Excess Capacity in a Fixed-Cost Economy *

Daniel Murphy
University of Virginia Darden School of Business
October 10, 2016

Abstract: This paper proposes a new mechanism that can explain persistent economic slack. The theory shows that when producers face negligible marginal costs and desired spending is below the economy’s capacity, the economy features slack in equilibrium, even when prices are flexible and there are no other frictions. A heterogeneous household version of the model demonstrates how an economy can enter a capacity trap in response to a temporary negative demand shock: when demand by some consumers falls temporarily, other consumers’ permanent income (and hence their desired consumption) also falls. Since output is determined by demand, the permanent fall in desired consumption causes a permanent state of excess capacity.

Keywords: Excess capacity, Flexible prices, Business cycles, Fixed costs, Capacity Traps

JEL: E12, E22, E32

* I am grateful to Alan Deardorff, Ed Leamer, Eric Leeper, Casey Lichtendahl, Elena Loutskina, Toshihiko Mukoyama, Pontus Rendahl, Kieran Walsh, Frank Warnock, and seminar participants at the University of Virginia for helpful comments and discussions, and to Justin Pierce and Norm Morin for helpful discussions about the Federal Reserve Board’s capacity utilization measure. An early draft of this paper was lightly circulated under the title, “A Shopkeeper Economy.”

Contact: University of Virginia Darden School of Business, Charlottesville, VA 22903. E-mail: murphyd@darden.virginia.edu
1. Introduction
Between 2009Q1 and 2013Q4, the difference between potential output and real GDP averaged $900 billion in 2009 dollars, or approximately 5.6% of potential output. The persistent output gap has inspired new theoretical models to help explain economic slack (e.g., Michaillat 2012; Michaillat and Saez 2015; Eggertsson and Krugman 2012; Rendahl 2016). A common feature of this theoretical work is that economic slack depends on price rigidity. In this paper I propose a new theory of economic slack in which prices are fully flexible.

The working definition of slack explored here is the value of goods and services that firms (or workers) could produce without incurring additional costs. This definition is consistent with the notion of excess capacity (which I use synonymously with slack) underlying the Federal Reserve Board’s index of capacity utilization (Figure 1). This notion of excess capacity is distinct from the notion of suboptimal production in canonical models of imperfect competition. In Blanchard and Kiyotaki (1987), for example, monopolists incur marginal production costs, and in that sense have no excess capacity even when their production levels are socially suboptimal.

Figure 1: Excess Capacity

Note: Data from the Board of Governors of the Federal Reserve based on the Quarterly Survey of Plant Capacity. The excess capacity series indicates how much more firms can produce without incurring additional costs.
This paper shows that when monopolistically competitive producers operate in regions of negligible marginal costs (NMC), economic slack can occur in equilibrium even when prices are flexible. I demonstrate this result in the simplest setting possible. The economy is static and NMC firms take as given their production capacity. Homogenous workers inelastically supply labor as operators of the firms. Additional output is costless to the firm when production is below capacity. If demand curves feature price-dependent demand elasticities, excess capacity exists in equilibrium when consumers’ preferences for consumption are sufficiently low relative to firms’ production capacity.

The result is quite different when firms are assumed to face marginal production costs. I show that when the setting is modified only slightly to incorporate marginal production costs (along the lines of a closed-economy version of Krugman [1980]), aggregate output depends on labor supply and production technology regardless of consumers’ level of utility from consumption or the nature of competition in the goods market. In other words, the assumption of marginal costs is equivalent to imposing Say’s Law that “supply creates its own demand.” In the presence of marginal costs, income depends on factor supply, whereas in the NMC economy, income depends on spending.

In the setting with positive marginal costs, there is no effect of monopoly power (relative to perfect competition) on equilibrium output, which may seem inconsistent with the standard distortionary effect of monopoly power in models such as Blanchard and Kiyotaki (1987). The key difference between my setting and that in Blanchard and Kiyotaki (1987) is that labor is inelastically supplied in my model and hence there is no effect of markups on the marginal rate of substitution between consumption and labor. While monopoly power in my model with marginal costs has the standard partial equilibrium effect of reducing firms’ output for a given wage, in general equilibrium, prices (and wages) adjust so that aggregate output equals the effective labor supply. The assumption of inelastic labor provides a clear notion of an economy’s capacity level and isolates the effect of desired spending on excess capacity from the effect of consumption preferences on labor supply.

In the economy with NMC firms, aggregate output depends entirely on a representative consumer’s taste for consumption, which is exogenous. In an extended version of the model, I show that incorporating heterogeneous households gives rise to endogenous aggregate demand which does not perfectly track an aggregate measure of preferences. When some agents receive a
large share of income (and others a small share), the economy can enter a capacity trap in response to a temporary shock to consumer preferences. Specifically, when rich agents (those receiving a large share of income) temporarily demand less, poor agents (those receiving a small share of income) choose to permanently lower their consumption each period to smooth their consumption over time. Since aggregate income is determined by aggregate (poor plus rich) demand, aggregate income falls permanently and excess capacity increases. The model with rich and poor households also predicts that an increase in inequality due to a decline in the income share of poor households causes an increase in excess capacity, thus providing a potential explanation for the upward trend in excess capacity (Figure 1) that coincided with increasing inequality in the U.S. during the latter part of the 20th century.

Michaillat and Saez (2015) is the most closely related paper to mine in that it also examines notions of excess capacity. There are a number of important distinctions between their work and mine. First, excess capacity in their framework relies on a matching friction between buyers and sellers. In the absence of such a friction, their economy achieves the optimal level of output. In my model, in contrast, there is no friction preventing buyers from meeting sellers; rather excess capacity results from low desired purchases by consumers. Second, excess capacity and output depend on consumer preferences in my model even when prices are flexible. In Michaillat and Saez (2015), aggregate demand has no effect on excess capacity or output when prices are flexible. In that sense, my model offers a new and complementary perspective on the forces that contribute to demand-determined output and excess capacity.

The remainder of the paper proceeds as follows. Section 2 presents a model with positive marginal costs as a baseline for comparison with the NMC model. I show that when labor supply is inelastic, the budget constraint pins down aggregate output regardless of price-setting behavior or the nature of competition in goods markets. Section 3 presents the model with NMC firms and a representative household. It also discusses the model’s assumptions. Section 4 presents the model with rich and poor households.

2. Implications of Marginal Costs for the Determination of Aggregate Output.
Here I present a one-country adaptation of the model in Krugman (1980) to demonstrate the implications of marginal costs for the determination of aggregate output. The model features an inelastic supply of labor, a perfectly competitive labor market, and firms that face positive
marginal production costs. There is a representative consumer who inelastically supplies \( L \) units of labor. The consumer owns a mass \( J \) of firms, indexed by \( j \in [0, J] \), each of which produces a differentiated good.

Consumer Preferences. A representative consumer maximizes the utility function

\[
U(q_j, \ldots q_J)
\]

subject to the budget constraint

\[
wL + \int_{j \in J} \Pi_j dj = \int_{j \in J} p_j q_j dj,
\]

where \( w \) is the wage, \( \Pi_j \) represents firm \( j \)'s profits, \( p_j \) is the price of good \( j \), and \( q_j \) is the quantity consumed of good \( j \). Goods enter the utility function symmetrically.

