to satisfy the following regularity conditions:

1. (*Standard properties*) The variable cost function C is non-negative, continuous, convex, increasing in terms of output, and decreasing in terms of capital.
2. (*Necessity of capital*) The domain of the variable cost function $\mathbf{dom}\,(C)$ is either given by $\{(y,k) \mid y \geq 0, k \geq \chi y\}$, where $\chi > 0$, or given by $\{(y,k) \mid y \geq 0, k > \chi y\} \cup$

$\{(0,0)\}$, where $\chi \geq 0$. In the latter case, if $y > 0$ and $k \to \chi y$, the variable cost function converges to infinity: $C(y,k;w) \to \infty$.

3. (*Constant returns to scale*) The variable cost function C is homogeneous of degree one.

4. (*Smoothly diminishing marginal returns to capital*) If for a given y, there exists an interval $[a,b]$ on which the variable cost function is strictly decreasing in capital, then C is twice differentiable in capital on the interval (a,b) such that $C_k(y,k;w) < 0$ and $C_{kk}(y,k;w) > 0$.

Note that if $\chi > 0$, there is a production limit which is the smallest upper bound on output: $y \leq \chi^{-1}k$. The Leontief production technology, the Cobb-Douglas production technology, and the production technology with a constant elasticity of substitution (CES) between capital and labor smaller than one, are all special cases of the technology described above.

Definition 3.1. (*SRAC*) The short run average total cost (SRAC) is defined as

$$\mathbf{SRAC}(y,k;w_t) \equiv \frac{C(y,k;w_t) + (r+\delta)k}{y}, \tag{3.1}$$

where the output level $y > 0$ is positive, $\delta > 0$ is the capital depreciation rate, $r = Q^{-1} - 1 > 0$ is the real interest rate in steady state, and $Q \in (0,1)$ is the stochastic discount factor in steady state.

Claim 3.2. There exists a unique capital level that minimizes the short run average total cost (SRAC):

$$k^*(y,w_t) \equiv \arg \min_{k \in \{x|(y,x) \in \mathbf{dom}(C)\}} \mathbf{SRAC}(y,k;w_t); \tag{3.2}$$

and that capital level must be positive: $k^*(y,w_t) > 0$.

Proof. If there is a feasible production limit, the SRAC is continuous and convex in terms of capital on the interval $[\chi y, \infty)$. When $k \to \infty$, we have $\mathbf{SRAC}(y,k;w_t) \geq \frac{(r+\delta)k}{y} \to \infty$. Hence, there exists $k^* > 0$ that minimizes the SRAC. If there is no feasible production limit, the SRAC is continuous and convex in terms of capital on the interval $(\chi y, \infty)$. When $k \to \chi y$ or $k \to \infty$, we have $\mathbf{SRAC}(y,k;w_t) \geq \max\left\{\frac{C(y,k;w_t)}{y}, \frac{(r+\delta)k}{y}\right\} \to \infty$. Hence, there exists $k^* > 0$ that minimizes the SRAC.

If there exists $k' \neq k^*$ that also minimizes the SRAC, any $k \in [a,b]$ minimizes the SRAC, where $a = \min\{k', k^*\} < b = \max\{k', k^*\}$. For any $k_1, k_2 \in [a,b]$ and $k_1 > k_2$, we have $C(y,k_1;w_t) = C(y,k_2;w_t) + (r_t+\delta)(k_2 - k_1) < C(y,k_2;w_t)$. Hence, by the assumed smoothly diminishing marginal returns to capital, C is twice differentiable in terms of capital on the interval (a, b) such that $C_{kk}(y, k; w_t) > 0$, a contradiction to the previous conclusion that any $k \in [a, b]$ minimizes the SRAC. Hence, k^* must be unique. □

Definition 3.3. (*LRAC*) The long run average total cost (LRAC) is the minimum of the SRAC when capital stock can be adjusted:

$$\mathbf{LRAC}(y; w_t) \equiv \min_{k \in \{x | (y,x) \in \mathbf{dom}(C)\}} \mathbf{SRAC}(y, k; w_t). \tag{3.3}$$

Claim 3.4. The LRAC curve exists and is flat.

Proof. By definition, the long run average total cost curve can be expressed as:

$$\begin{aligned}
\mathbf{LRAC}(y; w_t) &= \min_{k \in \{x | (y,x) \in \mathbf{dom}(C)\}} \left\{ C\left(1, \frac{k}{y}; w_t\right) + (r + \delta) \frac{k}{y} \right\} \\
&= \min_{\tilde{k} \in \{x | (1,x) \in \mathbf{dom}(C)\}} \left\{ C\left(1, \tilde{k}; w_t\right) + (r + \delta) \tilde{k} \right\} \\
&= C(1, k^*(1, w_t); w_t) + (r + \delta) k^*(1, w_t),
\end{aligned}$$

where $k^*(1, w_t)$ exists according to Claim 3.2. It is clear that the LRAC is independent of the output level $y > 0$. Hence, the LRAC curve is flat. □

Definition 3.5. (*Capacity*) Capacity is the unique output level $\bar{y}$ at which the SRAC curve is tangent to the LRAC curve as long as the capital level is positive ($k > 0$). If the capital level is zero ($k = 0$), capacity is zero.

Claim 3.6. The capacity of firm j is properly defined and varies proportionally with the firm's capital stock $k_{j,t} \geq 0$. Particularly, for a given aggregate capital level $k_{j,t} \geq 0$ and a real wage rate $w_t > 0$, the capacity of firm j is given by

$$\bar{y}_{j,t} = \frac{k_{j,t}}{k^*(1, w_t)}, \tag{3.4}$$

where k^* is a function that gives the capital level that minimizes the corresponding aggregate SRAC (see Claim 3.2).

Proof. Consider the case where the capital stock is positive ($k_{j,t} > 0$). Since the LRAC curve is flat, the tangent point $(\bar{y}_{j,t}, k_{j,t})$, if there is any, must minimize the SRAC curve:

$$\min_{y \in \{x > 0 | (x, k_{j,t}) \in \mathbf{dom}(C)\}} \left\{ C\left(1, \frac{k_{j,t}}{y}; w_t\right) + (r + \delta) \frac{k_{j,t}}{y} \right\}.$$

The above minimization problem is equivalent to finding the capital to output ratio that minimizes the SRAC:

$$\min_{\tilde{k}\in\{x|(1,x)\in\mathbf{dom}(C)\}}\left\{C\left(1,\tilde{k};w_t\right)+(r+\delta)\,\tilde{k}\right\}.$$

According to Claim 3.2, $k^*(1,w_t)$ is the unique capital to output ratio that minimizes the SRAC. Therefore, the tangent point exists uniquely and it is proper to define the capacity as the output level at the unique tangent point: $\bar{y}_{j,t}={}^{k_{j,t}}/_{k^*(1,w_t)}$. If $k_{j,t}=0$, capacity is also properly defined as zero. □

Demand System

Let us now move to the demand side. Suppose that the demand faced by firm j at time t is a function of nominal prices, capital levels, and the total demand requested $\mathcal{D}_t>0$:

$$D_{j,t}=D\left(P_{j,t},k_{j,t},\mathbf{P}_t,\mathbf{k}_t,\mathcal{D}_t\right), \tag{3.5}$$

where $P_{j,t}>0$ is the nominal price charged by firm j, $\mathbf{P}_t\equiv\left(P_{j,t}\right)_{j\in[0,1]}>0$ collects all the nominal prices charged by firms, and $\mathbf{k}_t\equiv\left(k_{j,t}\right)_{j\in[0,1]}\geq 0$ collects all the capital stock held by firms. D is assumed to have the following properties:

1. (*Non-negativity*) The demand function is non-negative.
2. (*Aggregation*) The sum of all demand requested is the total demand requested:

$$\int_0^1 D\left(P_{j,t},k_{j,t},\mathbf{P}_t,\mathbf{k}_t,\mathcal{D}_t\right)dj\equiv\mathcal{D}_t.$$

3. (*Neutral to absolute price-level*) The demand function is homogeneous of degree zero in terms of nominal prices:

$$\forall\alpha>0:D\left(\alpha P_{j,t},k_{j,t},\alpha\mathbf{P}_t,\mathbf{k}_t,\mathcal{D}_t\right)=D\left(P_{j,t},k_{j,t},\mathbf{P}_t,\mathbf{k}_t,\mathcal{D}_t\right).$$

4. (*Symmetric*) For any $\mathbf{P}'_t=\left(P_{\mathcal{I}(j),t}\right)_{j\in[0,1]}$ and $\mathbf{k}'_t=\left(k_{\mathcal{I}(j),t}\right)_{j\in[0,1]}$, where $\mathcal{I}:[0,1]\to[0,1]$ is a one-to-one mapping, we have

$$D\left(P_{j,t},k_{j,t},\mathbf{P}'_t,\mathbf{k}'_t,\mathcal{D}_t\right)=D\left(P_{j,t},k_{j,t},\mathbf{P}_t,\mathbf{k}_t,\mathcal{D}_t\right).$$

The real price charged by firm j at time t is the ratio of the nominal price to the aggregate price: $p_{j,t} \equiv {}^{P_{j,t}}/_{\mathcal{P}_t}$, where $\mathcal{P}_t > 0$ is the aggregate price. Because the demand function is homogeneous of degree zero in terms of nominal prices, the demand function could also be expressed in terms of real prices.

Profit Maximization

To focus on capital accumulation decisions, throughout this section, all prices will be treated as exogenous. I assume that firms are unable to hold any inventory. Hence, each unit of goods sold at time t must also be produced at time t. The gross profit of firm j who produces and sells y units of goods at time t is given by:

$$\pi\left(y; p_{j,t}, k_{j,t}, w_t\right) \equiv p_{j,t} y - C\left(y, k_{j,t}; w_t\right), \tag{3.6}$$

where the gross profit function π is concave in output y because C is convex. If there exists $y_1 > y_2$ such that $\pi\left(y_1; p_{j,t}, k_{j,t}, w_t\right) > \pi\left(y_2; p_{j,t}, k_{j,t}, w_t\right)$, we have that π is strictly increasing in y for all $y \leq y_1$. Let $\mathbb{V}\left(p_{j,t}, k_{j,t}, w_t\right)$ be the set of all such y_1 below which the gross profit is strictly increasing in output:

$$\begin{aligned}
&\mathbb{V}\left(p_{j,t}, k_{j,t}, w_t\right) \equiv \\
&\qquad \left\{\left(y_1, k_{j,t}\right) \in \mathbf{dom}\left(C\right) \mid \forall y_2 \in \left[0, y_1\right), \pi\left(y_1; p_{j,t}, k_{j,t}, w_t\right) > \pi\left(y_2; p_{j,t}, k_{j,t}, w_t\right)\right\}.
\end{aligned} \tag{3.7}$$

If $\mathbb{V}$ is empty, the firm would not want to sell at all. If $\mathbb{V} \neq \emptyset$, the firm would not want to sell any more when its output has reached the supremum of $\mathbb{V}$. The output level above which the firm has no incentive to sell can be expressed as a function of the real price charged, the capital stock held, and the real wage rate:

$$S\left(p_{j,t}, k_{j,t}; w_t\right) \equiv \begin{cases} \sup \mathbb{V}\left(p_{j,t}, k_{j,t}, w_t\right), & \mathbb{V}\left(p_{j,t}, k_{j,t}, w_t\right) \neq \emptyset, \\ 0, & \mathbb{V}\left(p_{j,t}, k_{j,t}, w_t\right) = \emptyset, \end{cases} \tag{3.8}$$

which captures the firm's willingness to supply.

Claim 3.7. The firm's willingness to supply $S\left(p_{j,t}, k_{j,t}; w_t\right)$ varies proportionally with the firm's capital stock $k_{j,t}$.

Proof. The statement follows directly from the assumption that C is homogeneous of degree one, i.e., the production technology is of constant returns to scale. □

Recall that firms are unable to hold any inventory. If the goods produced are not sold, they would have to be discarded. To ensure that each unit of goods produced can make a profit, output shall not exceed demand nor the firm's willingness to supply. Hence, the optimal output of the firm is the minimum of the demand and the supply:

$$Y\left(p_{j,t}, k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t; w_t\right) \equiv \min\left\{D\left(p_{j,t}, k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t\right), S\left(p_{j,t}, k_{j,t}; w_t\right)\right\}. \tag{3.9}$$

where $\mathbf{p}_t \equiv \left(p_{j,t}\right)_{j\in[0,1]}$ collects all the real prices charged by firms. For simplicity, let $Y_t\left(p_{j,t}, k_{j,t}\right)$ denote the output function of firm j at time t.

Claim 3.8. $Y_t\left(p_{j,t}, k_{j,t}\right)$ maximizes the gross profit $\pi\left(y; p_{j,t}, k_{j,t}, w_t\right)$ of the firm subject to the demand constraint: $y \leq D\left(p_{j,t}, k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t\right)$.

Proof. Let y^* be the output, S the supply, and D the demand. Since π is concave in terms of y, π is strictly increasing on $[0, S]$ and non-increasing for any $y > S$. Hence, if $y^* = S$, the gross profit is maximized. If $y^* = D < S$, the output is demand constrained. Any $y < y^*$ decreases the profit, while any $y > y^*$ is not feasible. Hence, y^* maximizes the gross profit. □

The maximized gross profit of the firm can be expressed as:

$$\Pi_t\left(p_{j,t}, k_{j,t}\right) \equiv p_{j,t} Y_t\left(p_{j,t}, k_{j,t}\right) - C\left(Y_t\left(p_{j,t}, k_{j,t}\right), k_{j,t}; w_t\right). \tag{3.10}$$

Capital Accumulation

Each firm aims to maximize its firm value, which is the present value of its dividend flows:

$$\forall j: \max_{\left\{k_{j,t+1}\geq 0\right\}_{t\geq 0}} \mathbb{E}_0\left(\sum_{t=0}^{\infty}\left(\prod_{\tau=1}^{t} Q_\tau\right)\left(\Pi_t\left(p_{j,t}, k_{j,t}\right) - k_{j,t+1} + (1-\delta) k_{j,t}\right)\right), \tag{3.11}$$

where Q_{t+1} is the stochastic discount factor that discounts the value at time $t+1$ to t.

Partial Equilibrium

For an exogenously given stable stochastic process of the real prices $\left\{\mathbf{p}_t\right\}_{t\geq 0}$, the real wage rate $\left\{w_t\right\}_{t\geq 0}$, the stochastic discount factor $\left\{Q_t\right\}_{t\geq 0}$, and the total demand $\left\{\mathcal{D}_t\right\}_{t\geq 0}$, the partial equilibrium is a stable stochastic process of the capital stock $\left\{\mathbf{k}_{t+1}\right\}_{t\geq 0}$ that solves the firm-value maximization problem (3.11) for all firms.

3.2.2 Capacity Competition Externality

The setup so far is standard, which includes the standard Dixit-Stiglitz (DS) monopolistic competition framework as a special case. In the rest of this section, I will introduce three assumptions to describe a goods market structure that features a capacity competition externality and chronic excess capacity.

Assumption 3.9. *(Scalability-demand) A firm that scales up (or down) its capital stock also scales up (or down) the demand allocated to the firm by the same proportion:*

$$\forall\alpha > 0 : D\left(p_{j,t}, \alpha k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t\right) = \alpha D\left(p_{j,t}, k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t\right).$$

Assumption 3.9 says that a firm can expand its market share proportionally by expanding its capital stock or capacity. For a given amount of total demand requested $\mathcal{D}_t$, if a firm expands its capacity while other firms do not, the firm is stealing demand from others. Therefore, in addition to the usual price competition, there is a capacity competition among firms.

