The Utilization Premium*

Fotis Grigoris and Gill Segal

July 22, 2019

Abstract

Firms that underutilize their capital are riskier. A quantitative model with production and flexible capacity utilization predicts a return spread between low and high utilization firms of above 5% p.a. Consistent with the model, we establish this utilization spread in the data as a novel empirical fact. Beyond the utilization premium, we show that a model without utilization yields many counterfactuals, such as investment's dispersion being too low, and its skewness bearing the wrong sign. Flexible utilization can addresses these moments by endogenously substituting large adjustment costs. Overall, utilization tightens the link between firms' production and valuation.

JEL classification: G12, E23, E32

Keywords: Production, Capacity, Utilization, Productivity, Asset Pricing

*Grigoris: Kenan-Flagler Business School, University of North Carolina at Chapel Hill, McColl Building, Chapel Hill, NC 27599, U.S.A. (e-mail: `Fotis_Grigoris@kenan-flagler.unc.edu`); Segal (corresponding author): Kenan-Flagler Business School, University of North Carolina at Chapel Hill, McColl Building, Chapel Hill, NC 27599, U.S.A. (e-mail: `Gill_Segal@kenan-flagler.unc.edu`). This paper benefited from comments and suggestions by Max Croce, Ric Colacito, Eric Ghysels, Christian Lundblad, Jun Li (discussant), Dimitris Papanikolaou and seminars participants at the 2019 MFA annual meeting and the University of North Carolina at Chapel Hill. All errors are our own.

Capacity utilization measures the degree to which a business is using its production potential. A flexible capacity utilization rate lets firms scale their production by choosing how much of their machinery to use. For example, instead of cutting production by selling machines, a manager can choose to keep some machines idle. At the *aggregate* level, existing studies show that the economy-wide utilization rate serves as a leading empirical predictor for macroeconomic growth, since an increase in utilization signals an increase in capital spending. However, the extent to which utilization quantitatively affects *firm* level investment and risk remains largely unexplored.

We incorporate the empirically motivated flexible utilization decision into a model with endogenous production. The model yields a previously undocumented prediction: lower utilization firms are riskier, and should earn higher expected returns. Quantitatively, the model predicts a spread of about 5% per annum between the returns of low and high utilization firms. Turning to the data, we establish this utilization premium as a novel empirical fact. We first confirm the sign and magnitude of the utilization spread in the data vis-à-vis the model. We also show empirically that utilization's explanatory power for risk premia is both incremental to other existing intensive margins, and orthogonal to other known production related spreads (e.g., premia associated with investment and hiring rates, book-to-market, productivity, and capital overhang).

Beyond the utilization premium, we show that flexible utilization has a key role for production models that jointly target risk premia. It allows matching asset-pricing and investment moments that fixed utilization models struggle with. Without utilization, the cross-sectional dispersion of investment and the value premium are too low, and investment's skewness bears a counterfactual negative sign. With flexible utilization these moments align with the data. For a model without utilization to get closer to the data for some of the aforementioned moments, higher value or higher dimensional capital adjustment cost parameters are required (e.g. piecewise adjustment costs). For instance, we show that the quadratic adjustment cost in a model with fixed utilization needs to be *twice* as large as that in a model with flexible utilization in order to match the value premium. Flexible utilization therefore offers a simple way to endogenize the impact of larger (or more) adjustment cost parameters using a micro-founded margin.

In the model, firms extend their production capacity by buying capital and decrease capacity by selling machines in the secondary market for capital. This market involves frictions. Specifically, the model features a fixed cost for capital disinvestment that makes selling machines a real option.[1] The key novel ingredient in our model is a variable capacity utilization rate that allows managers to control the extent to which installed capital is utilized. Increasing the utilization rate is not costless, as it makes capital depreciate faster. We summarize the theoretical intuition behind the asset-pricing and macroeconomic implications of the model below.

In an economy in which the capacity utilization rate is fixed, firms can only reduce the cyclicality of their payouts via investment decisions. If adjusting capital is costly, then the risk of each firm is determined entirely by the interaction between aggregate productivity and these capital adjustment costs.[2] However, when the capacity utilization rate is variable, firms have an additional mechanism by which to decrease the cyclicality of productivity shocks on payouts and mitigate risk.