Firms. Assume that each firm uses the production technology \( q_j = l_j / a \), where \( a \) is the unit labor requirement and \( l_j \) is the amount of labor hired by firm \( j \). Therefore marginal production costs are \( wa \). I drop firm subscripts henceforth since firms are identical. For simplicity I abstract from fixed costs. Then a firm’s profits can be written as \( \Pi = q(p - wa) \). Firms take the wage and demand curve as given and choose a price to maximize profits.

Equilibrium Output. Output can now be derived simply by substituting firm profits into the household’s budget constraint:

\[
wL + \int_{j \in J} q_j (p - wa) = \int_{j \in J} p_j q_j dj.
\]

Rearranging yields

\[
wL = wJa
\]

or

\[
q = \frac{L}{Ja}
\]

which states that equilibrium firm-level output is equal to effective labor supply per firm. Aggregate output \( Q \) can be defined as

\[
Q = \int_{j \in J} q_j = \frac{L}{a}
\]

Note that prices are yet to be determined, and for that, we would need to specify a utility function and the nature of competition between firms. Even without taking that extra step, we can see that in the presence of marginal costs, the budget constraint alone pins down aggregate output as a function of labor and production technology. We have examined this result in a simple context in
which firms each have identical constant-marginal-cost technologies and zero fixed costs, but the result can be shown to hold more generally.

Below we relax the assumption of positive marginal costs. Doing so requires more than simply setting unit costs to zero because, as equation (1) demonstrates, setting $a = 0$ would result in a level of output that is undefined when labor markets clear. In particular, the model below will relax the assumptions of positive marginal costs and competitive labor markets. In Section 3.3, I discuss modifications of the NMC model that are compatible with labor-market clearing.

3. Model of Negligible Marginal Costs.
This section presents the general-equilibrium implications of negligible marginal costs under flexible prices. I first present a static general-equilibrium model that delivers the paper’s main result (Section 3.1). The model specifies a utility function for concreteness. Section 3.2 demonstrates the generality of the result and derives the necessary and sufficient conditions on the utility function for there to be a flexible price equilibrium with excess capacity. Section 3.3 discusses the correspondence between the model’s assumptions and microfoundations of firm and worker behavior. It also discusses robustness of the model’s results to alternative setups such as firm entry and alternative forms of competition in labor markets.

3.1 General-Equilibrium Model of a NMC Economy.
Here I develop a static general-equilibrium model in which aggregate output depends on a demand parameter rather than on supply of factor inputs. There is a mass $J$ of workers indexed by $j \in [0, J]$, each of which belongs to the representative household. Each worker operates a firm of an identical index. Each firm has a capacity level of output $\bar{q}$ which represents the maximum level of output that the firm can produce. There are no marginal production costs when output is below the capacity level. A firm operator could represent a barber, for example, who internalizes no additional costs of providing haircuts for up to 40 hours in a week but does not work more than a 40-hour workweek.

Consumers. Workers earn revenues for operating firms and remit the revenues to the representative household. The household consumes on behalf of the workers. The household maximizes utility over the differentiated goods/services indexed by $j \in [0, J]$,.
where $\theta$ is a taste parameter and $\gamma$ is a parameter that dictates the elasticity of substitution between goods. For simplicity $\theta$ is assumed to be equal across goods and hence represents the household’s preference for aggregate consumption. Equation (2) is a modified version of the utility function used by Ottaviano, Tabucci, and Thisse (2002); Melitz and Ottaviano (2008); and Foster, Haltiwanger, and Syverson (2008). It leads to analytically tractable demand curves with price-dependent demand elasticities. It is important to note that utility functions yielding price-dependent demand elasticities will suffice for the existence of an equilibrium featuring excess capacity, but utility functions that yield constant price elasticities do not. Section 3.2 discusses this in more detail.

The representative consumer’s budget constraint is

$$\int_0^J \Pi_j dj = \int_0^J p_j q_j dj,$$

where $\Pi_j$ is the profit from ownership of firm $j$. Since the firm owner’s profit equals revenues, the budget constraint in the NMC economy is simply a restatement of the identity that revenues equal expenditure.\(^1\) Therefore, in contrast to the model from Section 2, here the budget constraint does not pin down equilibrium output. Consumer optimization yields the following demand curve:

$$q_j^d = \frac{1}{\gamma} \left( \theta - \lambda p_j \right),$$

where $\lambda$ is the multiplier on the agent’s budget constraint.

Firms. Firms are operated by a shopkeeper who can costlessly provide services up to the firm’s capacity level $\bar{q}$. Actual production requires a customer (the representative household) to make a purchase, just as the provision of a haircut requires a customer to purchase the service from a barber. Firms therefore do not have a production function in the traditional sense. Rather, the firm’s output can be written as $q_j = \min(q_j^d, \bar{q})$.

\(^1\) An implicit assumption is that the representative consumer does not internalize the fact that higher spending increases aggregate income when maximizing utility. The same results derived in this section can be shown to hold in an environment with a continuum of consumers with identical utility, each of which owns its own firm. In that case, individual consumers are correct to perceive that their spending has a negligible effect on their own income (even though aggregate income depends on aggregate spending). Since the results from both setups are identical, I focus on case of the representative consumer for ease of exposition.
Even though additional output is costless to the firm, it does not produce at capacity if it is not profit maximizing to do so. Each firm chooses a price to maximize revenues $\Pi_j = p_j q_j^d$. The profit-maximizing price is

$$p_j = \frac{\theta}{2\lambda}, \quad (5)$$

Given the price, the quantity demanded is

$$q_j^d = \frac{\theta}{2\gamma}, \quad (6)$$

where I assume that $\theta/2\gamma$ is strictly less than the firm’s capacity level $\bar{q}$. Note that the effect of the agent’s budget multiplier on the price exactly offsets its effect on the quantity demanded, so the resulting quantity does not depend on the multiplier.  

Definition of the Equilibrium: An equilibrium consists of a set of prices $p_j$ and quantities $q_j$ such that (i) consumers choose consumption to maximize utility subject to their budget constraint while taking prices as given, and (ii) firms maximize profits, taking consumers’ demand curves as given.

Equation (6) pins down quantities as a function of exogenous parameters. For completeness, we can write aggregate output as

$$Q = \int_{j \in J} \frac{\theta}{2\gamma} = J \frac{\theta}{2\gamma}. \quad (7)$$

Assuming that each firm is below capacity (and in particular that $\frac{\theta}{2\gamma} < \bar{q}$), additional output requires only additional demand in the form of a higher taste parameter. This result is in stark contrast to the case of non-negligible marginal production costs examined in Section 2.