If a firm operates with a production technology of constant returns to scale, an immediate corollary of Assumption 3.9 is that the gross profit of the firm is also scalable.

Corollary 3.10. *(Scalability-gross profit) A firm that scales up (or down) its capital stock also scales up (or down) its gross profit by the same proportion.*

Proof. Let y and k be the output and the capital stock of the firm respectively. If the capital stock of the firm is scaled by $\alpha > 0$, by Assumption 3.9 and Claim 3.7, the output of the firm is scaled by α. Because the production technology is of constant returns to scale, the gross profit of the firm is now given by $\alpha p_{j,t} y - \alpha C\left(y, k; w_t\right)$, which is also scaled by α. □

The scalability assumption implies that a firm can replicate its success in the market by expanding its capital or capacity, an implication that is quite reasonable for small firms. Since the gross profit function is linear in terms of capital, firms have no incentive to merge nor to split up. Having no incentive to merge nor to split up is *almost* equivalent to Assumption 3.9.

Claim 3.11. The scalability assumption (Assumption 3.9) is true for any $p_{j,t}$ and $k_{j,t}$ such that $D\left(p_{j,t}, k_{j,t}, \mathbf{p}_t, \mathbf{k}_t, \mathcal{D}_t\right) < S\left(p_{j,t}, k_{j,t}; w_t\right)$ if and only if firm j has no incentive to merge with others nor to split up.

Proof. For simplicity, let $D_t(p_{j,t}, k_{j,t})$ denote the demand and $S_t(p_{j,t}, k_{j,t})$ denote the supply. First, note that the maximized gross profit function Π_t is homogeneous of degree one in capital if and only if the firm has no incentive to merge with others nor to split up. Second, $\forall \alpha > 0$, we have:

$$\begin{aligned}\Pi_t(p_{j,t}, \alpha k_{j,t}) &= \pi(Y_t(p_{j,t}, \alpha k_{j,t}); p_{j,t}, \alpha k_{j,t}, w_t) = \alpha \Pi_t(p_{j,t}, k_{j,t}) \\ &= p_{j,t} \alpha Y_t(p_{j,t}, k_{j,t}) - C(\alpha Y_t(p_{j,t}, k_{j,t}), \alpha k_{j,t}; w_t) \\ &= \pi(\alpha Y_t(p_{j,t}, k_{j,t}); p_{j,t}, \alpha k_{j,t}, w_t).\end{aligned}$$

According to Claim 3.7, $Y_t(p_{j,t}, \alpha k_{j,t})$ and $\alpha Y_t(p_{j,t}, k_{j,t})$ both belong to the interval $0 \leq y \leq S_t(p_{j,t}, \alpha k_{j,t})$ on which π is strictly increasing and invertable. Hence, Y_t is homogeneous of degree one in capital if and only if Π_t is homogeneous of degree one in capital.

To prove sufficiency, note that if $D_t(p_{j,t}, k_{j,t}) < S_t(p_{j,t}, k_{j,t})$, by Corollary 3.10, we have $\forall \alpha > 0$, $Y_t(p_{j,t}, \alpha k_{j,t}) = \alpha Y_t(p_{j,t}, k_{j,t})$. If $D_t(p_{j,t}, k_{j,t}) \geq S_t(p_{j,t}, k_{j,t})$, one can prove by contradiction that $D_t(p_{j,t}, \alpha k_{j,t}) \geq S_t(p_{j,t}, \alpha k_{j,t})$. Thus, by Claim 3.7, we have $\forall \alpha > 0$, $Y_t(p_{j,t}, \alpha k_{j,t}) = \alpha Y_t(p_{j,t}, k_{j,t})$.

To prove necessity, note that if $D_t(p_{j,t}, k_{j,t}) < S_t(p_{j,t}, k_{j,t})$, we must have $\forall \alpha > 0$, $D_t(p_{j,t}, \alpha k_{j,t}) < S_t(p_{j,t}, \alpha k_{j,t})$. Otherwise, we have $Y_t(p_{j,t}, \alpha k_{j,t}) = S_t(p_{j,t}, \alpha k_{j,t}) = \alpha Y_t(p_{j,t}, k_{j,t}) = \alpha D_t(p_{j,t}, k_{j,t}) < \alpha S_t(p_{j,t}, k_{j,t})$,a contradiction to Claim 3.7. Hence, we have $D_t(p_{j,t}, \alpha k_{j,t}) = Y_t(p_{j,t}, \alpha k_{j,t}) = \alpha Y_t(p_{j,t}, k_{j,t}) = \alpha D_t(p_{j,t}, k_{j,t})$. □

Claim 3.11 shows that when output is demand determined, a firm has no incentive to merge with others nor to split up if and only if the demand faced by the firm varies proportionally with the firm's capital stock. Hence, like constant returns to scale, Assumption 3.9 sets a good benchmark.

Note that Assumption 3.9 is not satisfied in the usual DS monopolistic competition market structure, in which capital has no effect on demand and the only competition among firms is a price competition. In this case, the gross profit function is concave in terms of capital, and if possible, it is profitable for a big firm to be split into small sub-firms with different brand names to increase the variety in the market. By assuming scalability, we deviate from the usual DS monopolistic competition market structure.[1]

Having the ability to scale up via capacity expansion does not mean that firms have an

[1] Later in section 3.5, I show that Assumption 3.9 can be true in the standard DS framework if investment in capital not only increases the production potential but also expands variety.

incentive to do so because capacity expansion is costly. Firms will not want to participate in the capacity competition unless the net profit is positive.

Assumption 3.12. *(Profitability or pricing power) Each firm* $j \in [0,1]$ *in equilibrium is able to charge a price above the long run average total cost (LRAC):*

$$p_{j,t} > \mathbf{LRAC}\left(w_t\right) = C\left(1, k^*\left(1, w_t\right); w_t\right) + \left(r + \delta\right) k^*\left(1, w_t\right).$$

Hence, if all goods produced at full capacity can be sold, the firm can make a positive net profit.

Assumption 3.12, together with Assumption 3.9, ensures that it is profitable for firms to expand capacity. In this case, capacity expansion has a negative externality: if a firm expands its capacity while other firms do not, it steals profits from others by stealing demand. When Assumption 3.12 holds, this capacity competition externality is strong enough such that the desire to expand capacity will not disappear unless the capacity becomes underutilized. The following proposition shows that adding Assumption 3.12 to Assumption 3.9 is sufficient and necessary for firms to have capacity underutilization in the long run.

Proposition 3.13. *(Capacity utilization) Suppose that the scalability assumption, i.e., Assumption 3.9, is true. Consider the steady state of a symmetric equilibrium in which all firms charge the same price. If the price is higher than the LRAC, capacity will be underutilized. If the price is equal to the LRAC, capacity will be fully utilized. If the price is lower than the LRAC, firms will have no incentive to hold any capacity.*

The following lemma is useful to prove Proposition 3.13. The lemma says that if the price charged by the firm is sufficiently high, the firm would be happy to produce at full capacity. Hence, if capacity is underutilized as Proposition 3.13 predicts, it is not because the cost of utilization is too high, but because demand is not high enough compared to the capacity accumulated. The lemma shows that the capacity underutilization caused by the capacity competition externality is consistent with the fact that the vast majority of firms covered in the Quarterly Survey of Plant Capacity Utilization cites insufficient demand as the main reason for capacity underutilization (Boehm et al., 2017).

Lemma 3.14. *(Willingness to produce at full capacity) If the price charged by firm* $j \in [0,1]$ *is no smaller than the long run average total cost (LRAC), the firm is willing to produce at full capacity.*

Proof. According to Claim 3.7, the firm's willingness to supply varies proportionally with capital: $S(p_{j,t}, 1; w_t) k_{j,t}$. According to Claim 3.6, the capacity of the firm is also proportional to the capital stock: $\bar{y}_{j,t} = k_{j,t} k^* (1, w_t)^{-1}$. Let $a = k^* (1, w_t)^{-1} > 0$ and $b = S(p_{j,t}, 1; w_t)$. We need only prove that $a \leq b$.

Suppose that $a > b$. Because the price is no smaller than the LRAC, we have that $p_{j,t} \geq C\left(1, a^{-1}; w_t\right) + (r+\delta) a^{-1}$. Hence, $\pi(a; p_{j,t}, 1, w_t) \geq r + \delta$. By definition, the gross profit function π is maximized by b: $\pi(b; p_{j,t}, 1, w_t) \geq \pi(a; p_{j,t}, 1, w_t)$. Thus, b must be positive, otherwise, the gross profit will be zero, which cannot be larger than the gross profit obtained at full capacity. However,

$$\begin{aligned}\mathbf{SRAC}(a, 1; w_t) &= p_{j,t} - \frac{\pi(a; p_{j,t}, 1, w_t) - (r+\delta)}{a} \\ &\geq p_{j,t} - \frac{\pi(b; p_{j,t}, 1, w_t) - (r+\delta)}{b} = \mathbf{SRAC}(b, 1; w_t),\end{aligned}$$

a contradiction to Claim 3.2 that a is the unique capital level that minimizes the SRAC. □

Equipped with Lemma 3.14, we are now ready to give a proof for Proposition 3.13.

Proof. In the steady state of a symmetric equilibrium, we have that for all $j \in [0,1]$, $p_j = p$ and $k_j = k$. The optimal capital accumulation condition in steady state can be characterized as: $R + \mu = r + \delta$, where $R \equiv \partial\Pi(p,k)/\partial k$ is the return of capital and $\mu \geq 0$ is the Lagrangian multiplier for the non-negativity constraint $k \geq 0$. If the capital stock in steady state is positive ($k > 0$), we have $R = r + \delta$. If the return of capital is smaller than the cost of capital ($R < r + \delta$), we have $k = 0$ as firms have no incentive to hold any capital stock. Because of the scalability assumption (Assumption 3.9), the return of capital can be expressed as a function of the sales per unit of capital $\tilde{y}$:

$$R(\tilde{y}) = \pi(\tilde{y}; p, 1, w) = p\tilde{y} - C(\tilde{y}, 1; w).$$

which is strictly increasing for all $\tilde{y} \in [0, S(p, 1; w)]$.

First, consider the case where the price charged by firms is larger than the LRAC. If goods are produced and sold at full capacity, the return of capital will be larger than the cost of capital:

$$R\left(k^* (1, w)^{-1}\right) > r + \delta.$$

By Lemma 3.14, $k^* (1, w)^{-1} \leq S(p, 1; w)$. In addition, if the sales per unit of capital is zero ($\tilde{y} = 0$) the return of capital will be zero, $R(0) = 0 < r + \delta$, which is smaller than the cost of capital. Since $R(\tilde{y})$ is strictly increasing for all $\tilde{y} \in [0, S(p,1;w)]$. By the intermediate value theorem, there exists a unique sales per unit of capital $\tilde{y}^* \in \left(0, k^*(1,w)^{-1}\right)$ that solves the steady state equilibrium condition: $R(\tilde{y}^*) = r + \delta$. The capital stock in steady state must be positive. If $k = 0$, sales is supply constrained: $S(p,0;w) = 0 < \mathcal{D}$. Thus, if a firm chooses to invest a unit of capital, the sales gained is $S(p,1;w)$. The return of capital in this case is larger than the cost of capital: $R(S(p,1;w)) > r + \delta$, contradicting to the equilibrium condition. Since $k > 0$, the firm-level output is given by $y = \tilde{y}^* k < k^*(1,w)^{-1} k = \bar{y}$, which is smaller than the firm-level capacity, i.e., capacity is underutilized.

Second, consider the case where the price charged by firms is equal to the LRAC. If goods are produced and sold at full capacity, the return of capital will be equal to the cost of capital:

$$R\left(k^*(1,w,r)^{-1}\right) = r + \delta.$$

By Lemma 3.14, we have $k^*(1,w)^{-1} \leq S(p,1;w)$. Hence, $k^*(1,w)^{-1}$ is the unique sales per unit of capital $\tilde{y} \in (0, S(p,1;w)]$ that solves the steady state equilibrium condition. Therefore, the firm-level output is given by $y = k^*(1,w)^{-1} k = \bar{y}$, which means that capacity is fully utilized.

Finally, consider the case where the price charged by firms is smaller than the LRAC. If goods are produced and sold at full capacity the return of capital will be smaller than the cost of capital:

$$R\left(k^*(1,w)^{-1}\right) < r + \delta.$$

Since $k^*(1,w)^{-1}$ is already the point at which the SRAC is minimized, we have $p < \mathbf{LRAC}(w) \leq \mathbf{SRAC}(S(p,1;w), 1; w)$, which implies that

$$pS(p,1;w) - C(S(p,1;w), 1; w) < r + \delta.$$

Thus, for all $\tilde{y} \in [0, S(p,1;w)]$, the return of capital is always smaller than the cost of capital: $R(\tilde{y}) = p\tilde{y} - C(\tilde{y}, 1; w) < r + \delta$. By the equilibrium condition, we have that $k = 0$ and $y = 0$ as firms have no incentive to hold any capacity. □

3.2.3 Chronic Excess Capacity

However, capacity underutilization at the firm-level does not necessarily mean that capacity is in excess from an aggregate perspective.

Definition 3.15. (*Excess capacity and aggregate capacity*) Capacity is said to be in excess if and only if aggregate capacity is underutilized. Aggregate capacity is the aggregate output level at which the aggregate SRAC curve is tangent to the aggregate LRAC curve. The aggregate SRAC is given by

$$\mathcal{SRAC}\left(\mathcal{Y},k;w\right)\equiv\frac{\mathcal{C}\left(\mathcal{Y},k;w\right)+\left(r+\delta\right)k}{\mathcal{Y}},\tag{3.12}$$

where $\mathcal{Y}>0$ is the aggregate demand and $\mathcal{C}\left(\mathcal{Y},k;w\right)$ is the aggregate variable cost incurred to satisfy the aggregate demand. The aggregate LRAC is the minimum of the aggregate SRAC when aggregate capital can be adjusted:

$$\mathcal{LRAC}\left(\mathcal{Y};w\right)\equiv\min_{k\in\{x|(\mathcal{Y},x)\in\mathbf{dom}(\mathcal{C})\}}\frac{\mathcal{C}\left(\mathcal{Y},k;w\right)+\left(r+\delta\right)k}{\mathcal{Y}}.\tag{3.13}$$

In general, the aggregate variable cost function $\mathcal{C}$ is different from the variable cost function at the firm-level C as the former takes into account all the variable costs incurred to satisfy aggregate demand. For example, suppose that $\Phi\left(\mathbf{D},\mathbf{k}\right)$ units of labor are required to demand the goods produced by firms as it could be costly for buyers to search for supply, to process information, and to make orders, where $\mathbf{D}\equiv\left(D_j\right)_{j\in[0,1]}$ collects all the demand faced by firms. This purchasing cost shall be included in the aggregate variable cost.

Furthermore, aggregate output is a different object from the output at the firm-level. Let $\mathcal{A}$ denote the aggregation function, which aggregates the goods produced by firms into a final good:

$$\mathcal{Y}=\mathcal{A}\left(\mathbf{y},\mathbf{k}\right),\tag{3.14}$$

where $\mathbf{y}\equiv\left(y_j\right)_{j\in[0,1]}$ collects all the goods produced by firms. It is possible that different capital levels can cause the aggregate output to be different even though the firm-level output is the same.