To illustrate how capacity utilization is tied to firms' riskiness, consider an economy featuring convex and symmetric capital adjustment costs. A firm operating in a low productivity state has the incentive to reduce its capital stock, thereby exposing itself to potentially large adjustment costs. Simultaneously, the firm has an incentive to lower its capacity utilization rate as a substitute for this costly disinvestment. By lowering utilization the firm reduces its capital depreciation rate. This reduced depreciation not only conserves capital for future states that are more productive, but also reduces the adjustment cost of downsizing.[3] By similar logic, increasing utilization in good states reduces the adjustment cost for expanding capital by increasing depreciation. Thus, utilization and investment comove positively. This implies that both very high and very low utilization firms have high exposures to aggregate productivity. Both extremes reflect firms that incur high risk by greatly modifying their capital stock, and that simultaneously alter utilization to partially hedge their (dis)investment policies.

Two mechanisms in our model break the symmetry between high and low utilization firms. First, with a positive fixed adjustment cost, the ability to hedge bad states by downscaling is diminished, and disinvestment becomes a costly real option. As a result,

[1] These types of fixed costs feature prominently in both the macroeconomic and finance literatures (e.g., Bloom (2009), Carlson, Fisher, and Giammarino (2004), and Cooper (2006)).

[2] Firms that disinvest (invest) the most in low (high) aggregate productivity states are required to pay large capital adjustments costs. Consequently, since these firms are unable to fully absorb the impact of productivity shocks on their payouts, these firms are risky.

[3] In other words, lower utilization implies that the current depreciation, δ_t, falls. With quadratic adjustment frictions over net investment, the adjustment cost is proportional to the distance between i_t, the investment rate, and δ_t. As δ_t drops whenever i_t drops, the adjustment cost falls.

firms with moderately low levels of productivity substitute disinvestment by lowering utilization. These firms avoid the costly act of selling installed capital to "wait and see" if productivity recovers. Instead of selling capital, these firms temporarily downscale by reducing the utilization rate of their installed machines. As the friction in the market for selling capital is higher for these firms, they are riskier. Second, the model features a countercylical market price of risk (motivated by countercylical volatility). This implies that firms' whose valuations covary more with economic conditions during bad times command a risk premium. During bad states, in which the price of risk is high, low utilization firms are those with the higher productivity betas (riskier firms), and thus, tend to earn higher expected returns on average. Conversely, despite the fact that high utilization firms have high productivity betas in good states, this does not translate into a larger premium as the price of risk is low. Overall, our model yields a monotonic negative relation between utilization and risk premia, and predicts that the spread between the returns of low and high utilization firms is about 5% per annum.

We examine whether this quantitative prediction of the model that connects capacity utilization and risk premia holds in the data. To this end, we base our empirical analyses on publicly available utilization data that is published by the Federal Reserve Board (FRB). This data is available for a comprehensive cross-section of industries. We establish a novel empirical fact by sorting industries into portfolios on the basis of utilization rates: firms belonging to low capacity utilization industries earn average returns that are 5.7% per annum higher than those earned by firms belonging to high capacity utilization industries. Both the sign and the magnitude of this spread are consistent with our model. In line with our model's mechanism, we also document that each portfolio's exposure to aggregate productivity is monotonically decreasing in the portfolio's average utilization rate, regardless of which of three proxies for aggregate productivity we use. The differences in the productivity betas of the low and high utilization portfolios are also positive, and are economically and statistically significant.

Empirically, the relation between capacity utilization and risk premia extends beyond other well-established production-based characteristics. For instance, high and low utilization industries are indistinguishable in terms of size, profitability, productivity, and hiring. While low utilization industries typically have higher book-to-market ratios and lower investment rates, we show that the utilization premium is distinct from both the value and investment premia. We control for these effects using two methods:

conditional double sorts that control for either value or investment, and Fama and MacBeth (1973) regressions that control for value, investment, and an array of other production-related variables. In either case, the relation between utilization and risk premia remains negative and significant.