Excess Capacity. The shopkeeper setup provides a simple but straightforward notion of economic slack:

Definition: Economic slack is defined as

---

2 A common concern with quadratic utility is that it implies a bliss point for consumption, which may or may not be a feature of consumers’ preferences. The possibility that consumption could reach its bliss point is not a concern in my model, since prices are such that consumption is always below the bliss point. Specifically, consumption is always half of the bliss point level.
\[ S = \begin{cases} \sum_{j \in J} (\bar{q}_j - q_j) & \text{if } q_j < \bar{q} \\ 0 & \text{otherwise} \end{cases} \]

By substituting in for equilibrium output, we can derive the equilibrium level of slack as

\[ S = \begin{cases} f \left( \bar{q} - \frac{\theta}{2\gamma} \right) & \text{if } \frac{\theta}{2\gamma} < \bar{q} \\ 0 & \text{otherwise} \end{cases} \]

So far, we have assumed parameter values which lead to slack for all firms in equilibrium. It is straightforward to analyze a situation in which demand is sufficiently high that firms optimally choose to produce at capacity. For firms that are below capacity, the price is given by (5) and the resulting quantity is given by (6). If the resulting demand from the below-capacity price exceeds \( \bar{q} \), then the price simply adjusts so that the quantity demanded in (4) is equal to the upper bound on supply \( \bar{q} \). Therefore a firm’s optimal price is

\[ p = \begin{cases} \frac{\theta}{2\lambda} & \text{if } \frac{\theta}{2\gamma} < \bar{q} \\ \frac{1}{\lambda} (\theta - \gamma \bar{q}) & \text{if } \frac{\theta}{2\gamma} \geq \bar{q} \end{cases} \]  

The basic model of shopkeepers implies that demand stimulus is state-dependent. Increases in \( \theta \) are associated with increases in output whenever firms are below capacity. When firms are at capacity, there is no effect of \( \theta \) on output. This prediction of the model is similar to the prediction in the search-based model of Michaillat (2012) and is consistent with evidence of state-dependent fiscal multipliers (e.g., Auerbach and Gorodnichenko 2012; Demyanyk, Loutskina, and Murphy 2016).

3.2. Generality of the Results.

The baseline NMC model assumes a quasilinear demand curve, a fixed number of firms operated by workers who don’t compete in a labor market, and imperfect competition in the goods market. Here I demonstrate the sensitivity of the model’s implications to changes in each of these assumptions.

3.2.1. Demand Curves. In order for slack to occur as an equilibrium outcome, firms must maximize profits at a quantity below the capacity level. In the textbook case of monopoly power
under zero marginal costs, firms optimally produce at a point on the demand curve where the
price elasticity of demand is unity. Therefore demand curves must have price-dependent
elasticities of demand that cross unity. If the elasticity of demand is everywhere greater than
unity, firms’ revenues (and hence profits under NMC) are always decreasing in the price, and
hence firms choose a price such that they produce at capacity.

More formally, consider a monopolistically competitive firm facing a monotonically
decreasing demand function \( q_j(p_j) \) and revenue function \( R_j(p_j) = p_j q_j(p_j) \). Since the firm
goes no marginal costs, it maximizes profits by maximizing revenue. Let \( R_j(p_j) \) be continuously
differentiable and satisfy \( R_j''(p_j) < 0 \) for all \( p_j \). Let \( p_j^* = \arg\max_{p_j}\{R_j(p_j)|q_j(p_j) < \bar{q}_j\} \).

Slack is strictly positive if the revenue function satisfies \( R_j''(p_j) < 0 \) for all \( p_j \) and
\( R_j'(p_j) = 0 \) for some \( p_j \) such that \( q_j(p_j) < \bar{q}_j \). The sufficient condition for the existence of firm-
level slack is that the revenue function is concave and reaches its maximum at a quantity below
the capacity level.

**Proposition:** The sufficient condition for positive slack is that the price elasticity of demand,
\( \varepsilon(p_j) = \left[ \frac{\partial q_j}{\partial p_j} \right]_{p_j} q_j \), is unity at some \( q_j(p_j) < \bar{q}_j \).

**Proof:** This is a standard result in microeconomics. Revenue maximization implies that \( q_j(p_j) + p_j q_j'(p_j) = 0 \), or \( \frac{q_j(p_j)}{p_j} = |q_j'(p_j)| \). Substituting the revenue first-order condition into the
definition of the price elasticity of demand yields
\[
\varepsilon(p_j^*) = \left[ \frac{q_j(p_j^*)}{p_j^*} q_j(p_j^*) \right]_{p_j^*} = 1.
\]

For any downward-sloping demand curve with nonconstant price elasticities of demand, this
condition simply states that the elasticity of demand is increasing in the price. Examples of
theories analyzing these types of demand functions include Kimball (1995), Eichenbaum and
Fisher (2007), and Dotsey and King (2005). The commonly analyzed demand functions featuring
constant and elastic demand elasticities do not imply slack; in that case, the revenue function is
always decreasing in the price and firms optimally produce at capacity. Although demand
functions with constant elasticities are often analyzed due to their convenient analytical
properties, empirical evidence suggests that demand elasticities are increasing in the price (e.g., Nakamura and Zerom 2010; Foster, Haltiwanger, and Syverson 2008). The following proposition derives the conditions on the utility function under which price elasticities of demand cross unity.3

Proposition: Let \( u(q_j) \) be a representative consumer’s utility function which is increasing, concave, and differentiable over good \( q_j \). Then positive slack exists if and only if there exists a \( q_j^* < \bar{q}_j \) such that

\[
    u''(q_j^*) + u'(q_j^*)q_j^* = 0. \tag{9}
\]

Proof: Consumer optimization requires that \( u'(q_j) = \lambda p_j \), where \( \lambda \) is the multiplier on the budget constraint. The consumer’s demand curve is implicitly given by \( p_j(q_j) = u'(q_j)/\lambda \). Firms maximize \( \Pi_j = q_j \times p_j(q_j) = q_j u'(q_j)/\lambda \), which implies that the optimal quantity demanded must satisfy (9). To see that the demand elasticity is unity, note that substitution yields

\[
    \epsilon = \left| \frac{1}{u''(q_j)} \left( \frac{\partial u'(q_j)}{\partial q_j} \right) \right| = \left| \frac{u'(q_j)}{u''(q_j)q_j} \right|,
\]

which equals unity by (9).

3.2.2. Firm Entry and Competition in Labor Markets. The baseline model treats as fixed the number of firms in the economy. An alternative approach is to assume that heterogeneous firms face fixed labor costs, and that firms enter until labor markets clear. Here I outline modifications to the baseline NMC model that incorporates competitive labor markets and fixed costs.

As in Section 3.1, there is a representative household with utility given by (2), which implies that demand is given by (4). In contrast to the baseline model with a fixed number of firms operated by a fixed number of workers, here firms will enter until the labor market clears. Aggregate labor supply is \( L \).

Each firm requires \( l_j^F \) units of operating labor. Let there be a continuum of firms with fixed labor costs distributed over \([0, \infty)\). Firms that survive in equilibrium are those with positive

---

3 I am indebted to Pontus Rendahl for suggesting this proposition and for pointing out that equilibrium quantities are independent of the Lagrange multiplier for any utility function satisfying equation (9).
profits. Since labor is a fixed-only cost, it does not alter the firm’s pricing decision or its revenue. Prices and quantities are therefore given by equations (5) and (6). Profits, however, must net out the fixed cost. In the case of quadratic utility, profits become

$$\Pi_j = pq - wl_j^f = \frac{\theta^2}{4\lambda} - wl_j^f.$$  

Firms in $$J^* = \left\{j : w l_j^f < \frac{\theta^2}{4\lambda}\right\}$$ earn positive profits.

**Equilibrium.** We have already solved for prices and output for firms that exist in equilibrium. To derive aggregate output, we must also solve for the mass of firms that earn positive profits.

Let $$l_j^f$$ be uniformly distributed such that each firm’s labor requirement is equal to its index. Then the mass of firms can be written as

$$J^* = \frac{\theta^2}{w4\lambda'},$$  \hspace{1cm}(10)$$

which is a function of $$w\lambda$$. To solve for $$J^*$$ as a function of exogenous parameters, note that labor demand is the sum of individual firms’ fixed labor requirements:

$$L^D = \int_0^{J^*} j, dj = \frac{j^2}{2} \bigg|_0^{J^*} = \frac{1}{2} \left( \frac{\theta^2}{w4\lambda'} \right)^2.$$  

Labor-market clearing equates labor demand with the fixed supply $$L$$. Solving for $$w\lambda$$ yields

$$w\lambda = \left( \frac{1}{2L} \right)^5 \frac{\theta^2}{4}.$$  

Substituting for $$w\lambda$$ in (10) and solving for the mass of firms yields $$J^* = \sqrt{2L}$$, which implies that aggregate output is $$Q \equiv J^* q = \sqrt{2L} \frac{\theta}{2\gamma}$$.