Suppose that the aggregation function $\mathcal{A}$ and the purchasing cost function Φ satisfy the following properties:

1. (*Non-negativity*) $\mathcal{A}$ and Φ are non-negative.

2. (*Concavity/convexity*) The aggregation function $\mathcal{A}$ is concave and the purchasing cost function Φ is convex.

3. (*Monotone*) For any $\alpha_1>\alpha_2\geq0$, we have $\mathcal{A}\left(\alpha_1\mathbf{y},\mathbf{k}\right)>\mathcal{A}\left(\alpha_2\mathbf{y},\mathbf{k}\right)$ if $\mathcal{A}\left(\mathbf{y},\mathbf{k}\right)>0$, and $\Phi\left(\alpha_1\mathbf{D},\mathbf{k}\right)\geq\Phi\left(\alpha_2\mathbf{D},\mathbf{k}\right)$.

4. (*Symmetric*) For any $\mathbf{y}' = \left(y_{\mathcal{I}(j)}\right)_{j\in[0,1]}$, $\mathbf{k}' = \left(k_{\mathcal{I}(j)}\right)_{j\in[0,1]}$, and $\mathbf{D}' = \left(D_{\mathcal{I}(j)}\right)_{j\in[0,1]}$, where $\mathcal{I} : [0,1] \to [0,1]$ is a one-to-one mapping, we have $\mathcal{A}(\mathbf{y}', \mathbf{k}') = \mathcal{A}(\mathbf{y}, \mathbf{k})$ and $\Phi(\mathbf{D}', \mathbf{k}') = \Phi(\mathbf{D}, \mathbf{k})$.

The aggregate variable cost function can now be expressed as a result of the following cost minimization problem:

$$\mathcal{C}(\mathcal{Y}, k; w) = \min_{\mathbf{y},\mathbf{k},\mathbf{D}\geq 0} \int_0^1 C(y_j, k_j; w)\, dj + w\Phi(\mathbf{D}, \mathbf{k}), \tag{3.15}$$

subject to the aggregation function (3.14), the demand constraint $y_j \leq D_j$, the production limit $y_j \leq \chi^{-1} k_j$ if there is one, and the capital stock available $\int_0^1 k_j dj \leq k$. Since the problem is convex and symmetric, a symmetric allocation can minimize the aggregate variable cost. Hence, $y_j = D_j = y$ and $k_j = k$ for all $j \in [0,1]$. The optimal output at the firm-level is given by the inverse of the aggregation function (3.14): $y = \mathrm{a}(\mathcal{Y}, k)$. Therefore, the aggregate variable cost function is

$$\mathcal{C}(\mathcal{Y}, k; w) = C(\mathrm{a}(\mathcal{Y}, k), k; w) + w\Phi(\mathrm{a}(\mathcal{Y}, k), k), \tag{3.16}$$

where if there is a production limit, the aggregate demand $\mathcal{Y}$ must be feasible such that $\mathrm{a}(\mathcal{Y}, k) \leq \chi^{-1} k$.

Equation (3.16) shows that at both the individual firm level and the aggregate level, capital is valuable because it can expand the production potential of firms. First, it can save costs by increasing the productivity of labor. Second, it can expand the production limit if there is one. At the aggregate level, if $\mathrm{a}(\mathcal{Y}, k)$ or $\Phi(D, k)$ is strictly decreasing in terms of capital, capital has additional reasons to be valuable: it can increase the efficiency of the goods aggregation process or reducing the amount of labor services that has to be incurred by buyers to purchase the goods from firms. Because firms are small, these benefits will not be internalized. Thus, capacity underutilization at the firm-level *may* be justified at the aggregate level.

There are two types of externalities that work in opposite directions. On one hand, for a given amount of total demand requested, when a firm expands its capacity while other firms do not, the firm steals profitable demand from others. Because of this negative externality due to capacity competition, firms tends to hold too much capacity as we have shown in Proposition 3.13. On the other hand, capacity expansion could also increase the efficiency of the goods aggregation process or reducing the amount of labor services that has to be incurred by buyers to purchase the goods from firms. Because of this positive externality, firms tends to hold too little capacity. It is, therefore, possible to have a knife-edge condition, like the Hosios condition in the search and matching literature, with which these two types of externalities happen to cancel out.

By assuming away the positive externality in a symmetric equilibrium, the following assumption ensures that the capacity competition externality dominates.

Assumption 3.16. *(No positive externality) If all firms expand (or contract) their capital stock by the same proportion, both the aggregate output and the purchasing cost are not affected:* $\forall\alpha > 0$, $\mathcal{A}(\mathbf{y}, \alpha\mathbf{k}) = \mathcal{A}(\mathbf{y}, \mathbf{k})$ *and* $\Phi(\mathbf{D}, \alpha\mathbf{k}) = \Phi(\mathbf{D}, \mathbf{k})$.

With Assumption 3.16 added, the extra capacity installed in aggregate due to the capacity competition externality has no extra benefit and thus is a waste of resources from an aggregate perspective. The following proposition shows that because of this mechanism, the economy must exhibit chronic excess capacity.

Proposition 3.17. *(Chronic excess capacity) With Assumptions 3.9, 3.12, and 3.16, aggregate capacity must be underutilized in the steady state of a symmetric equilibrium.*

Proof. The aggregate variable cost function can now be written as

$$\mathcal{C}(\mathcal{Y}, k; w) = C(\mathrm{a}(\mathcal{Y}, 1), k; w) + w\Phi(\mathrm{a}(\mathcal{Y}, 1), 1).$$

Hence, for each $\mathcal{Y} > 0$, $k^*(1, w)\,\mathrm{a}(\mathcal{Y}, 1)$ is the unique aggregate capital level that minimizes the aggregate SRAC. Let $\bar{\mathcal{Y}}$ be the aggregate capacity. By definition, $\left(\bar{\mathcal{Y}}, k\right)$ is the point at which the aggregate SRAC is tangent to the aggregate LRAC; thus, $k = k^*(1, w)\,\mathrm{a}\left(\bar{\mathcal{Y}}, 1\right)$. By Proposition 3.13 and Lemma 3.14, capacity is underutilized at the firm-level and firms are willing to produce at full capacity: $y < \bar{y} \leq S(p, k; w)$. Thus, we have $y = \mathrm{a}(\mathcal{Y}, 1) < \bar{y} = k^*(1, w)^{-1} k = \mathrm{a}\left(\bar{\mathcal{Y}}, 1\right)$. Since the output at the firm-level $y = \mathrm{a}(\mathcal{Y}, 1)$ is strictly increasing in aggregate output, we have $\mathcal{Y} < \bar{\mathcal{Y}}$, which says that aggregate capacity is underutilized. □

3.3 Example I: Rational Inattention

There are several different ways to micro-found the goods market structure described in section 3.2. For example, in Chapter 1, I developed a basic capacity underutilization (CU) model where buyers are inattentive to prices and tend to search for capacity in an undirected way because it is costly for buyers to think about prices. In this section, I review a generalized version of the basic CU model and make some further discussions on the buyer's problem.

3.3.1 Model Setup

Production Technology

There is a continuum of firms indexed by $j \in [0,1]$. The goods produced by firms are perfect substitutes for each other and can be used either as consumption or as investment. The production technology is assumed to be Leontief:

$$y_{j,t} = \min\left\{\frac{l_{j,t}}{\alpha_v}, Ak_{j,t}\right\}, \tag{3.17}$$

where $y_{j,t}$ is the output of firm j, $k_{j,t}$ is the capital stock, and $\alpha_v > 0$ is the required labor per output. Each firm is a price taker in the labor market. Hence, the variable cost function of firm j at time $t \geq 0$ is given by

$$C\left(y_{j,t}, k_{j,t}; w_t\right) = w_t \alpha_v y_{j,t}, \tag{3.18}$$

where $w_t > 0$ is the real wage rate; and the domain of the variable cost function is given by

$$\mathbf{dom}\,(C) = \left\{(y,k) \mid y \geq 0, k \geq A^{-1}y\right\}, \tag{3.19}$$

where $A > 0$. Thus, there is a production limit such that $y_{j,t} \leq Ak_{j,t}$. The short run average total cost (SRAC) is

$$\mathbf{SRAC}\,(y,k;w_t) \equiv \frac{w_t \alpha_v y + (r+\delta)\,k}{y}, \tag{3.20}$$

where $y \in (0, Ak]$, $r > 0$ is the real interest rate in steady state, and $\delta > 0$ is the depreciation rate; and the long run average total cost (LRAC) is

$$\begin{aligned}\mathbf{LRAC}\,(w_t) &\equiv \min_{k \in \{x \mid (y,x) \in \mathbf{dom}(C)\}} \mathbf{SRAC}\,(y,k;w_t) \\ &= w_t \alpha_v + (r+\delta)\,A^{-1}.\end{aligned} \tag{3.21}$$

One can verify that the capacity of the firm is equal to its production limit: $\bar{y}_{j,t} = Ak_{j,t}$.

Demand System

The purchasing process takes two steps. First, a household or a firm decides how many goods $\mathcal{Y}_t$ shall be consumed or invested based on the aggregate price level $\mathcal{P}_t > 0$. Second, the household or the firm sends out her buyer to purchase these goods for her. All buyers are identical and the payoff for the buyer who purchases goods from firm j is a strictly decreasing function of the real price charged by the firm:

$$v_t(j) = \mathrm{v}\left(\frac{P_{j,t}}{\mathcal{P}_t}\right), \tag{3.22}$$

where $P_{j,t}$ is the nominal price charged by firm j and $\mathrm{v}' < 0$. The payoff function is kept general. It will become clear later that the main properties of the model do not depend on the particular functional form of the payoff function as long as the payoff is strictly decreasing in the real price.

In principle, buyers want to purchase only the cheapest goods. However, following the rational inattention literature, I assume that it is costly for the buyers to process price information in order to direct their actions towards the desired. Without exerting any information processing effort, the matching between the demand from buyers and the capacity supplied by firms is assumed to be completely random:

$$n_t^*(j) = \frac{\bar{y}_{j,t}}{\bar{y}_t}, \tag{3.23}$$

where $n_t^*(j)$ is the likelihood of purchasing from firm j and $\bar{y}_t \equiv \int_0^1 \bar{y}_{j,t} dj$ is the total capacity. n_t^* describes the most inattentive behavior of a buyer and is called the *default* purchasing behavior.

To deviate from the default, a buyer has to process some price information and incurs some information processing cost. Let n_t be the purchasing behavior eventually obtained by a buyer. $n_t(j)$ gives the likelihood that the buyer purchases from firm j. The amount of information that needs to be processed is given by the relative entropy of the behavior chosen by the buyer n_t with respect to the default n_t^*:

$$D_{KL}(n_t || n_t^*) \equiv \int_0^1 n_t(j) \ln\left(\frac{n_t(j)}{n_t^*(j)}\right) dj. \tag{3.24}$$

Intuitively, the more different n_t is from n_t^*, the more information is needed to be processed.

A representative buyer's problem can now be characterized as:

$$\max_{n_t \geq 0} \int_0^1 \mathrm{v}\left(\frac{P_{j,t}}{\mathcal{P}_t}\right) n_t(j)\, dj - \Lambda D_{KL}\left(n_t \| n_t^*\right), \tag{3.25}$$

subject to $\int_0^1 n_t(j)\, dj = 1$, where $\Lambda > 0$ is the unit cost of processing information.

As demonstrated in section 1.2.2 of Chapter 1, the behavior of the buyers yields the following demand curve for firm j:

$$y_{j,t} \leq D_{j,t} = n_t(j)\, \mathcal{D}_t = \frac{k_{j,t} e^{\frac{1}{\Lambda}\mathrm{v}\left(\frac{P_{j,t}}{\mathcal{P}_t}\right)}}{\int_0^1 k_{j,t} e^{\frac{1}{\Lambda}\mathrm{v}\left(\frac{P_{j,t}}{\mathcal{P}_t}\right)} dj} \mathcal{D}_t. \tag{3.26}$$

where $D_{j,t}$ is the demand for the goods produced by firm j and $\mathcal{D}_t$ is the total demand requested by buyers. Suppose that all demand requested can be fulfilled. The conjecture will be verified later. The total demand requested is simply the total amount of goods that households and firms want to buy:

$$\mathcal{D}_t = \int_0^1 y_{j,t} dj = \mathcal{Y}_t, \tag{3.27}$$

where the second equality is true because all goods are perfect substitutes.

Profit Maximization

Each firm would choose both its price and its output to maximize its gross profit:

$$\Pi_t\left(k_{j,t}\right) = \max_{P_{j,t} \geq 0, y_{j,t} \geq 0} \left(\frac{P_{j,t}}{\mathcal{P}_t} - w_t \alpha_v\right) y_{j,t}, \tag{3.28}$$

subject to the demand constraint (3.26) and the capacity constraint $y_{j,t} \leq A k_{j,t}$. Let $\mu_{j,t} \geq 0$ be the Lagrangian multiplier for the capacity constraint. The firm's pricing condition is given by:

$$\frac{P_{j,t}}{\mathcal{P}_t} = \frac{\varepsilon_{j,t}}{\varepsilon_{j,t} - 1}\left(w_t \alpha_v + \mu_{j,t}\right), \tag{3.29}$$

where $\varepsilon_{j,t}$ is the price elasticity of demand (3.26):

$$\varepsilon_{j,t} = -\frac{1}{\Lambda} \mathrm{v}'\left(\frac{P_{j,t}}{\mathcal{P}_t}\right) \frac{P_{j,t}}{\mathcal{P}_t}, \tag{3.30}$$

which is inversely related to the unit cost of processing information. I assume that $\Lambda < -\mathrm{v}'(1)$, a condition that is sufficient to prevent firms from charging an infinitely high price in equilibrium.

If the capacity constraint is not binding, the price charged by firm j is given by a markup on the marginal cost. If the capacity constraint is binding, the price charged by the firm will be raised up until demand is equal to capacity. Hence, in equilibrium, prices will be adjusted such that all demand requested is fulfilled, verifying our previous conjecture.

The aggregate price $\mathcal{P}_t$ must be such that solves $\mathcal{P}_t \mathcal{Y}_t = \int_0^1 P_{j,t} y_{j,t} dj$ to ensure that the money earned by firms is the same as the money spent on aggregate goods. According to the aggregation function (3.27), the aggregate price $\mathcal{P}_t$ is simply the average price of the goods purchased by buyers: $P_t = \int_0^1 P_{j,t} n_t (j)\, dj$.

Capital Accumulation

Given the gross profit function (3.28), each firm aims to maximize its firm value, which is the present value of the firm's dividend flows:

$$\max_{\{i_{j,t}, k_{j,t+1} \geq 0\}_{t=0}^{\infty}} \mathbb{E}_0 \left(\sum_{t=0}^{\infty} \left(\prod_{\tau=1}^{t} Q_\tau \right) \left(\Pi_t (k_{j,t}) - i_{j,t} \right) \right), \tag{3.31}$$

subject to the capital law of motion $k_{j,t+1} = (1-\delta) k_{j,t} + i_{j,t}$, where Q_{t+1} is the stochastic discount factor that discounts the value at time $t+1$ to time t and $i_{j,t}$ is the investment.