We conduct extensive checks to confirm the robustness of the utilization premium. For example, using different breakpoints, subsamples of industries, or a different timespan to form portfolios shows that low utilization industries still command a large risk premium. Moreover, although the FRB's utilization data is reported at the industry level, we demonstrate that the utilization spread does not reflect ex-ante heterogeneity between sectors in three ways. First, we double sort on the economic sector and show that an economically large utilization spread emerges among durable manufacturers only. Second, we proxy for firm-level utilization rates by projecting *industry-demeaned* utilization rates onto industry-level characteristics, and extrapolate the results of these projections using firm-level characteristics. The utilization premium also emerges when forming portfolios on these novel firm-level proxies for utilization. Finally, we show the utilization spread persists when we form portfolios using the growth rate of utilization, thereby eliminating any industry-specific fixed effects in utilization rates.

Our theoretical results suggest that flexible capacity utilization plays a key role for simultaneously matching investment and asset-pricing moments to the data. To illustrate how flexible utilization helps reconcile the model and data, we first consider the distribution of investment rates and risk premia in an economy with fixed utilization.

With fixed utilization, the cross-sectional dispersion and skewness of investment are less than half of their empirical magnitudes. Moreover, the time-series skewness of firm-level investment is negative, whereas this quantity is positive in the data. This happens because disinvestment in the model is a costly real option. During moderate economic slowdowns, firms "wait and see" if productivity will improve before opting to sell capital. Under fixed utilization, these firms do not alter their capital stocks and set their investment rates equal to the (constant) depreciation rate instead. Because a mass of waiting firms are lumped around the center of investment's distribution, the cross-section of investment rates is compressed. It features lower dispersion and cross-sectional skewness. If productivity further deteriorates, then investment drops sharply as these waiting firms shed their unproductive capital. These disinvestment jumps create the counterfactual negative sign for the time-series skewness of investment.

The distorted distribution of investment rates in the model with fixed utilization also has adverse implications for asset prices. As the cross-sectional distribution of investment becomes compressed, investment-related spreads, such as the value premium, shrink considerably. This is because the tails of investment's distribution feature fewer extreme outcomes (fewer firms with larger exposures to aggregate productivity).

When flexible utilization is introduced, firms can also respond to moderate drops in productivity by utilizing less capital. This causes depreciation to fall, and reduces the investment required to preserve the current capital stock. Since the natural (or preservation) rate of investment in this economy is time-varying, even firms that "wait and see" have to keep altering their investment rates to preserve their existing capital. Thus, the long periods of constant investment rates are eliminated. Time-varying depreciation rates that are (ex-post) heterogeneous between firms also imply that the cross-sectional dispersion of investment and risk premia also increase. This is because waiting firms are no longer massed at the same investment rate. Moreover, since firms utilize their machines more intensively in good times, depreciation increases in these periods. Thus, larger investments are needed to expand capital, causing the time-series and cross-sectional skewness of investment to rise, turn positive, and match the data.

Importantly, the problem of matching moments to the data without utilization is not simply alleviated by recalibrating the model. For instance, the diminished value premium in the model without utilization can be raised by increasing the convex capital adjustment costs. However, the adjustment costs required to match the value premium with fixed utilization are 100% higher than those with flexible utilization. The alternative calibration also has counterfactual implications for investment's dispersion and skewness. Both of these moments move further away from the data. More generally, we explain why flexible utilization allows to "save" on the degree of adjustment costs in the model. For the same adjustment cost parameter, flexible utilization yields more realistic investment, while keeping the amount of friction (risk) the same as under fixed utilization. This happens because of the endogenously time-varying depreciation rate.

Taken together, our empirical and theoretical results emphasize the economically important relation between capacity utilization, investment, and risk premia, and tighten the connection between firms' production dynamics and their valuations.

Related literature. The paper contributes to the literatures on the role of capacity utilization in RBC models, costly reversibility, and production-based asset pricing.