When labor markets are competitive, aggregate output depends on consumer demand, but the equilibrium mass of firms depends only on labor supply and firms’ fixed labor requirements. One way to permit consumer demand to affect firm entry is to replace competitive labor markets with revenue sharing between workers and firms. Appendix B presents a model of this type.

**3.3.3. Discussion of Assumptions.** The approach in this paper is motivated by the observation that many firms and workers operate in regions of negligible marginal costs. For example, a barber can provide additional haircuts without incurring marginal costs, up until the point at which he reaches capacity. Existing literature supports this notion. Workers often supply
their labor in increments, with typical hours clumped at 40 per week (e.g., Card 1990; Faber 2005), and firms likewise often operate in regions of fixed-only costs in the form of overhead capital and worker salaries (e.g., Brown 1992; Oi 1962; Rotemberg and Summers 1990). The existence of negligible marginal costs is also consistent with evidence of excess capacity and associated idleness (Figure 1). Indeed, the excess capacity series measures how much more firms could produce without incurring additional costs.4

One objection to the assumption that some firms face negligible marginal costs is that intermediate inputs are marginal costs even if labor is a fixed-only cost. The theoretical framework developed here can easily be extended to incorporate intermediate inputs and/or investment. In an extended model, the macroeconomy will feature excess capacity when suppliers somewhere along the chain of production face fixed-only costs. Therefore, even if final-goods firms face marginal costs of intermediate inputs, the intermediate-goods producers may have excess capacity if they produce under regions of negligible marginal costs. In particular, the mechanism is equally relevant with monopolistically competitive workers who face negligible marginal costs and firms that hire workers (or intermediate inputs) as marginal costs.

The model assumes that a firm’s capacity level (the maximum level of output it can produce) is fixed. I do not explicitly model the microfoundations for why producers may face (or behave as if they face) predetermined levels of capacity. Let me briefly discuss a number of potential microfoundations. First, the nature of production may imply indivisible capacity levels. I have already discussed the example of a barber who has no disutility from work during a certain amount of time (e.g., 40 hours). Similarly, consider a convenience store for which demand fluctuates between 5 and 10 customers per hour and for which the per-hour profits are positive even if only 5 customers arrive. The cashier is indifferent between serving few or many customers in an hour and is capable of serving as many as 15. The firm cannot choose a capacity level of less than a single cashier simply based on the nature of the service provided. Even if the store could perfectly contract with the cashier to only work when customers arrive, the economy

---

4 In addition to the fact that the excess capacity index is explicitly designed by the Federal Reserve Board to reflect how much more firms can produce without incurring additional costs, it is informative that the predominant reason that survey respondents give for excess capacity is “insufficient orders” for their output. Other possible reasons for slack include “insufficient supply of local labor force skills,” “lack of sufficient fuel or energy,” “equipment limitations,” and “logistics/transportation constraints.” On average, 80% of respondents with excess capacity cite “insufficient orders” as the primary reason, and nearly 90% cited insufficient orders during the Great Recession (Stahl and Morin 2013).
would feature excess capacity if the cashier experiences no marginal disutility from serving additional customers (e.g., inelastically supplies more labor than is demanded). Indeed, many workers prefer to be busy simply because they are less bored and feel an intrinsic sense of value in their work that may offset the opportunity cost of leisure.

Second, even if a firm can choose its capacity level (and higher capacity is associated with additional costs) and it can perfectly forecast demand, it may choose a level of capacity such that it experiences slack the majority of the time. Consider a manufacturing firm deciding what size plant to build. 90% of the time the firm has orders for 8 widgets and the remaining 10% of the time it has orders for 10. Inventory storage is prohibitively costly. The firm must choose prior to production whether to build a plant that can accommodate 8 widgets per month or 10. The firm will build the larger plant if the additional revenues from the extra two widgets cover the costs of a larger plant. In that case, 90% of the time the firm is only producing at 80% of its capacity and could costlessly accommodate additional orders for widgets.5

Finally, many services can be costlessly reproduced and therefore have nearly unbounded capacity. Perhaps the most stark and increasingly relevant example of firms with fixed-only costs is technology and media companies. Once new software is developed or a new movie filmed it can be costlessly distributed. For a firm in these industries the number of addition sales it can accommodate without incurring additional costs (and hence its capacity level) is nearly limitless.

4. Application: Temporary Demand Shocks and Capacity Traps.
In the representative agent NMC economy, demand is determined by an exogenous parameter $\theta$. Here I show that incorporating heterogeneous households gives rise to endogenous aggregate demand that does not perfectly track an aggregate measure of preferences. When some agents receive a large share of income (and others a small share), the economy can enter a capacity trap in response to a temporary shock to consumer preferences. Specifically, when rich agents (those receiving a large share of income) temporarily demand less, poor agents (those receiving a small share of income) choose to permanently lower their consumption each period to smooth their consumption over time. Since aggregate income is determined by aggregate (poor plus rich) demand, aggregate income falls permanently and excess capacity increases.

5 See Fine and Freund (1990) for a general formalization of optimal capacity investment under demand uncertainty.
The model extension below formalizes this mechanism. Differential income shares can arise for a number of reasons, including different weights over high-skilled and low-skilled labor in the production of the fixed cost. For simplicity I assume that agents do not change their income shares over time. Incorporating this channel would change the persistence of the capacity trap without changing the result that a temporary shock can have long-lasting effects.

To pin down the interest rate, the model also features a good that is endowed each period. For simplicity, I assume that the rich agents also own the endowment. The endowment represents land-or-capital-intensive production where the factors of production are owned by the rich. An extensive body of research documents the strong relationship between wealth and income across households in the data (e.g., Saez and Zucman 2014). In addition to serving as a modeling device, the assumption about the ownership of endowment income is consistent with the relationship in the data.

For simplicity, and to facilitate derivation of analytical results, I assume that all uncertainty is resolved after the initial period. Without loss of generality, each agent type has a net bond position of zero in the initial period. Agents subsequently trade bonds to satisfy their desired time paths of consumption, subject to a no-Ponzi constraint that the present value of their asset position must be weakly greater than zero.

4.1. A Model of Capacity Traps.

Consumers. Rich and poor households, denoted by \( h \in \{R, P\} \), each maximize utility,

\[
U^h = \sum_{t=0}^{\infty} \beta^t \left( y^h_t + \theta^h q^h_t - \frac{Y}{2} (q^h_t)^2 \right),
\]

subject to

\[
\Pi^h_t + B^h_t + e^h_t = p_t q^h_t + Q_t B^h_{t+1} + y^h_t,
\]

where \( q^h_t \) is agent \( h \)’s consumption from the NMC sector in period \( t \), \( \Pi^h_t \) is agent \( h \)’s income from the NMC sector of the economy, \( e^h_t \) and \( y^h_t \) are \( h \)’s endowment and consumption of the numeraire, and \( Q_t \) is the price of a bond \( B_{t+1} \) that pays a unit of the numeraire in period \( t + 1 \). Agents must satisfy the no-Ponzi condition

\[
p_0 q^h_0 + y^h_0 + \sum_{t=1}^{\infty} Q_t (p_t q^h_t + y^h_t) \leq \Pi^h_0 + e^H_0 + \sum_{t=1}^{\infty} Q_t (\Pi^h_t + e^H_t),
\]
which states that the present value of their consumption is no greater than the present value of future income.