Households

There is a unit mass of identical households. The representative household maximizes her expected lifetime utility

$$\max_{\{c_t, l_t\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{c_t^{1-\gamma} - 1}{1-\gamma} - \bar{\omega} l_t \right), \tag{3.32}$$

subject to the resource constraint

$$c_t = w_t l_t + d_t, \tag{3.33}$$

where $\beta \in [0,1)$ is the subjective discount factor, $\gamma^{-1} > 0$ gives the elasticity of intertemporal substitution, $\bar{\omega} > 0$ is the marginal dis-utility of labor, c_t is the amount of consumption goods purchased, l_t is the labor supply, and $d_t = \int_0^1 \Pi_t (k_{j,t}) - i_{j,t} dj$ is the amount of dividends received from firms. The labor market is perfectly competitive. Hence, the household takes the real wage rate w_t as exogenous. The household's problem implies that the stochastic discount factor is given by $Q_{t+1} = \beta \left({}^{c_{t+1}}/_{c_t} \right)^{-\gamma}$.

3.3.2 Symmetric Equilibrium

In a symmetric equilibrium, if a variable is of the form $x_{j,t}$, we have $x_{i,t} = x_{j,t}$ for all $i, j \in [0, 1]$. Hence, we can omit the subscripts that index a variable to a particular firm after obtaining the equilibrium condition. The symmetric equilibrium is determined by a stable process of five variables (c_t, w_t, k_{t+1}, μ_t, and y_t) that satisfies the household's consumption leisure trade-off:

$$c_t^{-\gamma} w_t = \bar{\omega}, \tag{3.34}$$

the firm's pricing condition

$$1 = \frac{\varepsilon}{\varepsilon - 1} \left(w_t \alpha_v + \mu_t \right), \tag{3.35}$$

where $\varepsilon = -\Lambda^{-1} \mathrm{v}'(1) > 1$, the firm's investment condition

$$1 = \mathbb{E}_t \left(\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \left(\frac{1}{\varepsilon} A u_{t+1} + A \mu_{t+1} + 1 - \delta \right) \right), \tag{3.36}$$

where $u_t \equiv y_t / Ak_t$ measures of the capacity utilization rate, the complementary slackness condition for the capacity constraint

$$\mu_t \left(Ak_t - y_t \right) = 0, \tag{3.37}$$

where $\mu_t \geq 0$ and $Ak_t - y_t \geq 0$, and the aggregate resource constraint combined with the law of motion for capital

$$c_t + k_{t+1} - (1 - \delta) k_t = y_t. \tag{3.38}$$

Note that the equilibrium conditions of the model are exactly the same as those of the basic CU model developed in Chapter 1. The shape of the buyer's payoff function only affects the demand elasticity (see equation (3.30)). Since the equilibrium is symmetric, the effect is homogeneous and can be offset by a proper choice of the unit cost of processing information Λ: for each $\mathrm{v}'(1) < 0$, there exists a $\Lambda > 0$ such that the demand elasticity $\varepsilon = -\Lambda^{-1} \mathrm{v}'(1)$ remains unchanged.

3.3.3 Chronic Excess Capacity

Because it is costly to think about prices, buyers tend to purchase in an undirected way. Hence, firms with a larger capacity are more likely to be visited by buyers. As the demand function (3.26) is linear in terms of capital, the purchasing behavior of the buyers ensures the scalability of the demand, i.e., Assumption 3.9.

The rational inattention assumption also allows firms to enjoy some monopolistic power. The price elasticity of demand is inversely related to the unit cost of processing information. As long as buyers are not fully attentive to prices $\Lambda > 0$, each firm will be able to charge some positive net markup. If the unit cost of processing information is high enough such that

$$1 > \frac{\Lambda}{-\mathrm{v}'(1)} > (r+\delta) A^{-1}, \tag{3.39}$$

one can verify that the real price charged by firms in equilibrium will be larger than the LRAC of the firms:

$$\mathbf{LRAC}(w_t) = w_t \alpha_v + (r+\delta) A^{-1} \leq 1 + \frac{\Lambda}{\mathrm{v}'(1)} + (r+\delta) A^{-1} < 1, \tag{3.40}$$

where the first inequality comes from the firm's pricing condition (3.35) and the fact that the value of the Lagrangian is non-negative: $\mu_t \geq 0$. Hence, the profitability assumption, i.e., Assumption 3.12, is satisfied.

Scalability and profitability together imply that firms are involved in a capacity competition that has a negative externality. According to Proposition 3.13, the capacity of firms will be underutilized in steady state. Indeed, according to the firm's investment condition (3.36), the capacity utilization rate of a representative firm in steady state is given by

$$u = \frac{y}{\bar{y}} = \frac{(r+\delta) A^{-1}}{1 - w\alpha_v} < 1, \tag{3.41}$$

which is smaller than one because of the profitability condition (3.40).

Since the aggregation function (3.27) is simply the summation of all the goods produced and no purchasing cost is paid by households or firms, the invested capital has no positive externality and Assumption 3.16 is satisfied. According to Proposition 3.17, the economy exhibits chronic excess capacity.

3.3.4 Further Discussions on the Buyer's Problem

Although buyers purchase on behalf of households and firms, the buyers have their own payoffs. This assumption is convenient as it greatly simplifies the aggregation problem by allowing buyers to be homogeneous across potentially different households and firms.

In this sub-section, I show that a homogeneous purchasing problem can be directly obtained as a sub-problem of households and firms. If the information processing cost is paid by households and firms in terms of goods or services and is proportional to the amount of goods purchased, households and firms would share the same purchasing problem that is to minimize the unit cost of acquiring goods. In this case, the payoff of purchasing a unit of goods from firm j decreases linearly in the real price charged by the firm. However, the main properties of the model remain unchanged.

Purchasing Problem

Suppose that households and firms have to pay an information processing cost when purchasing. Without paying any information processing cost, households and firms purchase in an undirected way as described in equation (3.23). This default behavior is summarized by the probability density function n_t^*.

To deviate from the default purchasing behavior, households and firms have to process some price information. Suppose that the amount of information processed is proportional to the amount of goods purchased. For each unit of goods purchased, the information processing cost is given by $w_t \phi D_{KL}\left(n_t || n_t^*\right)$, where n_t is the probability density function that describes the purchasing behavior obtained by a household or a firm, $D_{KL}\left(n_t || n_t^*\right)$ measures the relative entropy of n_t with respect to n_t^*, and $\phi > 0$ is the amount of labor required to process a unit of information. The expected cost per unit of goods acquired is given by

$$\varsigma_t \equiv \int_0^1 \frac{P_{j,t}}{\mathcal{P}_t} n_t\left(j\right) dj + w_t \phi D_{KL}\left(n_t || n_t^*\right), \tag{3.42}$$

where $n_t\left(j\right)$ gives the likelihood of purchasing from firm j. Hence, households and firms all aim to minimize the expected unit cost of acquiring goods, or to maximize the minus expected unit cost of acquiring goods $-\varsigma_t$:

$$\max_{n_t \geq 0} -\int_0^1 \frac{P_{j,t}}{\mathcal{P}_t} n_t\left(j\right) dj - w_t \phi D_{KL}\left(n_t || n_t^*\right), \tag{3.43}$$

subject to $\int_0^1 n_t\left(j\right) dj = 1$.

The above purchasing problem is of the same structure as the buyer's problem described previously in equation (3.25). Once the amount of goods that needs to be purchased has been determined, households and firms act as buyers who decide where the goods should be purchased from. The payoff function of purchasing from firm j is now linearly decreasing in the real price of the goods purchased:

$$\mathrm{v}\left(\frac{P_{j,t}}{\mathcal{P}_t}\right) = -\frac{P_{j,t}}{\mathcal{P}_t}, \tag{3.44}$$

and the unit cost of processing information is related to the labor required to process a unit of information and the real wage rate: $\Lambda_t = w_t\phi$.

The rest of the model setup is exactly the same as before. The solution to the purchasing problem (3.43) yields the following demand system:

$$\forall j: \; y_{j,t} \leq D_{j,t} = n_t(j)\,\mathcal{Y}_t = \frac{k_{j,t}e^{-\frac{1}{w_t\phi}\frac{P_{j,t}}{\mathcal{P}_t}}}{\int_0^1 k_{j,t}e^{-\frac{1}{w_t\phi}\frac{P_{j,t}}{\mathcal{P}_t}}\,dj}\mathcal{Y}_t. \tag{3.45}$$

Equation (3.45) shows that the price elasticity of demand is given by $\varepsilon_{j,t} = (w_t\phi)^{-1}\,{}^{P_{j,t}}/_{\mathcal{P}_t} > 1$, which must be larger than one in equilibrium. If prices are too high, the real wage rate will be depressed and the demand elasticity increases, preventing firms from charging an infinitely high price. In equilibrium, all firms would charge the same price. The firm's pricing condition in a symmetric equilibrium is now given by:

$$1 = \frac{1}{1 - w_t\phi}\left(w_t\alpha_v + \mu_t\right), \tag{3.46}$$

where $\mu_t \geq 0$ is the Lagrangian multiplier for the capacity constraint.

Main Properties

As households and firms tend to purchase in an undirected way, the demand for the goods produced by firm j is proportional to the firm's capital stock (see, equation (3.45)). Hence, the scalability assumption, i.e., Assumption 3.9, is satisfied.

If the labor required to process a unit of information is high enough such that

$$\phi > \frac{\alpha_v\,(r+\delta)\,A^{-1}}{1-(r+\delta)\,A^{-1}}, \tag{3.47}$$

one can verify that the real price charged by firms in equilibrium will be larger than the LRAC of the firms:

$$\mathbf{LRAC}\,(w_t) = w_t\alpha_v + (r+\delta)\,A^{-1} \leq \frac{\alpha_v}{\alpha_v+\phi} + (r+\delta)\,A^{-1} < 1, \tag{3.48}$$

where the first inequality comes from the firm's pricing condition (3.46) and the fact that the value of the Lagrangian is non-negative: $\mu_t \geq 0$. Hence, the profitability assumption, i.e., Assumption 3.12, is satisfied.

The aggregation function (3.27) is simply the summation of all the goods produced. In addition, the amount labor required for purchasing due to the information processing cost is given by

$$\Phi(\mathbf{D}, \mathbf{k}) = \phi D_{KL}(n_t \| n_t^*) \mathcal{D}_t = \phi \int_0^1 D_{j,t} \ln \left(\frac{D_{j,t}/\mathcal{D}_t}{k_{j,t}/k_t} \right) dj, \tag{3.49}$$

which is homogeneous of degree zero in terms of capital levels. Hence, Assumption 3.16, i.e., the no positive externality, is satisfied.

According to Propositions 3.13 and 3.17, the capacity at the firm-level is underutilized in steady state and the economy exhibits chronic excess capacity.

When capacity is in excess, the real wage rate and the unit cost of processing information in equilibrium are both constant: $w_t = (\phi + \alpha_v)^{-1}$ and $\Lambda = w_t \phi = \phi (\phi + \alpha_v)^{-1}$. Furthermore, since all firms charge the same price in equilibrium, the actual information processing cost incurred is zero. Hence, in equilibrium, no extra hours are worked. With a strictly decreasing payoff function (3.44) for buyers, the dynamic properties of the model in the case where capacity is in excess are exactly the same as before.

3.4 Example II: Random Matching and Nash Bargaining

This section presents an alternative way to micro-found the goods market structure described in section 3.2. I introduce the standard labor market search and matching framework into the goods market. Prices are determined via a static Nash bargaining process as in Michaillat and Saez (2015). With the same production technology and the same household's problem as in section 3.3, I show that the dynamic properties of the model locally around the steady state where capacity is in excess are the same as those of the basic CU model reviewed in section 3.3.

3.4.1 Model Setup

Production Technology

There is a continuum of firms indexed by $j \in [0, 1]$. The goods produced by firms are identical and are perfect substitutes for each other. The goods can be used either as consumption or as investment. The production technology is the same as described in section 3.3.

Demand System

Households and firms are price takers in the labor market. They need to hire some labor for purchasing goods. For each unit of demand requested, $\phi > 0$ units of labor are required. As in the standard labor market search and matching literature (e.g., Mortensen and Pissarides 1994), I assume that when households and firms purchase, they search for capacity in an undirected way. The random matching process between the total demand requested $\mathcal{D}_t$ and the total capacity $\bar{y}_t = \int_0^1 \bar{y}_{j,t} dj$ supplied yields the following demand system:

$$\forall j: \; D_{j,t} = \frac{\bar{y}_{j,t}}{\bar{y}_t} \mathcal{D}_t, \tag{3.50}$$

where $D_{j,t}$ is the demand requested for the goods produced by firm j. The matching technology is assumed to be Leontief:

$$y_{j,t} \leq \min \left\{ D_{j,t}, \bar{y}_{j,t} \right\}, \tag{3.51}$$

which means that the goods that could be sold by firm j is constrained by the demand requested and the capacity supplied.

Suppose that all demand matched can be fulfilled in equilibrium. This conjecture will be verified later. The probability that a unit of demand is successfully fulfilled is simply the ratio of the demand matched to the demand requested:

$$q(x_t) = \min \left\{ 1, \frac{\bar{y}_{j,t}}{D_{j,t}} \right\} = \min \left\{ 1, \frac{\bar{y}_t}{\mathcal{D}_t} \right\} = \min \left\{ 1, x_t^{-1} \right\}, \tag{3.52}$$

where $x_t \equiv \mathcal{D}_t / \bar{y}_t$ measures the market tightness and $q(x_t)$ denotes the demand fulfilling probability as a decreasing function of market tightness. Note that the demand fulfilling probability is not firm-specific because of the random matching process. Each buyer, who could be a household or a firm, is infinitely small and will take $q(x_t)$ as exogenous.

Because all goods are perfect substitutes, the amount of aggregate goods produced $\mathcal{Y}_t$ is given by

$$\mathcal{Y}_t = \int_0^1 y_{j,t} dj; \tag{3.53}$$

and the total demand that has to be requested in order to satisfy the aggregate demand

$\mathcal{Y}_t$ is given by:

$$\mathcal{D}_t = \frac{\mathcal{Y}_t}{q\left(x_t\right)}. \tag{3.54}$$

Nash Bargaining

Let $P_{j,t}$ be the price charged by firm j. $P_{j,t}$ will be determined via a static Nash bargaining process between the buyer and the firm after matching has taken place.

For each successful matching, a unit of goods can be purchased to satisfy a unit of aggregate demand. Since all goods are perfect substitues, the value of this matching for the buyer is equal to the buyer's outside option, which is the cost that would have to be incurred if the buyer chooses to search again. Hence, the marginal surplus to the buyer who purchases from firm j is given by

$$\mathscr{B}_t\left(P_{j,t}\right) = 1 + w_t\phi\frac{1}{q\left(x_t\right)} - \frac{P_{j,t}}{\mathcal{P}_t}, \tag{3.55}$$

where $q\left(x_t\right)^{-1}$ is the demand that has to be requested in order to purchase successfully a unit of goods and $\mathcal{P}_t$ is the aggregate price. The marginal surplus to the firm who sells a unit of goods to the buyer is simply given by the marginal gross profit:

$$\mathscr{F}_t\left(P_{j,t}\right) = \frac{P_{j,t}}{\mathcal{P}_t} - w_t\alpha_v. \tag{3.56}$$

The total surplus of the matching must be positive. If not, since purchasing is costly ($\phi > 0$), no purchasing will take place and no labor will be hired, violating the labor market clearing condition. The Nash bargaining process between the buyer and the firm is to find the price that splits the total surplus according the bargaining power of each party:

$$\max_{P_{j,t}\geq 0} \mathscr{B}_t\left(P_{j,t}\right)^{1-b} \mathscr{F}_t\left(P_{j,t}\right)^{b}, \tag{3.57}$$

where $b \in [0,1]$ is a parameter that captures the bargaining power of firms. Hence, the bargained price is give by

$$\frac{P_{j,t}}{\mathcal{P}_t} = b\left(1 + w_t\phi\frac{1}{q\left(x_t\right)} - w_t\alpha_v\right) + w_t\alpha_v, \tag{3.58}$$

which shows that if the firm has some positive bargaining power ($b > 0$), the price charged by the firm will be above the marginal cost $w_t\alpha_v$, creating effectively a positive net markup.