Our paper is tied to studies that examine the effects of time-varying capacity utilization in the macroeconomic literature. As a leading indicator, aggregate utilization data is studied extensively in relation to business cycle fluctuations. For instance, studies show how variable utilization is useful for matching macroeconomic growth dynamics to the data (e.g., Lucas (1970), Greenwood, Hercowitz, and Huffman (1988), Kydland and Prescott (1988), and Jaimovich and Rebelo (2009)). Additionally, Burnside and Eichenbaum (1996) show that variable utilization rates can propagate shocks over the business cycle, and amplify the impact of technology shocks. The macroeconomic literature utilizes several empirical proxies for utilization. Burnside, Eichenbaum, and Rebelo (1995) use electricity usage, while Basu, Fernald, and Kimball (2006) use hours per worker to proxy for all unobserved intensive margins. Similar to our empirical approach, Comin and Gertler (2006) use the FRB's measure of capacity utilization to study business cycle fluctuations over the medium-term.

In contrast to the macroeconomic literature, the relation between capacity utilization and asset prices has received considerably less attention. This is both despite the fact that capacity utilization is conceptually related to firm-level production decisions, and despite the fact that the FRB regularly reports granular data on the cross-section of utilization rates for various manufacturing and mining industries, and utilities.[4]

Of the small number of papers that also study capacity utilization in the context of asset pricing, most focus on aggregate asset-pricing moments. For instance, Garlappi and Song (2017b) include capacity utilization in a general equilibrium production-based asset pricing model and show that varying utilization is important to reconcile the market price of risk of investment-specific technology (IST) shocks with the data. Da, Huang, and Yun (2017) use industrial electricity usage as a proxy for utilization and find that higher electricity usage in the current period predicts lower stock market returns in the future. This latter result is broadly consistent with our utilization premium, but the findings in Da et al. (2017) pertain only to the time-series of aggregate market returns, and not to the cross-section of equities that we study.

Additionally, while the model in Cooper, Wu, and Gerard (2005) also includes capacity utilization, the authors' focus on explaining the value premium. Although the authors find a *qualitatively* negative relation between utilization and industry-

[4]While the U.S. manufacturing sector is of modest size, the sector still influences the macroeconomy to a large degree (Andreou, Gagliardini, Ghysels, and Rubin, 2019). Consequently, capacity utilization figures are routinely analyzed by both the Federal Reserve Bank (FRB) and other market participants.

level stock returns using OLS regressions in their empirical analysis of the book-to-market effect, we emphasize the *quantitative* contribution of capacity utilization to cross-sectional risk premia. We do this both theoretically, via a calibrated model, and empirically, by establishing a novel spread that is incremental to the book-to-market effect, as well as a host of other production-based characteristics.

The notion of costly reversibility, or the assumption that firms face higher costs to contract rather than expand their capital stocks, continues to influence research in macroeconomics and finance.[5] In macroeconomics, recent studies such as Bloom (2009) and Bloom, Floetotto, Jaimovich, Saporta-Eksten, and Terry (2018) combine costly reversibility and uncertainty shocks to explain the dynamics of real quantities over the business cycle. In finance, costly reversibility has become a standard feature in many models that seek to rationalize patterns in expected returns.

In an approach popularized by Zhang (2005), studies including Carlson et al. (2004) and Cooper (2006) explain the value premium by assuming that capital is at least partially irreversible. While the exact nature of costly reversibility differs between these studies, each employs a degree of capital inflexibility to match asset-pricing moments with the data. Recently, however, Clementi and Palazzo (2019) highlight that the capital adjustment frictions production-based models use to target risk premia may distort the distributions of model-implied investment rates. These distorted distributions feature too much investment inaction, and too little disinvestment, compared to the data. In this paper we show that flexible utilization provides a potential way to address this critique by Clementi and Palazzo (2019). Specifically, our model with flexible utilization simultaneously produces realistic dispersions in investment rates and risk premia jointly. A similar fit to the data cannot be achieved in a model with fixed utilization.