A convenient feature of the utility function is that agents consume only the good from the NMC sector when their income is sufficiently low. This feature, along with the assumption that poor agents are not endowed with the numeraire, \( e^p_t = 0 \ \forall \ t \), simplifies the analysis without loss of generality.

**Firms.** Output in the NMC sector is produced by a single monopolist who hires high-skilled labor (owned by rich households) and low-skilled labor (owned by poor households) as fixed costs. The firm’s output is equal to consumption by the rich and poor households: \( q_t = q^p_t + q^R_t \). As in the baseline model, firm production equals desired consumption (under the assumption that firms are below capacity). The poor receive a share \( \alpha \) of revenues (so that \( \Pi^p_t = \alpha p_t q_t \)), and the rich receive the remaining \( (1 - \alpha) \) share. I do not explicitly model why income shares differ, although one possibility is differences in bargaining power between managers and workers. The proposition below demonstrates that when the poor’s share of income is sufficiently low, the poor consumes only from the NMC sector.

**Proposition A:** If \( \theta^h_t > 1 \ \forall \ t, h \) then there exists a threshold value \( \bar{\alpha} \), such that for all \( 0 < \alpha < \bar{\alpha} \) the poor consume only output from the NMC sector. In this case, the poor’s consumption is determined by their budget constraint and the Euler equation

\[
\theta^p_t - \gamma q^p_t = \frac{p_t}{p_{t+1}} (\theta^p_{t+1} - \gamma q^p_{t+1}).
\]

(11)

**Proof:** Appendix A.

The precise timing of poor consumption depends on changes in the price of the consumption good, and when prices are invariant across time, the poor perfectly smooth consumption.

To see how quantitatively important the income constraint can be for the consumption of the poor, consider the poor household’s budget constraint in a deterministic steady state:

\[
\sum \beta^t p_t q^p_t = \sum \beta^t \alpha p_t (q^R_t + q^p_t).
\]

If all exogenous variables are constant across time, then the budget constraint reduces to \( q^p = \frac{\alpha}{1-\alpha} q^R \), which implies that when income shares are constant, a percent change in consumption by the rich causes an equal percent change in consumption by the poor household.
**Price Setting.** The monopolist sets a price in each period to maximize revenue $p_t(q_t^p + q_t^R)$. The demand curve of the rich is $q_t^R = \frac{1}{\gamma}(\theta_t^R - p_t)$, which is based on the rich household’s first-order conditions. The poor household’s demand curve is

$$q_t^p = \frac{1}{\beta_t^p p_t}(l - \sum_{s \neq t} \beta_s^p q_s^p p_s),$$

where $l$ is the poor household’s permanent income (defined explicitly in the Appendix). The demand curve captures the fact that a current increase in the price of the consumption good acts effectively as a reduction in the poor household’s permanent income. The implicit assumption is that the monopolist does not internalize the positive effect of a price increase on the income (and hence demand) of the poor. Given the demand curves, the monopolist maximizes profits by choosing the price

$$p_t = \frac{\theta_t^R}{2}.$$

(12)

**Equilibrium.** The time paths of consumption, prices, and output are fully determined by the time paths of the preference parameters and the income share of the poor. The assumption that all uncertainty is resolved after the initial period $t = 0$ permits the analytical derivation of the following comparative statics:

**Proposition B:** When $\alpha < \bar{\alpha}$, a fall in the consumption preference of the rich leads to a permanent fall in aggregate output and income

$$\frac{dq_t}{d\theta_0^R} = \frac{\alpha}{1 - \alpha} \frac{1}{2\gamma} + \frac{1}{\gamma} \frac{\theta_t^P}{\theta_0^R}(1 - \beta) > 0, \forall t > 0.$$

A fall in the income share of the poor causes output and income to fall in all periods:

$$\frac{dq_t}{d\alpha} < 0, \forall t.$$

**Proof:** Appendix A.

The details of the proof are left for the Appendix, but the intuition driving the result is straightforward: When the income of the poor falls (either due to a fall in their share of revenues or a fall in total revenues), the poor household reduces its consumption in future periods. Since
aggregate output and income depend on aggregate consumption, the decline in the poor household’s consumption causes a permanent fall in aggregate output.

The consumption of the poor household may initially increase in response to a fall in $\theta_0^R$ if the relative price difference between the initial period and other periods is sufficiently high. There are a number of extensions to the model that would both improve the model’s realism and prevent an initial consumption increase. One is to incorporate firm heterogeneity and amend the utility function along the lines of Melitz and Ottaviano (2008) to permit competitive effects on prices. This would mitigate the effect of relative preferences on relative prices (e.g., markups would be less procyclical), but at the cost of model simplicity. An alternative extension is to impose that household debt cannot exceed income accrued over a short period of time (rather than over a household’s lifetime). If the debt constraint is sufficiently strong, consumption falls in all periods simply because there is not enough income earned to permit an initial consumption increase. The debt constraint acts as in Eggertsson and Krugman (2012) to depress aggregate demand and hence aggregate output. An important and interesting feature of the model here, relative to the insights in Eggertsson and Krugman (2012), is that debt constraints can depress aggregate output even when prices are flexible and the interest rate is above the zero lower bound. Furthermore, debt constraints contribute to permanent effects of temporary negative demand shocks.

Inequality also has an interesting effect in this model. As $\alpha$ and the income share of the poor household fall, so does capacity utilization. Capacity utilization has been trending downward since the late-1960s, from nearly 90% to just over 80% in 2005, and even lower over the subsequent decade (Figure 1). This decline in utilization coincided with a well-documented increase in inequality over the same time period. An interesting avenue for future work is to examine this relationship in more detail, perhaps expanding the current model to an open-economy framework to see how inequality and demand in one country affect capacity utilization in its trading partners.

5. Conclusion
This paper demonstrates that the assumption of marginal costs in flexible-price general-equilibrium models is equivalent to imposing Say’s Law—that income depends on production. When firms or workers face negligible marginal costs and firms’ capacity is sufficiently high,
income depends on spending and spending is independent of production capacity. In the NMC model, spending depends on consumers’ tastes for goods and services, although the taste parameter can be interpreted broadly to capture expectations of economic growth and other determinants of spending.

The NMC model may shed light on a number of important phenomena. It offers a new mechanism to explain persistent excess capacity, even when prices are flexible and the interest rate is positive. A simple heterogeneous agent version of the model shows how an economy can end up in a persistent state of excess capacity in response to a temporary negative demand shock. It also demonstrates how increasing inequality can cause excess capacity (consistent with trends in the data over the past decade) and low aggregate output.

A key implication is that productivity growth alone may not suffice to restore output to its potential. When some agents’ consumption is constrained by low income, restoring output to its potential requires either higher spending by the rich or alternative forms of demand stimulus.

References


Appendix A.

Proof of Proposition A: Assume that the poor household’s income is sufficiently low that it consumes only the output of the NMC sector. I will derive the equilibrium and verify the parameter set under which the equilibrium consumption of the poor household yields marginal utility less than unity such that the household is indeed at a corner solution.