Since it is always profitable to sell, the firms will be happy to fulfill all the matched demand, verifying our previous conjecture. Thus, the goods produced by the firm is given by

$$y_{j,t} = \min\{D_{j,t}, \bar{y}_{j,t}\} = u(x_t)\,\bar{y}_{j,t}, \tag{3.59}$$

where $u(x_t)$ is the capacity utilization probability or the capacity utilization rate:

$$u(x_t) = \frac{y_{j,t}}{\bar{y}_{j,t}} = \min\left\{\frac{\mathcal{D}_t}{\bar{y}_t}, 1\right\} = \min\{x_t, 1\}; \tag{3.60}$$

and the gross profit function of firm j can be written as:

$$\Pi_t(k_{j,t}) = b\left(1 + w_t\phi\frac{1}{q(x_t)} - w_t\alpha_v\right)u(x_t)\,Ak_{j,t}. \tag{3.61}$$

Note that because of the random matching process, the capacity utilization rate, like the demand fulfilling probability, is not firm-specific.

The aggregate price $\mathcal{P}_t$ must solve $\mathcal{P}_t\mathcal{Y}_t = \int_0^1 P_{j,t}y_{j,t}dj$ to ensure that the money earned by firms is the same as the money spent on aggregate goods. According to the aggregation condition (3.27), the aggregate price $\mathcal{P}_t$ is simply the average price of the goods purchased: $\mathcal{P}_t = \int_0^1 P_{j,t}\left(y_{j,t}/\int_0^1 y_{j,t}dj\right)dj$.

Capital Accumulation

The capital accumulation decision of firm j is to choose the firm's investment $i_{j,t}$ and the capital stock $k_{j,t+1}$ to maximize the firm's value:

$$\max_{\{i_{j,t},k_{j,t+1}\geq 0\}_{t=0}^{\infty}} \mathbb{E}_0\left(\sum_{t=0}^{\infty}\left(\prod_{\tau=1}^{t}Q_\tau\right)\left(\Pi_t(k_{j,t}) - \left(1 + w_t\phi\frac{1}{q(x_t)}\right)i_{j,t}\right)\right), \tag{3.62}$$

subject to the capital law of motion $k_{j,t+1} = (1-\delta)\,k_{j,t} + i_{j,t}$, where Q_{t+1} is the stochastic discount factor that discounts the value at time $t+1$ to time t.

Households and Aggregate Demand

There is a continuum of identical households. The representative household's problem, which determines the household's consumption c_t and the labor supply l_t, is to maximize the life-time utility as described by equations (3.32) in section 3.3, subject to the resource constraint

$$\left(1 + w_t\phi\frac{1}{q(x_t)}\right)c_t = w_tl_t + d_t. \tag{3.63}$$

The household's problem implies that the stochastic discount factor is given by $Q_{t+1} = \beta \left(c_{t+1}/c_t \right)^{-\gamma} \left(\Phi_t/\Phi_{t+1} \right)$, where $\Phi_t \equiv \left(1 + w_t \phi \frac{1}{q(x_t)} \right)$.

3.4.2 Symmetric Equilibrium

The symmetric equilibrium is determined by a stable process of five variables (c_t, w_t, k_{t+1}, x_t, and y_t) that satisfies the household's consumption leisure trade-off

$$c_t^{-\gamma} \frac{w_t}{1 + w_t \phi \frac{1}{q(x_t)}} = \bar{\omega}, \tag{3.64}$$

the Nash bargaining condition

$$1 = b \left(1 + w_t \phi \frac{1}{q\left(x_t\right)} - w_t \alpha_v \right) + w_t \alpha_v. \tag{3.65}$$

the firm's investment condition

$$1 = \mathbb{E}_t \left(\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \left(\frac{1 - w_{t+1} \alpha_v}{1 + w_{t+1} \phi \frac{1}{q(x_{t+1})}} A u\left(x_{t+1}\right) + (1 - \delta) \right) \right), \tag{3.66}$$

the aggregate matching technology

$$y_t = u\left(x_t\right) A k_t, \tag{3.67}$$

and the aggregate resource constraint (3.38) as obtained in section 3.3.

3.4.3 Chronic Excess Capacity

As shown by the linearity of the demand function (3.50), the random matching assumption ensures the scalability of the demand, i.e., Assumption 3.9.

If the bargaining power of the firms is sufficiently strong such that

$$1 \geq b > \frac{(r + \delta) A^{-1} \left(\alpha_v + \phi\right)}{\phi + (r + \delta) A^{-1} \alpha_v}, \tag{3.68}$$

one can verify that the real price charged by firms in equilibrium will be larger than the

LRAC of the firms in equilibrium:

$$\begin{aligned}\mathbf{LRAC}\left(w_t, x_t\right) &= w_t\alpha_v + \left(r+\delta\right)A^{-1}\left(1 + w_t\phi\frac{1}{q\left(x_t\right)}\right)\\ &\leq \frac{\left(r+\delta\right)A^{-1}\phi\left(1-b\right) + \alpha_v\left(1-b\right)}{\phi b + \alpha_v\left(1-b\right)} + \left(r+\delta\right)A^{-1} < 1,\end{aligned} \tag{3.69}$$

where the first inequality comes from the Nash bargaining condition (3.65) and the fact that the demand fulfilling probability is no larger than one: $q\left(x_t\right) \leq 1$. Hence, the profitability assumption, i.e., Assumption 3.12, is satisfied.

Scalability and profitability together imply that firms are involved in a capacity competition that has a negative externality. According to Proposition 3.13, the capacity of firms will be underutilized in steady state. Indeed, according to the firm's investment condition (3.66), the capacity utilization rate of a representative firm in steady state is given by

$$u\left(x\right) = \frac{y}{\bar{y}} = \frac{\left(r+\delta\right)A^{-1}}{1 - w\alpha_v}\left(1 + w\phi\frac{1}{q\left(x\right)}\right) < 1, \tag{3.70}$$

which is smaller than one because of the profitability condition (3.68).

Since the aggregation function (3.53) is simply the summation of all the goods produced and the cost of purchasing is independent of the capital stock, Assumption 3.16, i.e., the no positive externality assumption, is satisfied. According to Proposition 3.17, the economy exhibits chronic excess capacity.

We can verify this conclusion. Note that the aggregate variable cost function is given by

$$\mathcal{C}\left(\mathcal{Y}_t, k_t; w_t\right) = \min_{\left\{y_{j,t}, k_{j,t}, D_{j,t} \geq 0\right\}_{j\in[0,1]}} w_t\left(\int_0^1 \alpha_v y_{j,t}dj + \int_0^1 \phi D_{j,t}dj\right), \tag{3.71}$$

subject to the aggregation function (3.53), the matching constraint (3.51), and the capital stock available $\int_0^1 k_{j,t}dj \leq k_t$. The problem is symmetric and convex. Hence, the aggregate variable cost is given by $\mathcal{C}\left(\mathcal{Y}_t, k_t; w_t\right) = w_t\left(\alpha_v + \phi\right)\mathcal{Y}_t$, for all $\mathcal{Y}_t \leq Ak_t$. Thus, the aggregate capacity is given by the production limit, which is the same as the total capacity at the firm-level: $\bar{\mathcal{Y}}_t = Ak_t = \bar{y}_t$. The aggregate output is also the same as the total output at the firm-level because of the aggregation function (3.53). Therefore, as Proposition 3.17 predicts, aggregate capacity is underutilized in steady state ($\mathcal{Y} < \bar{\mathcal{Y}}$) as long as the capacity at the firm-level is underutilized in steady state ($y < \bar{y}$).

Finally, locally around the steady state where capacity is in excess, all demand is matched $q\left(x_t\right) \equiv 1$ and the real wage rate w_t is a constant. Hence, the dynamic properties of the model developed here are exactly the same as those of the basic CU model reviewed in section 3.3.[2]

3.5 Example III: Investment as Variety Expansion

This section shows that the Dixit-Stiglitz (DS) monopolistic competition framework can be adapted to provide a micro-foundation for the goods market structure described in section 3.2. Particularly, I present a model in which investment in capital not only expands production potential but also expands variety. With a standard DS setup, if a firm increases the variety of the goods sold, the firm increases its competitive advantage, which allows the firm to steal demand from others. Hence, Assumption 3.9 holds. Since goods are imperfect substitutes, firms can charge a positive net markup. If the markups charged by firms are high enough, Assumption 3.12 is satisfied. Finally, the invested capital has no positive externality in equilibrium as the value of variety can be internalized by firms. I show that the economy exhibits chronic excess capacity as Proposition 3.17 predicts. With the same household's problem as in section 3.3, I show that the dynamic properties of the model are exactly the same as those of the basic CU model reviewed in section 3.3.

3.5.1 Model Setup

Production Technology

There is a continuum of firms indexed by $j \in [0,1]$. Each firm j operates with a continuum of production lines. Let $N_{j,t}$ be the amount of production lines owned by firm j. Each production line is indexed by a pair (j,v), where $v \in [0, N_{j,t}]$. Each production line requires a certain amount of capital stock $\bar{k} > 0$ to be installed and produces a certain type of differentiated goods. The amount of production lines owned by firm j is proportional to the capital stock owned by the firm: $k_{j,t} = \bar{k} N_{j,t}$.

Let $y_{j,v,t}$ be the amount of goods produced by the production line v of firm j. The aggregate goods produced are a CES aggregation of all the differentiated goods produced:

$$\mathcal{Y}_t = N_t^{-\varphi} \left(\int_0^1 \int_0^{N_{j,t}} y_{j,v,t}^{\frac{\varepsilon-1}{\varepsilon}} dv dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3.72}$$

[2] The total hours worked in the random matching model is larger than that in the basic CU model because of the cost of purchasing. However, the percentage changes in total hours will be the same in both models as sizes are normalized when calculating percentages.

where $\mathcal{Y}_t$ is the amount of aggregate goods, $N_t = \int_0^1 N_{j,t} dj$ is the total mass of production lines, $\varepsilon > 1$ is the elasticity of substitution among differentiated goods, and $\varphi \geq 0$ is a parameter that captures the potential crowd-out effect due to variety expansion.

Firms do not sell the goods produced by their production lines directly. Instead, they sell a combination of the goods produced by their production lines:

$$y_{j,t} = N_{j,t}^{-\varphi} \left(\int_0^{N_{j,t}} y_{j,v,t}^{\frac{\varepsilon-1}{\varepsilon}} dv \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3.73}$$

where $y_{j,t}$ is the effective output of firm j. Hence, the CES aggregation (3.72) can be rewritten as

$$\mathcal{Y}_t = \left(\int_0^1 \left(\frac{N_{j,t}}{N_t} \right)^{\varphi \frac{\varepsilon-1}{\varepsilon}} y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{3.74}$$

which shows that firms who own more production lines than others are more attractive to customers. Note that the CES aggregation at the firm-level (3.73) resembles the economy-wide CES aggregation (3.72), and allows firms to take into account both the value of variety expansion and the potential crowd-out effect due to variety expansion. However, as equation (3.74) shows, firms with more production lines still enjoy a competitive advantage. This advantage is not justified in a symmetric equilibrium: if all firms expand their variety by the same proportion without changing their effective output, the aggregate output will be the same.

To produce goods, each production line operates with a Leontief production technology, uses the capital stock $\bar{k}$ installed, and hires some variable labor $l_{j,v,t}$:

$$y_{j,v,t} = \min \left\{ \frac{l_{j,v,t}}{\alpha_v}, A\bar{k} \right\}. \tag{3.75}$$

The variable cost of the firm is given by the following cost minimization problem:

$$C\left(y_{j,t}, k_{j,t}; w_t\right) = \min_{\{y_{j,v,t} \geq 0\}_{v \in [0, N_{j,t}]}} w_t \alpha_v \int_0^{N_{j,t}} y_{j,v,t} dv, \tag{3.76}$$

subject to the CES aggregation at the firm-level (3.73), the production limit $y_{j,v,t} \leq A\bar{k}$, and the amount of production lines determined by the capital stock available $\bar{k}N_{j,t} = k_{j,t}$. Since the problem is strictly convex and symmetric, the optimal solution must be symmetric. Hence, the variable cost function is

$$C\left(y_{j,t}, k_{j,t}; w_t\right) = w_t \alpha_v y_{j,t} \left(\frac{k_{j,t}}{\bar{k}} \right)^{1+\varphi-\frac{\varepsilon}{\varepsilon-1}}, \tag{3.77}$$

for $y_{j,t} \leq A\bar{k}\left(k_{j,t}/\bar{k}\right)^{\frac{\varepsilon}{\varepsilon-1}-\varphi}$. In the rest of this section, I assume that $\varphi = (\varepsilon-1)^{-1}$; thus, the effective production technology is of constant returns to scale and the variable cost function is homogeneous of degree one in terms of effective output and capital. In this case, one can verify that the capacity of the firm is simply the production limit, which is given by $\bar{y}_{j,t} = Ak_{j,t}$.

Demand System

Suppose that all demand requested can be fulfilled. The conjecture will be verified later. If $\mathcal{Y}_t$ units of aggregate goods are demanded, the demand requested $D_{j,t}$ for the goods produced by firm j is generated by minimizing the total cost of the goods purchased:

$$\min_{\{D_{j,t}\geq 0\}_{j\in[0,1]}} \int_0^1 P_{j,t} D_{j,t} dj, \tag{3.78}$$

subject to the aggregation function (3.74), where $P_{j,t} > 0$ is the price charged by firm j. The solution to the above minimization problem gives:

$$\forall j:\ y_{j,t} \leq D_{j,t} = \frac{N_{j,t}}{N_t}\left(\frac{P_{j,t}}{P_{I,t}}\right)^{-\varepsilon}\mathcal{Y}_t = \frac{k_{j,t}}{k_t}\left(\frac{P_{j,t}}{P_{I,t}}\right)^{-\varepsilon}\mathcal{Y}_t, \tag{3.79}$$

where $P_{I,t} \equiv \left(\int_0^1 \frac{N_{j,t}}{N_t} P_{j,t}^{1-\varepsilon} dj\right)^{\frac{1}{1-\varepsilon}}$ is a price index and $k_t = \int_0^1 k_{j,t} dj$ is the total capital stock. Note that the demand faced by firm j is linear in terms of the capital stock of the firm because capital investment also works as a variety expansion that makes the firm more attractive to customers.