More broadly, our paper is related to asset-pricing studies that connect production-related firm characteristics to expected returns (e.g., Zhang (2005), Belo and Lin (2012), Jones and Tuzel (2013)). Consequently, the conceptual relation between capacity utilization and asset prices is of particular interest to this growing literature that examines the joint dynamics of firm-level investment and dispersion in risk premia.

Prior studies in this literature include Belo, Lin, and Bazdresch (2014), who study the impact of labor market frictions on asset prices and find that firms with low hiring

[5] While the literature on costly reversibility is voluminous, some key studies include Dixit and Pindyck (1994), Abel and Eberly (1996), and Cooper and Haltiwanger (2006a).

rates earn higher returns. We show empirically that there is no statistical difference in hiring rates between low and high utilization industries. Similarly, Imrohoroglu and Tuzel (2014) examine firm-level total factor productivity (TFP), and theoretically and empirically show that low TFP firms earn a significant productivity premium. This is important for our study because TFP and capacity utilization are linked by the fact that TFP can be decomposed into three distinct components: utilization, markups, and technology. While utilization is a component of TFP, we show empirically that most of the productivity premium stems from the markups and technology components of TFP. That is, controlling for capacity utilization, the productivity premium persists. Conversely, the utilization spread also persists after controlling for TFP.

In a recent study, Aretz and Pope (2018) estimate firm-level capacity overhang, or the difference between a firm's installed and optimal production capacity, and show that overhang has sizable implications for cross-sectional risk premia. Although capacity utilization and capacity overhang are conceptually similar, we show that utilization and overhang result in theoretically and empirically distinct spreads. In particular, both portfolio double sorts and Fama and MacBeth (1973) regressions show that the utilization spread survives controlling for overhang, and vice versa.

We add to this literature and focus on the utilization rate of productive units, such as machines, since flexible utilization provides managers with a major degree of freedom with which to smooth the cyclicality of their profits. The quantitative impacts of this utilization margin on both a firm's risk and its cost of capital remain unexplored. In this paper we address these open questions and demonstrate that capacity utilization is an important determinant of expected returns.

The rest of the paper is organized as follows. Section 1 outlines our model with flexible capacity utilization. Section 2 examines the relation between utilization and risk premia through the lens of the model, and tests the model's key predictions in the data. Section 3 explores the theoretical implications of flexible utilization for firm-level investment dynamics and adjustment costs. Section 4 provides concluding remarks.

1 The model

This section describes a production-based model featuring endogenous capacity utilization. The economy consists of competitive firms that produce a homogeneous good using capital and labor. Firms face convex adjustment costs when altering their capital stocks and time-varying capital depreciation rates. When firms choose to increase utilization, their capital depreciates faster. In addition, firms pay a fixed cost to reduce their capital. This makes disinvestment a real option. Firms may freely adjust labor and, following Belo et al. (2014), pay wages according to an exogenous and stochastic wage function. Risk in the economy originates from persistent aggregate productivity shocks. Finally, the stochastic discount factor is specified exogenously in the spirit of Berk, Green, and Naik (1999) and Zhang (2005).

1.1 Economic environment

Technology. The economy is populated by a continuum of firms that produce a homogeneous good using capital ($K_{i,t}$) and labor ($L_{i,t}$). All firms are subject to the same aggregate productivity shocks, and each firm is subject to its own idiosyncratic firm-specific productivity shocks. The production function for firm i is given by:

$$Y_{i,t} = \exp(x_t + z_{i,t})\big(u_{i,t}K_{i,t}\big)^{\theta\alpha_K}\big(L_{i,t}\big)^{\theta\alpha_L}, \tag{1}$$

where $\alpha_K \in (0,1)$ and $\alpha_L \in (0,1)$ control the shares of capital and labor in the production function, respectively, and $\alpha_K + \alpha_L = 1$. The parameter $\theta \in (0,1]$ sets the degree of returns to scale associated with the production function.

The control variable $u_{i,t} > 0$ represents the capacity utilization rate of the firm. This variable controls the intensity with which the firm utilizes its capital. In other words, the presence of $u_{i,t}$ in equation (1) provides firms with the flexibility to scale production in response to productivity shocks, while keeping the capital stock fixed.