The poor household’s permanent income is

$$l = \alpha \left( \sum_{t=0}^{\infty} \beta^t p_t (q_t^R + q_t^P) \right),$$

and the present value of its consumption is

$$C = \sum_{t=0}^{\infty} \beta^t p_t q_t^P,$$

where I use the fact that $Q_t = \beta$, which is derived from the rich household’s first-order conditions. The Ponzi condition, $C = l$, implies that we can write

$$\sum_{t=0}^{\infty} \beta^t p_t q_t^P = \alpha \left( \sum_{t=0}^{\infty} \beta^t p_t (q_t^R + q_t^P) \right).$$

Some algebra yields

$$p_0 q_0^P (1 - \alpha) = \alpha \left( p_0 q_0^R + \sum_{t=1}^{\infty} \beta^t p_t (q_t^R + q_t^P) \right) - \sum_{t=1}^{\infty} \beta^t p_t q_t^P. \quad (13)$$

The assumption that all parameters are predetermined and constant subsequent to the initial period, $\theta_t^P = \theta_{t+1}^P \equiv \theta^P \forall t$, implies that the Euler equation becomes

$$\theta^P - \gamma q_t^P = \frac{p_t}{p_{t+1}} (\theta^P - \gamma q_{t+1}^P) \forall t; \quad q_t = q_{t+1}, \forall t > 0.$$

The fact that prices are flexibly set each period implies that $p_t = p_{t+1} \forall t > 0$. Substituting the equality of prices and output for $t > 0$ into (13) yields

$$p_0 q_0^P (1 - \alpha) = \alpha \left( p_0 q_0^R + \frac{\beta}{1-\beta} p_1 q_1^R \right) - (1 - \alpha) \frac{\beta}{1-\beta} p_1 q_1^P. \quad (14)$$

and substituting the $t = 0$ Euler equation,
\[ q_1^P = \frac{1}{\gamma} \left( 1 - \frac{p_1}{p_0} \right) \theta^P + \frac{p_1}{p_0} q_0^P, \]  

(15)

for \( q_1^P \) and rearranging yields

\[ q_0^P (1 - \alpha) \left( p_0 + \frac{p_1^2}{p_0} \right) = \alpha \left( p_0 q_0^R + \frac{\beta}{1 - \beta} p_1 q_1^R \right) + (1 - \alpha) \frac{\beta}{1 - \beta} p_1 \frac{1}{\gamma} \left( \frac{p_1 - p_0}{p_0} \right) \theta^P. \]  

(16)

The price \( p_t \) is given by (12) and the resulting demand \( q_t^R \) is

\[ q_t^R = \frac{\theta^R}{2\gamma}, \]  

(17)

which are derived from the Rich consumer’s first-order conditions and profit maximizing by the monopolist. Substituting (12) and (17) into (16) and rearranging yields

\[ q_0^P = \frac{\alpha}{1 - \alpha} \frac{\theta^R_0 \left( \theta^R_0 \right)^2 + \frac{\beta}{1 - \beta} \left( \theta^R_0 \right)^2}{(\theta^R_0)^2 + 2 \frac{\beta}{1 - \beta} \theta^R_1} + \frac{\beta}{1 - \beta} \theta^R_1 \frac{1}{\gamma} \left( \frac{\theta^R_1 - \theta^R_0}{(\theta^R_0)^2 + 2 \frac{\beta}{1 - \beta} \theta^R_1} \right) \theta^P. \]  

(18)

To verify that the poor household is indeed at a corner solution and consumes only the good from the NMC sector, it suffices to show that the marginal utility of consumption is greater than unity,

\[ \theta^P - \gamma q_t^P > 1. \]

The sufficient condition in the first period is

\[ \theta^P > 1 + \frac{\alpha}{1 - \alpha} \frac{\theta^R_0 \left( \theta^R_0 \right)^2 + \frac{\beta}{1 - \beta} \left( \theta^R_0 \right)^2}{(\theta^R_0)^2 + 2 \frac{\beta}{1 - \beta} \theta^R_1} + \frac{\beta}{1 - \beta} \theta^R_1 \left( \frac{\theta^R_1 - \theta^R_0}{(\theta^R_0)^2 + 2 \frac{\beta}{1 - \beta} \theta^R_1} \right) \theta^P. \]

This condition implicitly solves for \( \alpha \), below which the poor household in equilibrium is constrained by its budget constraint, the equilibrium price is given by (12), the consumption of the rich household in each period is given by (17), the consumption of the poor household in the initial period is given by (18), and the consumption of the poor household in subsequent periods is given by

\[ q_t^P = \frac{1}{\gamma} \left( \frac{\theta^R_0 - \theta^R_1}{\theta^R_0} \right) \theta^P + \frac{\theta^R_1}{\theta^R_0} q_0^P. \]

\[ \blacksquare \]
Proof of Proposition B: To demonstrate that output falls permanently in response to a temporary adverse demand shock, it suffices to show that \( \frac{d q_1^P}{d \theta_0^R} > 0 \). Solve for \( q_1^P = q_1^P \) by substituting the Euler equation (15) for \( q_0^P \) in the budget constraint (14). Some algebra yields

\[
q_1^P = \frac{\alpha}{1 - \alpha} \left( q_0^R + \frac{\beta}{1 - \beta} p_1 q_1^R \right) \left( 1 + \frac{\beta}{1 - \beta} \left( \frac{p_1}{p_0} \right)^2 \right) + \frac{\beta p_1}{1 - \beta p_0} \left[ \frac{1}{\gamma} \left( \frac{p_1}{p_0} - 1 \right) \theta_1^p \right] - \left[ \frac{1}{\gamma} \theta_1^p \left( 1 - \frac{p_0}{p_1} \right) \right] + \frac{\beta}{1 - \beta} \left( \frac{p_1}{p_0} \right)^2 \left[ \frac{1}{\gamma} \theta_1^p \left( 1 - \frac{p_0}{p_1} \right) \right] + \frac{\beta}{1 - \beta} \left( \frac{p_1}{p_0} \right)^2 \left[ \frac{1}{\gamma} \theta_1^p \left( 1 - \frac{p_0}{p_1} \right) \right].
\]

Substituting in the equilibrium values for \( q_1^R \) and \( p_t \) yields

\[
q_1^P = \frac{\alpha}{1 - \alpha} \frac{1}{2 \gamma} \left( \theta_0^R + \frac{\beta}{1 - \beta} \theta_1^R \right) \left( 1 + \frac{\beta}{1 - \beta} \left( \frac{\theta_1^R}{\theta_0^R} \right)^2 \right) + \frac{\beta \theta_1}{\theta_0} \left[ \frac{1}{\gamma} \left( \frac{\theta_1^R}{\theta_0^R} - 1 \right) \theta_1^p \right] - \left[ \frac{1}{\gamma} \theta_1^p \left( 1 - \frac{\theta_0^R}{\theta_1^R} \right) \right],
\]

where I have removed the superscript from the preference parameter of the rich household. Differentiation around a steady state in which the preferences of the rich are constant over time, \( \theta_0 = \theta_1 \), yields

\[
\frac{d q_1^P}{d \theta_0^R} = \frac{\alpha}{1 - \alpha} \frac{1}{2 \gamma} + \frac{1}{\gamma} \theta_0 \left( 1 - \beta \right).
\]

Appendix B. Implications for Business Cycle Comovement.