Profit Maximization

Each firm chooses both its price and its effective output to maximize its gross profit, which is given by the revenue minus the variable cost:

$$\Pi_t\left(k_{j,t}\right) = \max_{P_{j,t}\geq 0, y_{j,t}\geq 0}\left(\frac{P_{j,t}}{\mathcal{P}_t} - w_t\alpha_v\right) y_{j,t}, \tag{3.80}$$

subject to the demand constraint (3.79) and the capacity constraint $y_{j,t} \leq Ak_{j,t}$. The firm's pricing condition is given by:

$$\frac{P_{j,t}}{\mathcal{P}_t} = \frac{\varepsilon}{\varepsilon-1}\left(w_t\alpha_v + \mu_{j,t}\right), \tag{3.81}$$

where $\mu_{j,t} \geq 0$ be the Lagrangian multiplier for the capacity constraint. If the capacity constraint is not binding, the price charged by firm j is given by a constant markup on the marginal cost. If the capacity constraint is binding, the price charged by the firm will be raised up until demand is equal to capacity. Hence, in equilibrium, prices will be adjusted such that all demand requested is fulfilled, verifying our previous conjecture.

The aggregate price $\mathcal{P}_t$ must solve $\mathcal{P}_t \mathcal{Y}_t = \int_0^1 P_{j,t} y_{j,t} dj$ to ensure that the money earned by firms is the same as the money spent on aggregate goods. According to the demand system (3.79) and the aggregation function (3.74), we have that the aggregate price is equal to the price index: $\mathcal{P}_t = P_{I,t}$.

Capital Accumulation

The capital accumulation decision is based on the same firm-value maximization problem (3.31) as described in section 3.3, a problem that determines the firm's investment $i_{j,t}$ and the capital stock $k_{j,t+1}$ for the next period.

Households and Aggregate Demand

There is a continuum of identical households. The representative household's problem, which determines the household's consumption c_t and labor supply l_t, is the same as described by equations (3.32) and (3.33) in section 3.3.

3.5.2 Symmetric Equilibrium

In a symmetric equilibrium, for all $j \in [0,1]$, we have that $P_{j,t} = P_{I,t} = \mathcal{P}_t$ and $y_{j,t} = y_t = \mathcal{Y}_t$.

The symmetric equilibrium of the model is characterized by a stable stochastic process of five variables (c_t, w_t, k_{t+1}, μ_t, and y_t) that satisfies the same household's consumption leisure trade-off (3.34), the firm's pricing condition (3.35), the firm's investment condition (3.36), the complementary slackness condition for the capacity constraint (3.37), and the aggregate resource constraint (3.38) as obtained in section 3.3 for the basic CU model.

3.5.3 Chronic Excess Capacity

Note that capital investment in this model allows firms to install more production lines and expand the variety of the goods supplied in the market. Firms who can supply a higher variety can attract more demand. As shown by the linearity of the demand function (3.79), the scalability of the demand, i.e., Assumption 3.9, is satisfied.

The imperfect substitution between goods allows firms to have some monopolistic power. If the elasticity of substitution between the differentiated goods is small enough such that

$$1 < \varepsilon < A\left(r+\delta\right)^{-1}, \tag{3.82}$$

one can verify that the real price charged by firms in equilibrium will be larger than the LRAC of the firms at full capacity:

$$\mathbf{LRAC}\left(w_t\right) = w_t\alpha_v + \left(r+\delta\right)A^{-1} \leq 1 - \frac{1}{\varepsilon} + \left(r+\delta\right)A^{-1} < 1, \tag{3.83}$$

where the first inequality comes from the firm's pricing condition (3.35) and the fact that the value of the Lagrangian is non-negative: $\mu_t \geq 0$. Hence, the profitability assumption, i.e., Assumption 3.12, is satisfied.

Scalability and profitability together imply that firms are involved in a capacity competition that has a negative externality. According to Proposition 3.13, the capacity of firms will be underutilized in steady state. Indeed, according to the firm's investment condition (3.36), the capacity utilization rate of a representative firm in steady state is given by

$$u = \frac{y}{\bar{y}} = \frac{\left(r+\delta\right)A^{-1}}{1 - w\alpha_v} < 1, \tag{3.84}$$

which is smaller than one because of the profitability condition (3.83).

Since the aggregation function (3.74) that aggregates the effective output of firms is homogeneous of degree zero in terms of capital levels and there is no purchasing cost, the no positive externality assumption, i.e., Assumption 3.16, is satisfied. According to Proposition 3.17, the economy exhibits chronic excess capacity.

We can verify this conclusion. The aggregate variable cost function is given by

$$\mathcal{C}\left(\mathcal{Y}_t, k_t; w_t\right) = \min_{\left\{y_{j,v,t}\geq 0, N_{j,t}\geq 0\right\}_{j\in[0,1], v\in\left[0,N_{j,t}\right]}} w_t\alpha_v \int_0^1 \int_0^{N_{j,t}} y_{j,v,t}\,dvdj, \tag{3.85}$$

subject to the CES aggregation (3.72), the production limit $y_{j,v,t} \leq A\bar{k}$, and the capital stock available $\bar{k} \int_0^1 N_{j,t} dj \leq k_t$. The problem is symmetric and strictly convex. Hence, the aggregate variable cost is given by $\mathcal{C}(\mathcal{Y}_t, k_t; w_t) = w_t \alpha_v \mathcal{Y}_t$, for all $\mathcal{Y}_t \leq Ak_t$; and the aggregate capacity is given by the production limit, which is the same as the total capacity at the firm-level: $\bar{\mathcal{Y}}_t = Ak_t = \bar{y}_t$. According to the aggregation function (3.74), the aggregate output is also the same as the total output at the firm-level because the equilibrium is symmetric. Therefore, as Proposition 3.17 predicts, aggregate capacity is underutilized in steady state ($\mathcal{Y} < \bar{\mathcal{Y}}$) as long as the capacity at the firm-level is underutilized in steady state ($y < \bar{y}$).

Finally, note that the variety expansion model developed here and the basic CU model reviewed in section 3.3 share the same equilibrium conditions. Hence, the dynamic properties of the two models are exactly the same.

3.6 Conclusions

This chapter discusses what kind of goods market structure features the capacity competition externality that can cause chronic excess capacity. Three key assumptions are identified. First, each firm is able to expand its market share proportionally by expanding its relative capacity. Hence, in addition to the usual price competition, there is a capacity competition among firms. Second, each firm is able to charge a sufficiently high price to make a positive net profit. Hence, there is a negative capacity competition externality: if one firm expands its capacity while other firms do not, it steals profitable demand from others. Third, the invested capital has no positive externality. Hence, any capacity underutilized at the firm-level is in excess.

The scalability, the profitability, and the no positive externality assumptions together describe a goods market structure in which the capacity competition externality dominates, and ensure that the economy exhibits chronic excess capacity. These three assumptions capture the key mechanism that drives the results of the capacity underutilization models studied in this thesis.

There are several different ways to micro-found this goods market structure. Firms can make a positive net profit because of a monopolistic power or because of a bargaining power. The market share of a firm is proportional to the firm's capital stock because buyers are inattentive to prices and thus search for capacity in an undirected way or because capital investment expands variety and thus increases the competitive advantage of the firm. Different models correspond to different interpretations but the main properties of the models are the same. The assumptions identified in this chapter are potentially testable. To what extent these three assumptions are true is an important and interesting question for future research.

Bibliography

Andrle, M., Bruha, J., Solmaz, S. 2017. On the sources of business cycles: implications for DSGE models. ECB Working Paper.

Angeletos, G.-M., Collard, F., Dellas, H. 2018. Business cycle anatomy.Technical report, National Bureau of Economic Research.

Ascari, G., Phaneuf, L., Sims, E. 2019. Can New Keynesian Models Survive the Barro-King Curse?.

Bai, Y., Rios-Rull, J.-V., Storesletten, K. 2012. Demand shocks as productivity shocks. Federal Reserve Board of Minneapolis.

Barro, R. J., King, R. G. 1984. Time-separable preferences and intertemporal-substitution models of business cycles. The Quarterly Journal of Economics, 99, 817–839.

Basu, S., Kimball, M. S. 1997. Cyclical productivity with unobserved input variation.Technical report, National Bureau of Economic Research.

Best, M. C., Cloyne, J. S., Ilzetzki, E., Kleven, H. J. 2020. Estimating the elasticity of intertemporal substitution using mortgage notches. The Review of Economic Studies, 87, 656–690.

Bils, M., Cho, J.-O. 1994. Cyclical factor utilization. Journal of Monetary Economics, 33, 319–354.

Bils, M. J. 1985. Real wages over the business cycle: evidence from panel data. Journal of Political economy, 93, 666–689.

Boehm, C. E., Flaaen, A., Pandalai-Nayar, N. 2017. Are supply curves convex? implications for state-dependent responses to shocks. Working paper.

Brandolini, A. 1995. In search of a stylised fact: do real wages exhibit a consistent pattern of cyclical variability? Journal of Economic Surveys, 9, 103–163.

Buch, C. M., Döpke, J., Strotmann, H. 2009. Does export openness increase firm-level output volatility? World Economy, 32, 531–551.

Bulow, J., Geanakoplos, J., Klemperer, P. 1985. Holding idle capacity to deter entry. The Economic Journal, 95, 178–182.

Burnside, C., Eichenbaum, M., Rebelo, S. 1993. Labor hoarding and the business cycle. Journal of Political Economy, 101, 245–273.

Christiano, L. J., Eichenbaum, M., Evans, C. L. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of political Economy, 113, 1–45.

Comin, D., Philippon, T. 2005. The rise in firm-level volatility: Causes and consequences. NBER macroeconomics annual, 20, 167–201.

Cooley, T. F., Hansen, G. D., Prescott, E. C. 1995. Equilibrium business cycles with idle resources and variable capacity utilization. Economic Theory, 6, 35–49.

Corrado, C., Mattey, J. 1997. Capacity utilization. The Journal of Economic Perspectives, 11, 151–167.

De Veirman, E., Levin, A. 2018. Cyclical changes in firm volatility. Journal of Money, Credit and Banking, 50, 317–349.

Deng, S., Yano, C. A. 2006. Joint production and pricing decisions with setup costs and capacity constraints. Management Science, 52, 741–756.

Dixit, A. 1980. The role of investment in entry-deterrence. The economic journal, 90, 95–106.

Eiteman, W. J. 1947. Factors determining the location of the least cost point. The American Economic Review, 37, 910–918.

Eiteman, W. J., Guthrie, G. E. 1952. The shape of the average cost curve. The American economic review, 42, 832–838.

Fagnart, J.-F., Licandro, O., Portier, F. 1999. Firm heterogeneity, capacity utilization, and the business cycle. Review of Economic Dynamics, 2, 433–455.

Fagnart, J.-F., Licandro, O., Sneessens, H. R. 1997. Capacity utilization and market power. Journal of Economic Dynamics and Control, 22, 123–140.

Florian, M., Klein, M. 1971. Deterministic production planning with concave costs and capacity constraints. Management Science, 18, 12–20.

Gilchrist, S., Williams, J. C. 2000. Putty-clay and investment: a business cycle analysis. Journal of Political Economy, 108, 928–960.

Greenwood, J., Hercowitz, Z., Huffman, G. W. 1988. Investment, capacity utilization, and the real business cycle. The American Economic Review, 402–417.

Greenwood, J., Hercowitz, Z., Krusell, P. 2000. The role of investment-specific technological change in the business cycle. European Economic Review, 44, 91–115.

Gruber, J. 2013. A tax-based estimate of the elasticity of intertemporal substitution. The Quarterly Journal of Finance, 3, p. 1350001.

Gupta, S., Starr, M. 2014. Production and operations management systems. CRC Press.

Hall, R. E. 1986. Chronic excess capacity in US industry.

Hall, R. E. 1988. Intertemporal substitution in consumption. Journal of political economy, 96, 339–357.

Hansen, G. D. 1985. Indivisible labor and the business cycle. Journal of monetary Economics, 16, 309–327.

Havránek, T. 2015. Measuring intertemporal substitution: The importance of method choices and selective reporting. Journal of the European Economic Association, 13, 1180–1204.

Hayashi, F. 1982. Tobin's marginal q and average q: A neoclassical interpretation. Econometrica: Journal of the Econometric Society, 213–224.

Jaimovich, N., Rebelo, S. 2009. Can news about the future drive the business cycle? American Economic Review, 99, 1097–1118.

Johansen, L. 1959. Substitution versus fixed production coefficients in the theory of economic growth: a synthesis. Econometrica: Journal of the Econometric Society, 157–176.

Johnson, P. A. 1994. Capital utilization and investment when capital depreciates in use: some implications and tests. Journal of Macroeconomics, 16, 243–259.

Justiniano, A., Primiceri, G. E., Tambalotti, A. 2010. Investment shocks and business cycles. Journal of Monetary Economics, 57, 132–145.

Kalish, S. 1983. Monopolist pricing with dynamic demand and production cost. Marketing Science, 2, 135–159.

Kelly, B., Lustig, H., Van Nieuwerburgh, S. 2013. Firm volatility in granular networks.Technical report, National Bureau of Economic Research.

Khan, H., Tsoukalas, J. 2011. Investment shocks and the comovement problem. Journal of Economic Dynamics and Control, 35, 115–130.

King, R. G., Rebelo, S. T. 1999. Resuscitating real business cycles. Handbook of macroeconomics, 1, 927–1007.

Köberl, E. M., Lein, S. M. 2011. The nircu and the phillips curve: an approach based on micro data. Canadian Journal of Economics/Revue canadienne d'économique, 44, 673–694.

Kumar, S. A., Suresh, N. 2006. Production and operations management. New Age International.

Kydland, F. E., Prescott, E. C. 1982. Time to build and aggregate fluctuations. Econometrica: Journal of the Econometric Society, 1345–1370.

Kydland, F. E., Prescott, E. C. 1988. The workweek of capital and its cyclical implications. Journal of Monetary Economics, 21, 343–360.

Matějka, F., McKay, A. 2015. Rational inattention to discrete choices: A new foundation for the multinomial logit model. American Economic Review, 105, 272–98.

Mattsson, L.-G., Weibull, J. W. 2002. Probabilistic choice and procedurally bounded rationality. Games and Economic Behavior, 41, 61–78.

Michaillat, P., Saez, E. 2015. Aggregate demand, idle time, and unemployment. The Quarterly Journal of Economics, 130, 507–569.

Moen, E. R. 1997. Competitive search equilibrium. Journal of political Economy, 105, 385–411.

Morrison, C. J. 1985. Primal and dual capacity utilization: an application to productivity measurement in the us automobile industry. Journal of Business & Economic Statistics, 3, 312–324.

Mortensen, D. T., Pissarides, C. A. 1994. Job creation and job destruction in the theory of unemployment. The review of economic studies, 61, 397–415.

Nekarda, C. J., Ramey, V. A. 2013. The cyclical behavior of the price-cost markup. National Bureau of Economic Research.

Nelson, R. A. 1989. On the measurement of capacity utilization. The Journal of Industrial Economics, 273–286.

Rogerson, R. 1988. Indivisible labor, lotteries and equilibrium. Journal of monetary Economics, 21, 3–16.

Smets, F., Wouters, R. 2007. Shocks and frictions in us business cycles: A bayesian dsge approach. American economic review, 97, 586–606.

Solon, G., Barsky, R., Parker, J. A. 1994. Measuring the cyclicality of real wages: how important is composition bias? The quarterly journal of economics, 109, 1–25.

Spence, A. M. 1977. Entry, capacity, investment and oligopolistic pricing. The Bell Journal of Economics, 534–544.

Stevenson, W. J. 2002. Operations management. McGraw-Hill/Irwin Boston.

Walters, A. A. 1963. Production and cost functions: An econometric survey. Econometrica: Journal of the Econometric Society, 1–66.

Woodford, M. 2005. Firm-specific capital and the new-keynesian phillips curve.Technical report, National Bureau of Economic Research.

Chapter 4

Appendix

4.1 Calibrations of the Standard RBC Models

I do not use the observed average capacity utilization rate as a calibration target because capacity is fully utilized in the steady state of the standard RBC models. The other calibration targets are the same as in the basic CU model (see section 1.3.5). For the CD-RBC model, the capital productivity A is not identified because for each value of $A > 0$, there exists a value of α_v such that the total factor productivity (TFP) $\alpha_v^{-\alpha} A^{1-\alpha}$ is the same. To pin down the value of A, I choose A such that y/Ak is normalized to 1. Table 4.1 summarizes the calibrated parameter values of the Leontief-RBC model and the CD-RBC model.

Table 4.1: Parameters and Calibration Targets – RBC Models

Leontief-RBC		CD-RBC		Target
Parameter	Value	Parameter	Value	
δ	0.0210	δ	0.0210	Quarterly depreciation rate 0.021
γ	1.0000	γ	1.0000	Elasticity of inter-temporal substitution 1
ϕ	0.8300	ϕ	0.8300	Price in utils 1
$\bar{\omega}$	0.6200	$\bar{\omega}$	0.6200	Output 1
α_v	0.9120	α_v	0.9120	Labor underutilization rate 0.088
β	0.9747	β	0.9747	Investment to output ratio 0.17
A	0.1235	A	0.1235	Labor share of income 0.62
		α	0.6200	y/Ak normalized to 1

4.2 Description of the Data

The data on consumption, investment, government spending, exports, imports, output, hours, and capital is from the National Income and Product Accounts (NIPA) published by the U.S. Bureau of Economic Analysis (BEA). The data on employment, unemployment, labor force, wage rate, and labor share is from the U.S. Bureau of Labor Statistics (BLS). The data on capacity utilization rate is from the Federal Reserve Board (FRB).

The FRB publishes a capacity utilization rate for manufacturing and a capacity utilization rate for total industries. The latter covers manufacturing, mining, and electric and gas utilities. These two capacity utilization rates are very close to each other. The manufacturing capacity utilization rate, which has a longer history, is used to compare with the model fitted values in Figure 1.7.

The BEA data on capital is annually. To get the quarterly data, I use linear interpolation to impute the BEA annual data on capital. Let K_Y be the capital stock at the end of year Y. The capital stock at the end of year Y and quarter Q is taken to be

$$K_{Y,Q} = \exp\left(\ln K_{Y-1} + \frac{Q}{4}\left(\ln K_Y - \ln K_{Y-1}\right)\right).$$

The BEA data on hours worked by full-time and part-time employees is also annually. To get the quarterly data, I impute the BEA annual data on hours based on the information provided by the BLS. The BLS issues data on hours worked in business sectors at a quarterly frequency. Let H_Y be the hours worked by full-time and part-time employees in year Y and $H^B_{Y,Q}$ be the hours worked in business sectors in year Y and quarter Q. The hours worked by full-time and part-time employees in year Y and quarter Q is taken to be

$$H_{Y,Q} = H_Y \frac{H^B_{Y,Q}}{H^B_{Y,1} + H^B_{Y,2} + H^B_{Y,3} + H^B_{Y,4}}.$$

To convert the nominal variables into real ones, I divide the nominal variables by the GDP deflator obtained from the BEA.

All variables are detrended before they are used for estimations or to calculate business cycle statistics. I estimate a quadratic trend for the log of labor productivity measured by output to hours ratio. I also estimate a quadratic trend for the log of hours worked per capita.

Let $\overline{LP}_{Y,Q}$ be the trend of labor productivity, $\bar{h}_{Y,Q}$ be trend of hours worked per capita, and $\bar{L}_{Y,Q}$ be the labor force in year Y and quarter Q. The real wage rate is assumed to be of the same trend as the labor productivity. $\bar{h}_{Y,Q}\bar{L}_{Y,Q}$ captures the potential hours that could be worked in the economy, which is treated as the trend of hours worked by full-time and part-time employees. $\bar{h}_{Y,Q}\bar{L}_{Y,Q}\overline{LP}_{Y,Q}$ captures the potential output that could be produced in the economy, which is treated as the trend of consumption, investment, government spending, exports, imports, output, and capital.

Figure 4.1 shows the logarithms of the detrended consumption, investment, output, hours, capital, and real wage rate.

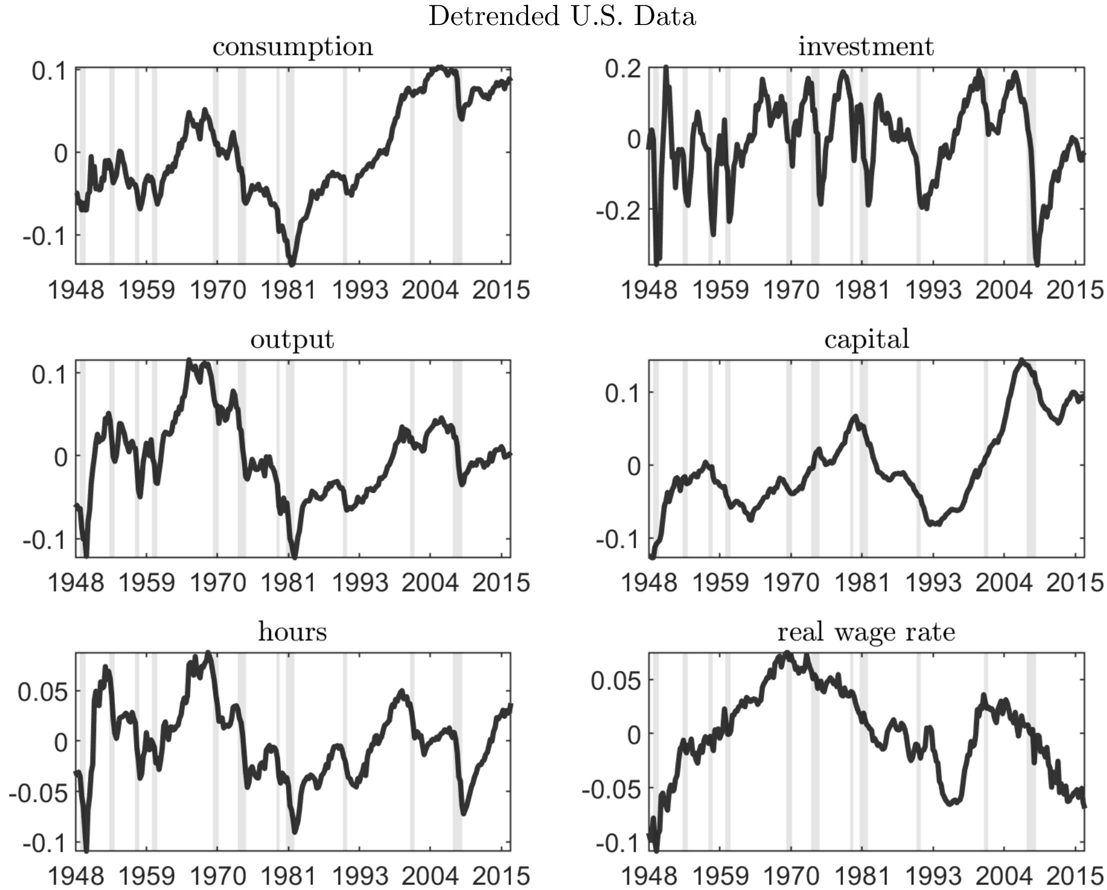

Figure 4.1: All variables are logarithms of the original series. The shaded areas are NBER dated recessions.

4.3 Estimation of the Full VU Model

I use Bayesian estimation techniques to estimate the full VU model. Thirteen calibration targets are chosen to be matched throughout this estimation procedure. The elasticity of inter-temporal substitution (EIS) is calibrated to 0.5. The exogenous expenditure to output ratio in steady state is matched to the average ratio of the exogenous expenditure to GDP from 1947 to 2016 in the BEA NIPA. The quarterly real interest rate in steady state is calibrated to 1%. The other ten calibration targets are the same as described in section 1.4.3, which calibrates the standard VU model.

Because of these targets, thirteen parameter values are *not* free to pick. Eleven parameters are fixed as their values follow directly from the thirteen calibration targets. Two parameters can be expressed as *functions* of other parameters and the thirteen calibration targets:

$$\alpha = \frac{\bar{\omega}\,(l - \alpha_g g)}{\bar{\omega}\,(l - \alpha_g g) + (\beta^{-1} - 1 + \delta)\,\frac{i}{\delta}}, \tag{4.1}$$

$$\varepsilon = \frac{c + i}{c + i - \left(\bar{\omega}\,(l - \alpha_g g) + (\beta^{-1} - 1 + \delta)\,\frac{i}{\delta}\right)}. \tag{4.2}$$

Table 4.2 summarizes these thirteen parameters and their mostly associated calibration targets.

The other parameters are estimated. The priors of parameters σ_c, σ_i, σ_g, σ_l, ρ_c, ρ_g, $\rho_{i,1}$, $\rho_{i,2}$, ρ_l, ρ_{ci}, ϕ_l, ϕ_k, ϕ_i, and α_g are assumed to be the same as in the full CU model. The parameter $\xi \equiv a''(\theta)\theta / a'(\theta)$, which captures the convexity of the capital utilization cost function in steady state, is assumed to follow a Normal distribution with a mean of 1 and a standard deviation of 0.5.

Table 4.3 summarizes the priors and shows the mode, the mean, and the 5th and 95th percentiles of the posterior distribution of the parameters obtained by the Metropolis-Hastings (MH) algorithm. The total number of MH draws is 100,000 and the acceptance ratio is about 20.5%.

Based on the posterior modes of the structural parameters and the thirteen calibration targets, the Cobb-Douglas (CD) labor share α is 0.68 and the elasticity of the demand curve ε is 55.57.

Table 4.2: Parameters Pinned down by the Calibration Targets – Full VU Model

Parameter	Value	Target
δ	0.0210	Quarterly depreciation rate 0.021
γ	2.0000	Elasticity of Inter-temporal substitution 0.5
β	0.9900	Quarterly real interest rate 0.01
g	0.2000	Exogenous expenditure to output ratio 0.2
α_v	0.9120	Labor underutilization rate 0.088
$\bar{\omega}$	0.6798	Output normalized 1
ϕ	0.6300	Price in utils normalized to 1
A	0.0988	y/Ak normalized to 1
θ	1.0000	θ normalized to 1
$a(\theta)$	0.0000	$a(\theta)$ normalized to 0
α	Eq. (4.1)	Labor share of income 0.62
ε	Eq. (4.2)	Investment to output ratio 0.17

Table 4.3: Bayesian Estimation – Full VU Model

	Prior Distribution			Posterior Distribution			
Parameter	Distribution	Mean	Std Dev	Mode	Mean	5th Percentile	95th Percentile
σ_c	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
σ_i	Invgamma	0.03	∞	0.03	0.02	0.02	0.03
σ_g	Invgamma	0.03	∞	0.02	0.02	0.02	0.03
σ_l	Invgamma	0.03	∞	0.01	0.01	0.01	0.01
ρ_c	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
$\rho_{i,1}$	Uniform	0.00	0.58	0.98	0.98	0.97	0.99
$\rho_{i,2}$	Normal	0.00	0.20	-0.23	-0.25	-0.35	-0.16
ρ_g	Uniform	0.00	0.58	0.99	0.99	0.98	1.00
ρ_l	Uniform	0.00	0.58	0.96	0.95	0.93	0.97
ρ_{ci}	Normal	0.00	0.20	-0.14	-0.15	-0.21	-0.08
ϕ_l	Uniform	0.00	0.58	-0.02	-0.03	-0.18	0.12
ϕ_k	Normal	2.00	1.00	1.88	2.09	0.85	3.29
ϕ_i	Normal	0.20	0.10	0.15	0.16	0.01	0.21
α_g	Normal	0.60	0.30	0.63	0.64	0.53	0.76
ξ	Normal	1.00	0.50	0.97	1.11	0.37	1.79

Note: Std Dev stands for standard deviation (of the priors). The sample period is from the first quarter of 1948 to the first quarter of 2017.

4.4 Additional Estimation Results

Figure 4.2 shows the consumption demand $z_{c,t}$, the investment demand $z_{i,t}$, the exogenous expenditure $z_{g,t}$, and the labor productivity $z_{l,t}$ series extracted from the data. Except for the labor productivity, the other exogenous stochastic processes extracted from the data are quite similar in both models.

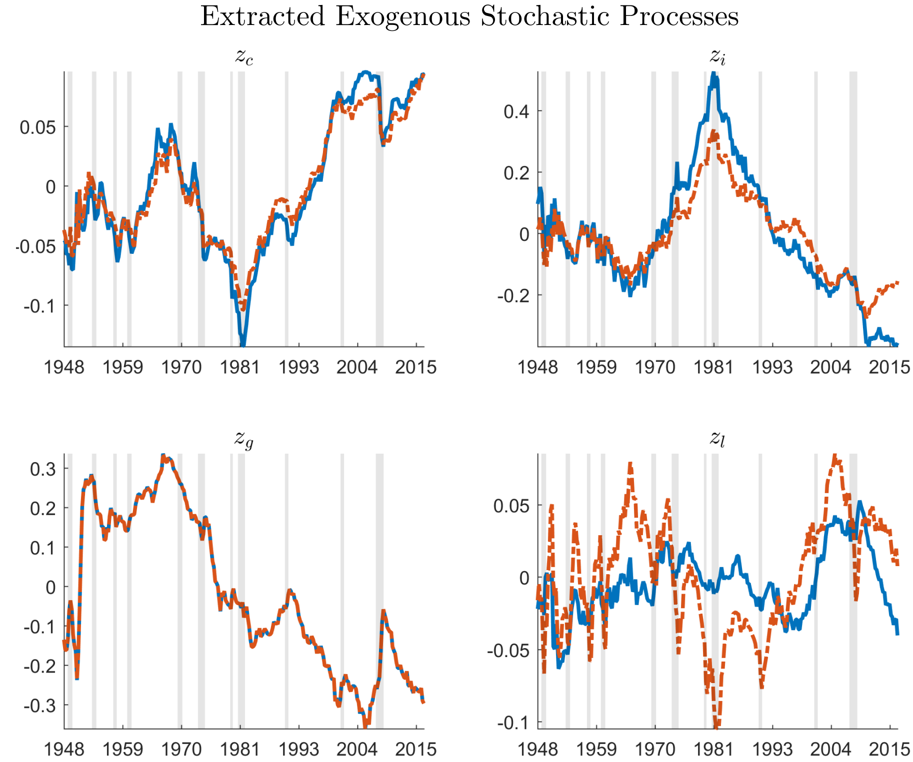

Figure 4.2: The consumption demand $z_{c,t}$, the investment demand $z_{i,t}$, the exogenous expenditure $z_{g,t}$, and the labor productivity $z_{l,t}$ extracted from the data. Shaded areas indicate the NBER dated recessions.

Figure 4.3 shows the model fitted values under investment demand shocks $e_{i,t}$. Since the estimated ρ_{ci} is negative, there is a negative correlation between the consumption and the investment. According to both models, a persistent increase in consumption shall cause a persistent increase in capital and a persistent increase in capital shall not cause a persistent

Table 4.4: Cyclicality of Consumption Under Investment Demand Shocks

	Full CU		Full VU	
	Correlation	Covariance	Correlation	Covariance
Consumption-Output	-0.004	-0.001	-0.147	-0.046
Consumption-Hours	-0.204	-0.064	-0.443	-0.132

Note: The covariance between consumption and output (or hours) is reported relative to the variance of output (or hours). The U.S. data is from the BEA and the BLS. All variables are Hodrick-Prescott (HP) filtered logarithms of the original series.