Can the theory of fixed-only costs account for business-cycle patterns? Here I infer the nature of economic fluctuation based on an extended version of the model that incorporates wage bargaining in the labor market and exogenous changes in labor efficiency and capacity. I first show that when only a single demand parameter is permitted to vary across time, the model explains the procyclicality and comovement of capacity utilization, firm entry, and markups. Prior studies have proposed theories that match the procyclicality of individual series but not the joint movement of all three.\(^6\)

I then permit time variation in labor efficiency and capacity to demonstrate the fit between the model and the data when technology, rather than demand, is permitted to vary across time. The analytical approach is similar to that in Michaillat and Saez (2015), who infer

---

the sources of economic fluctuations based on their model’s comparative statics. The general conclusion, that demand fluctuations are the primary drivers of the business cycle, is consistent with their findings and with the evidence in Galì (1999) and Basu, Fernald, and Kimball (2006). A key distinction between my model and previous studies is that my model implies a different source of frictions that lead to demand-driven output.

The model here specifies wage determination in a noncompetitive labor market. A determination of firms’ wage costs permits a determination of the number of firms that survive in equilibrium. It should be noted that there are no frictions that last between periods. Agents can perfectly reoptimize at the start of each period, so the past evolution of macro aggregates has no bearing on the current optimizing decision. Within a period, however, firms must hire labor in fixed increments.

Workers are randomly matched with a firm. The wage is determined by Nash Bargaining over revenues. The household incentivizes workers to supply indivisible labor (rather than receiving the utility value of not working and insured income) by offering contracts which specify that workers’ income is insured only under the condition that they accept sufficiently high wage offers (which are observed by the household).

Workers’ bargaining power derives from their ability to shirk. Even though there is no benefit to workers from shirking (e.g., if effort is costless) once a job is accepted, the ability of a worker to destroy firm revenues generates bargaining power that is increasing in the amount of revenues that it can affect. In existing models of holdup, firm revenue depends on the number of hours worked or workers’ effort levels (see Malcolmson [1999] and the references therein). Here I extend the intuition in these models to permit the value of an employment relationship to explicitly depend on firm-level demand rather than on worker effort or hours.

There are multiple workers per firm, each performing a unique task. The assumption of multiple workers per firm permits productivity (the inverse of the number of necessary tasks) to affect firm entry over the business cycle, although the basic insights regarding demand-induced comovement among utilization, markups, and entry hold in a simpler setup with one worker per firm. My treatment of the labor market is stylized for the sake of parsimony, but it captures the
basic features of a labor market in which (1) wages are determined by bargaining, and (2) higher productivity induces firm entry.

Model. A household consists of a mass $L$ of workers, each of which maximizes

$$U = \sum_{t=0}^{\infty} \beta^t u_t$$

where

$$u_t = b_t + \int_0^J \theta_j q_j dj - \frac{1}{2} \gamma \int_0^J \left(q_{jt}\right)^2 dj$$

(20)

and

$$b_t = \begin{cases} 0, & \text{work} \\ 1, & \text{don't work.} \end{cases}$$

(21)

Equation (21) captures the notion of indivisible labor that was formally introduced by Hansen (1985). The utility value of not working is normalized to unity, so the preference parameters $\theta_j$ affect the utility value of consumption relative to the utility value of not working. The household’s within-period budget constraint is

$$\int_0^L w_l + \int_0^J \Pi_j dj = \int_0^J p_j q_j dj,$$

(22)

where $w_l$ is the wage of worker $l \in [0, L]$.\(^7\)

Household optimization implies the same within-period demand as in (4). A firm’s optimal price is given by (5), and the resulting quantity is given by (6). Firm $j$ has revenue

$$R_{jt} = \frac{\theta_j^2}{4\gamma \lambda_t},$$

(23)

where $\lambda_t$ is the multiplier on the household’s period-$t$ budget constraint.

Labor Market. Each firm needs $N_t$ tasks to produce output (output is a Leontief technology over tasks), and the total amount of output that the $N_t$ workers can produce is the capacity level $\bar{q}_t$. In each period, each of the tasks across firms is randomly matched with a worker. If a wage contract is agreed upon, the employment relationship lasts for the duration of the period. There is only one opportunity to match with a firm each period, so matched workers’

---

\(^7\) The budget constraint (22) implies that all firm revenues are returned to the household as dividends each period.
opportunity cost of accepting a wage offer is the reservation utility $b$. The benefit to the firm of agreeing on an employment contract is real revenues minus the real wage,

$$V_{jt}^F = \lambda_t (R(\theta_{jt}) - w_{jt}),$$

where $w_{jt}$ is the wage paid by firm $j$ to a worker with which it is matched. Firms value profits at the household’s marginal utility of income since all profits are returned to the household within a period. The benefit to the household of accepting a contract is

$$V_{jt}^W = \lambda_t w_{jt} - b_t.$$

Workers likewise value income at the household’s marginal utility of income. I assume that the household does not coordinate bargaining between firms and workers even though it collects income from both.\(^8\)

Workers and firms Nash bargain over the surplus. The equilibrium wage maximizes the product of the value to the worker and the value to the firm:

$$w_{jt} = \arg\max_w \left\{ (\lambda_t w_{jt} - b_t)^\psi_t \lambda_t^{1-\psi_t} (R(\theta_{jt}) - w_{jt})^{1-\psi_t} \right\},$$

where $\psi_t$ is the workers’ bargaining power at time $t$. At an interior optimum, the resulting wage is

$$w_{jt} = \psi_t \left[ R(\theta_{jt}) + \frac{b_t}{\lambda_t} \right]. \quad (24)$$

Firms receive a new random draw of the preference parameter each period. Firms for which revenues do not exceed workers’ reservation wage will shut down, and those with sufficiently high revenues share the surplus with workers. Specifically, if $R(\theta) < Nb/\lambda$, then no wage contract is signed, the firm shuts down, and the worker is unemployed for the period. The revenue equation (23) implies that the firm will shut down when

$$\theta_{jt} < 2\sqrt{N\gamma b_t}. \quad (25)$$

Equilibrium. Equilibrium output is

$$Q_t = \int_{j \in J_t} \frac{\theta_{jt}}{2\gamma^\gamma}. \quad (26)$$

---

\(^8\) An alternative and equivalent assumption to a single household is that each worker is its own household but consumption is perfectly insured. The assumption of perfect consumption insurance simplifies the analysis because demand does not depend on the distribution of wages across workers.
where \( J_t^* \) is the set of firms satisfying the threshold revenue requirement at time \( t \). It is straightforward to check that this equilibrium quantity satisfies the household’s budget constraint as well as consumer and firm optimization.

**B.2. Correspondence between Model and the Data.**

Figure 2 shows historical data for the utilization rate, firm entry, and markups. The utilization rate is from the Federal Reserve Board, and the net entry data is from the Statistics of U.S. Businesses at the U.S. Census. The markup series is equal to the ratio of corporate profits to output, from the Bureau of Economic Analysis.\(^9\) The data are annual from 1989 through 2012, which is the frequency and availability of the net entry data. Each series is plotted relative to a linear time trend and normalized by its standard deviation. The series are strongly correlated (Table 1) and strongly procyclical.

\(^9\) See below for how markups are inferred from the ratio of corporate profits to output.
Here I explore the model’s ability to account for the joint movement of these series. To do so, I impose a functional form for the distribution of firms’ taste parameters (and hence firm size). It is well documented that the distribution of firm size is closely approximated by the Pareto distribution (e.g., Axtell 2001). Therefore let \( \theta_{jt} \) be distributed Pareto with a lower support equal to unity and shape parameter \( \alpha_t \). Equation (25) implies that only firms with \( \theta_{jt} \geq 2\sqrt{N_t \gamma b} \equiv \kappa_t \) survive.