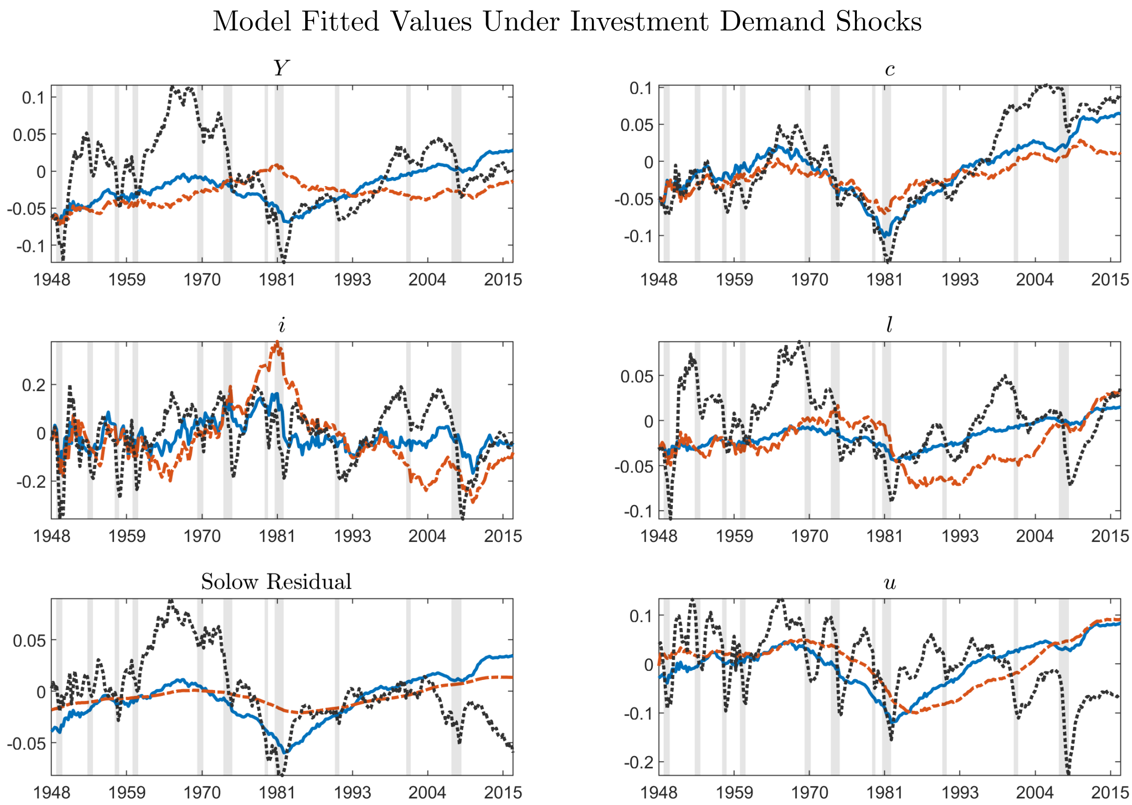

Figure 4.3: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

decline in consumption. However, from 1968 to 1993, the detrended consumption and the detrended capital move in opposite directions in the U.S. (see Figure 4.1). This divergence is unable to be captured by the internal mechanisms of both models but can be captured by the parameter ρ_{ci} that allows the consumption demand to be negatively affected by the investment demand (see Figure 4.4). The investment demand shocks, however, do not drive business cycles as consumption and investment move in opposite directions. Moreover, the movement of consumption under investment demand shocks is slightly countercyclical and does not exhibit large declines during the NBER dated recessions (see Table 4.4).

Figure 4.5 shows the model fitted values under exogenous expenditure shocks $e_{g,t}$. By construction, $e_{g,t}$ has no effect on consumption, investment, and capacity utilization rate.

Figure 4.6 shows the model fitted values under labor productivity shocks $e_{l,t}$. Labor productivity shocks play a small role in the full CU model but an important role in the full VU model to drive cyclical movements in investment, the Solow residual, and the capacity utilization rate. Labor productivity shocks, however, do not explain the fluctuations of hours very well and consumption is a bit too smooth under the labor productivity shocks in the full VU model.

Consumption-Capital Divergence Captured by Investment Demand Shocks

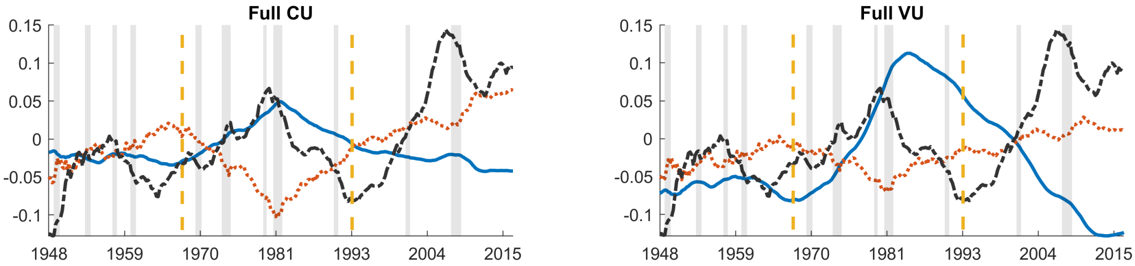

Figure 4.4: The solid lines are the model fitted capital. The dotted lines are the model fitted consumption. The dash-dotted lines are capital in the U.S. data. Shaded areas indicate the NBER dated recessions. The vertical dashed lines indicate the year 1968 and the year 1993. The data is from the BEA. All variables are logarithms of the original series.

Figure 4.7 shows the model predicted real wage rate, which is not targeted in estimation procedures. The overall trend of the real wage rate in both models is roughly consistent with that of the data. However, the real wage rate is a bit too volatile in the full VU model.

Model Fitted Values Under Exogenous Expenditure Shocks

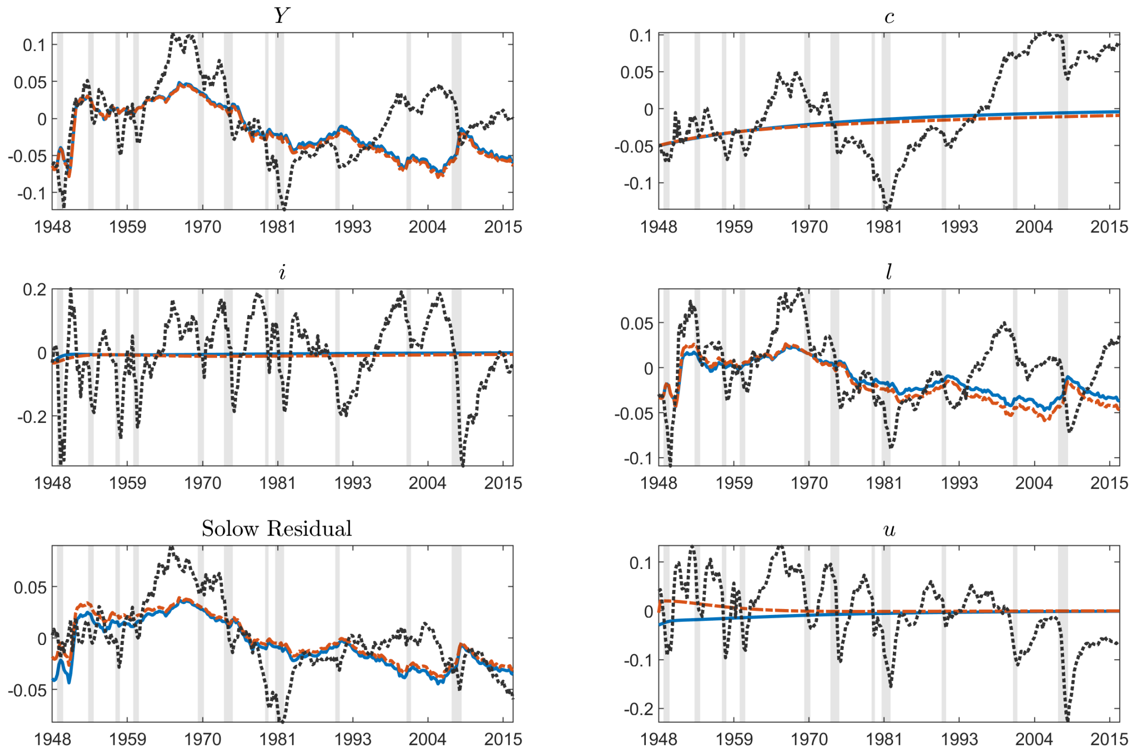

Figure 4.5: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Model Fitted Values Under Labor Productivity Shocks

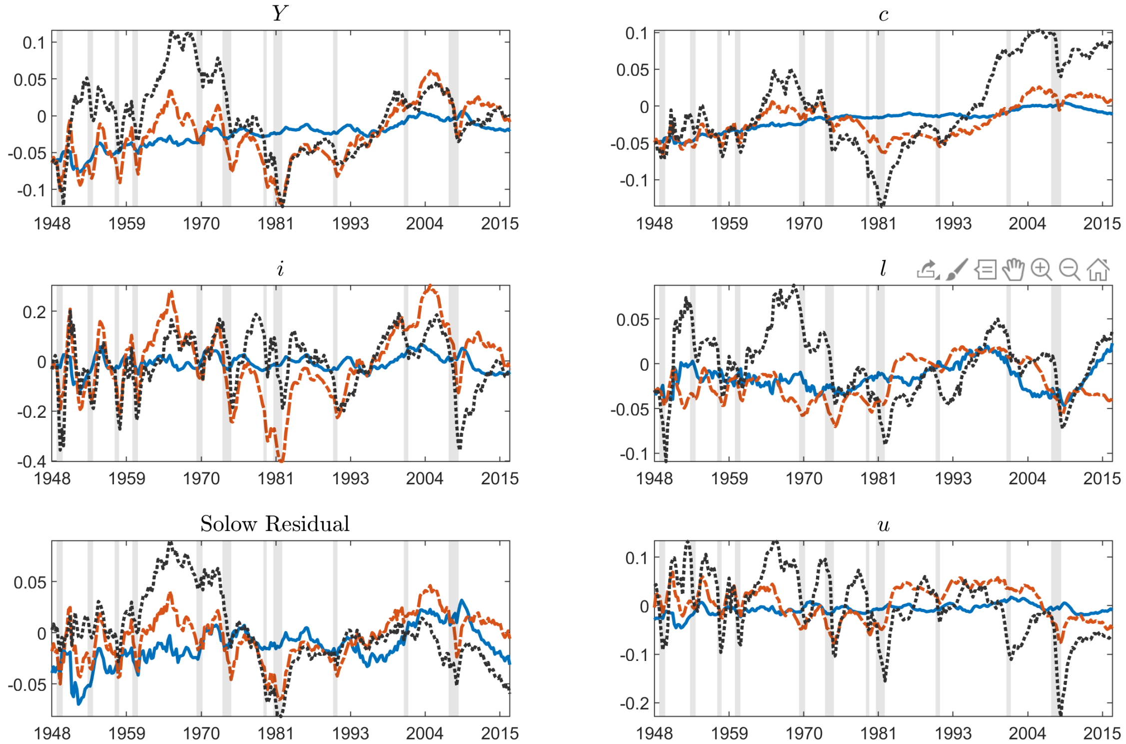

Figure 4.6: The solid lines are for the full CU model. The dash-dotted lines are for the full VU model. The dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA, the BLS, and the FRB. All variables are logarithms of the original series.

Model Predicted Real Wage Rate and the U.S. Data

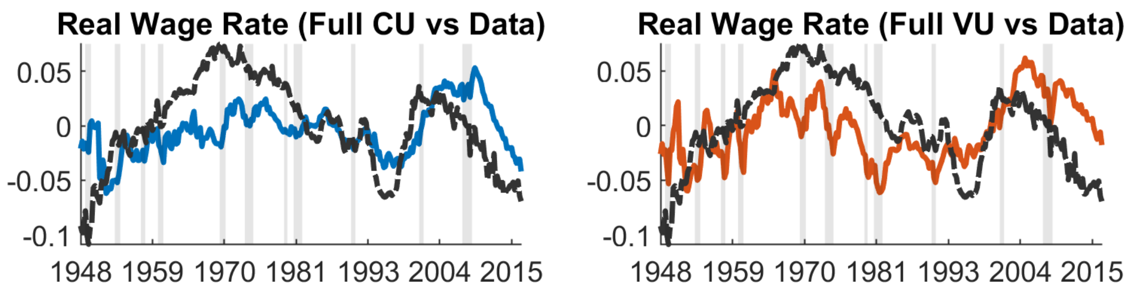

Figure 4.7: The solid lines are the model predicted values. The dash-dotted lines are the U.S. data. Shaded areas indicate the NBER dated recessions. The data is from the BEA and the BLS. All variables are logarithms of the original series.

4.5 Parametric Bootstrapping

The confidence intervals shown in this thesis are estimated using the following parametric bootstrapping procedure.

Suppose there are two observed variables $(y_t, x_t)_{t=1}^{T}$. I assume that the joint stochastic process of the two variables is characterized by a simple linear relationship:

$$y_t = \beta x_t + \epsilon_t, \tag{4.3}$$

where x_t and ϵ_t are independent, and each follows an auto-regressive (AR) process with Gaussian white noise.

First, I use the ordinary least square (OLS) method to estimate β. The estimated coefficient is denoted by $\hat{\beta}$. The estimated residuals are obtained based on the estimated coefficient: $\hat{\epsilon}_t \equiv y_t - \hat{\beta} x_t$. I then use the maximum likelihood method to estimate the AR process followed by x_t and the AR process followed by ϵ_t using the observed x_t and the estimated residuals $\hat{\epsilon}_t$. The Bayesian information criterion (BIC) is used to select the best number of lags.

Based on the above estimated parameter values, I re-sample the sample data 5,000 times to find the standard error and the 95% confidence interval of the statistics that I am interested in.

4.6 Cyclicality of the Average Real Wage Rates

Table 4.5 shows the cyclicality of the average real wage rates of different sectors including the business sector, the no-farm sector, and the non-financial corporations. The average nominal wage rates are deflated using different deflators, e.g., the GDP deflator, the consumer price index (CPI), and the own sector deflator. The data is from the BEA and the BLS and the sample period is from the first quarter of 1948 to the first quarter of 2017. The results show that the average real wage rates are roughly acyclical.

Table 4.5: Cyclical Properties of the Average Real Wage Rates

	GDP Deflator	CPI	Own Sector Deflator
Business	0.08 $[-0.03, 0.18]$	0.08 $[-0.05, 0.21]$	0.05 $[-0.07, 0.17]$
No-farm	0.05 $[-0.05, 0.15]$	0.05 $[-0.07, 0.18]$	0.11 $[0.00, 0.22]$
Non-financial Corp.	0.04 $[-0.07, 0.15]$	0.04 $[-0.09, 0.18]$	0.02 $[-0.11, 0.14]$

Note: The table shows the covariance between the real wage rate and the GDP relative to the variance of the GDP. The brackets are the 95% confidence intervals calculated using a parametric bootstrapping method. The U.S. data is from the BEA and the BLS. All variables are Hodrick-Prescott (HP) filtered logarithms of the original series.