**Utilization, Net Entry, and Markups.** Define aggregate utilization as

\[
U_t = \frac{1}{J^*_t} \int_{J^*_t} q_{jt}/\bar{q}_t,
\]

and let \( \bar{\theta} \) be the value of the firm taste parameter such that output is exactly equal to capacity, \( q_{jt} = \bar{q} \). Then the utilization rate is equal to unity for all firms with \( \theta_{jt} \geq \bar{\theta} \) and it equals \( q_{jt}/\bar{q} = \theta_{jt}/\bar{\theta} \) for firms with \( \theta_{jt} < \bar{\theta} \). The aggregate utilization rate can therefore be written as

\[
U_t = \frac{1}{\bar{\theta}} \int_{\bar{\theta}}^\infty \theta_{jt} f(\theta_{jt}) d\theta_{jt} + \int_{\bar{\theta}}^\infty f(\theta_{jt}) d\theta_{jt}.
\]

Under the assumed distribution of taste parameters, the aggregate utilization rate is

\[
U_t = \kappa^\alpha_t \left[ \frac{\alpha_t}{\bar{\theta}} \kappa^\alpha_t + \frac{1}{\alpha_t - 1} \left( \kappa^{\alpha_t + 1} - \bar{\theta}^{\alpha_t + 1} \right) + (\bar{\theta}^{\alpha_t}) \right].
\]

The utilization rate is decreasing in the shape parameter. As \( \alpha \) decreases toward unity, more firms have high demand and high capacity utilization.

Given the distribution of taste parameters, the mass of surviving firms is

\[
J^*_t = \kappa^{-\alpha_t},
\]

which implies that firm entry is

\[
NE_t = \kappa^{-\alpha_t} - \kappa^{-\alpha_{t-1}}.
\]

Markups are equal to prices for firms that are below capacity. Since we do not directly observe markups in the data, we must instead derive an alternative statistic that is observed in the data. Nekarda and Ramey (2013) propose a measure of markups based on the labor share of income. In my model, the labor share depends on workers’ bargaining power and need not co-move with markups. As an alternative statistic, which is closely related to that in Nekarda and Ramey (2013), consider the ratio of profits (available from the BEA) to output. For a firm in my model, this ratio is proportional to the taste parameter, \( \frac{n_{jt}}{q_{jt}} \sim \theta_{jt} \), and therefore co-moves with the firm’s markup. The ratio of aggregate profits to aggregate output (the ratio that corresponds to the data available from the BEA) is
\[
\frac{\Pi_t}{Q_t} = C \left(1 - \Psi \right) \frac{\int f(\theta) \Pi(\theta) f(\theta) d\theta}{\int f(\theta) q(\theta) f(\theta) d\theta}
\]

\[
= C \left(1 - \Psi \right) \left[\frac{\alpha_t}{2 - \alpha_t} \left(\bar{\theta}^{2-\alpha_t} - \kappa^{2-\alpha_t}\right) + \frac{1}{2\gamma} \left(\alpha_t - 1\right) \bar{\theta}^{2-\alpha_t} - \bar{\theta}^{1-\alpha_t}\right]
\]

where \(C\) is a constant.\(^{10}\)

The three key parameters in the model are listed below. \(\alpha\) is the aggregate demand parameter that summarizes the average preference for consumption. The remaining parameters are related to the firm’s cost structure and therefore represent technology parameters.

| \(\alpha_t\) | Shape parameter (determines \(E[\theta_j]\)) |
| \(\bar{\theta}_t\) | Threshold preference parameter for which firms produce at capacity level \(\bar{\theta}_t\) |
| \(\kappa\) | Lower bound on \(\theta_j\) for which firms survive (depends on technology parameter \(N\)) |

**Calibration.** Here I examine how well the model can fit the data by permitting only a single parameter, \(\alpha_t\) (which controls average consumer demand across firms), to vary across time. To pin down values for the constant parameters, I first choose the shape parameter \(\alpha_0\) to equal the point estimate of the distribution for firm size in Axtell (2001), \(\alpha_0 = 1.2\). I then choose \(\bar{\theta}\) to match the fraction of firms in the Quarterly Survey of Plant Capacity that are at capacity in the initial period, \(0.2 = \bar{\theta}^{-\alpha_0}\).\(^{11}\) I choose \(\kappa\) to match the utilization rate in the first period, given the chosen values for \(\alpha_0\) and \(\bar{\theta}\). With values of \(\kappa\) and \(\bar{\theta}\) pinned down, I then permit \(\alpha_t\) to vary at each point in time to match the time variation in the utilization rate.

The calibrated values of \(\bar{\theta}, \kappa,\) and \(\alpha_t\) yield model-based predictions for the evolution of net entry and the profit-to-output ratio. Figure 3 shows the model-implied series alongside the

\(^{10}\) Note that for firms with draws \(\theta_j > \bar{\theta}\), the price is given by (8) and \(\Pi_j \sim \frac{1}{2\gamma} \left(\theta_j - \bar{\theta}\right)^{\alpha_t - 1}\). Integrating profits over firms that are at capacity yields the second term in (28).

\(^{11}\) Approximately 80% of plant managers report “insufficient orders” as the primary reason for operating below the plant’s capacity output (Stahl and Morin 2013).
data. Table 1 shows the correlations from the model and from the data. In all cases, the model matches the strong positive comovement among the variables.

Figure 3: Time Series in the Model and the Data

<table>
<thead>
<tr>
<th>Table 1-Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Net Entry</td>
</tr>
<tr>
<td>Profit-to-Output Ratio</td>
</tr>
</tbody>
</table>

*Model Fit with Time Variation in Technology.* How well does the model account for the data when technology, rather than demand, is permitted to vary across time? Changes in technology can affect NMC firms in two ways. First, it can increase the amount of output at
which firms reach capacity (an increase in $\bar{q}_t$). Second, a technological improvement can
decrease the labor operating cost (for a given level of current capacity), which amounts to a
decrease in $N_t$ (and hence $\kappa_t$).

Figure 4 illustrates both of these situations for a NMC firm. An increase in efficiency
corresponds to a decrease in $N_t$ (and hence $\kappa_t$), which, by equation (27), causes firms to enter.
Firm entry also lowers aggregate utilization, as more firms with low utilization rates become
profitable. An increase in capacity corresponds to an increase in $\bar{\theta}_t$, which lowers utilization but
does not affect firm entry.

Table 2 summarizes the comparative statics arising from changes in demand and changes
in the two forms of technological improvement. It is clear that only changes in consumer demand
can generate the positive comovement in the data among utilization, net entry, and markups.
Even if both types of technology are permitted to vary with time (to match the time series of net
entry and utilization), the model cannot generate a procyclical profit-to-output ratio without
variation in demand. Figure 5 shows the model’s predictions when only technology is permitted
be time varying.

Figure 4: Two types of technological improvement
Table 2-Model Comparative Statics

<table>
<thead>
<tr>
<th>Increase in:</th>
<th>Output</th>
<th>Utilization</th>
<th>Net Entry</th>
<th>Profit-to-Output Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer preferences</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Technology (level of capacity)</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Technology (decrease in number of required tasks)</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 5: Correspondence between data and model with time variation in technology.