Each firm's capital stock evolves over time according to the following law of motion:

$$K_{i,t+1} = \big(1 - \delta(u_{i,t})\big)K_{i,t} + I_{i,t}. \tag{2}$$

Here $I_{i,t}$ represents gross investment and $\delta(u_{i,t})$ is the depreciation rate of the firm's capital stock. The depreciation rate depends on the degree to which capital is utilized at time t, and we assume that $\delta'(u_{i,t}) > 0$. Intuitively, this means that if the firm chooses to employ more machines in production, its capital depreciates at a faster rate.

Productivity. Aggregate productivity is denoted by x_t and evolves over time as a stationary AR(1) process:

$$x_{t+1} = \rho_x x_t + \varepsilon^x_{t+1}, \tag{3}$$

where $\varepsilon^x_{t+1} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_x^2)$. The idiosyncratic productivity process for firm i is denoted by $z_{i,t}$ and also evolves according to a stationary AR(1) process given by:

$$z_{i,t+1} = \bar{z}(1-\rho_z) + \rho_z z_{i,t} + \varepsilon^z_{i,t+1}, \tag{4}$$

where $\varepsilon^z_{i,t+1} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2)$. We assume that $\varepsilon^z_{i,t+1}$ and $\varepsilon^z_{j,t+1}$ are uncorrelated for $i \neq j$ and that idiosyncratic shocks are uncorrelated with ε^x_{t+1}. $\bar{z}$ is a scaling parameter.

Depreciation, adjustment costs, and wages. Production is subject to three different costs: variable capital depreciation rates, capital adjustment costs, and wages.

We follow Jaimovich and Rebelo (2009) and Garlappi and Song (2017b) and specify a depreciation function that features a constant elasticity of marginal depreciation with respect to capacity utilization as follows:

$$\delta(u_{i,t}) = \delta_k + \delta_u \left[\frac{u_{i,t}^{1+\lambda} - 1}{1+\lambda} \right]. \tag{5}$$

Here, δ_k represents the depreciation rate when $u_{i,t} = 1$ (the model's steady state). δ_u measures the additional cost of capital depreciation as the utilization rate is increased. The parameter λ controls the elasticity of depreciation with respect to utilization and determines how costly it for a firm to alter its utilization rate in response to exogenous shocks. Holding all else constant, larger values of λ make increasing the capacity utilization rate more costly and ensures that firms choose a finite level of utilization.

Capital adjustment costs are given by the following function:

$$G_{i,t} \equiv G(I_{,t}, K_{i,t}, u_{i,t}) = \frac{\phi}{2}\left(\frac{I_{i,t}}{K_{i,t}} - \delta(u_{i,t})\right)^2 K_{i,t} + f\mathbf{1}_{\left\{\left(\frac{I_{i,t}}{K_{i,t}} - \delta(u_{i,t})\right) < 0\right\}} K_{i,t}, \tag{6}$$

where $\phi > 0$, $f > 0$, and $\mathbf{1}_{\{.\}}$ is an indicator function that takes on a value of one when a firm reduces capacity. The adjustment cost function features two components. The first term is the standard neoclassical convex cost governed by ϕ. The second term is a fixed cost associated with disinvestment only and is governed by f. This fixed cost reflects frictions in the secondary market for capital, such as the cost of matching with a counterparty (buyer). We introduce the second term for two reasons. First, structural estimations of adjustment cost functions highlight the existence of non-convex adjustment costs (e.g., Cooper and Haltiwanger (2006a)). The fixed cost also makes disinvestment a real option. This component is crucial to feature a negative relation between investment and uncertainty (e.g., Bloom (2009)).[6] Second, the fixed

[6]Put differently, $f > 0$ is crucial for the production model to match the sign of investment's response to uncertainty shocks. While our model does not exhibit stochastic volatility, we confirm that when $f = 0$, the stochastic mean of investment rises in the model when σ_z increases. This counterfactual (i.e., $\partial i/\partial\sigma_z > 0$) does not happen when $f > 0$ because of the standard real option

cost of disinvestment is motivated by the large literature that emphasizes the relatively larger costs associated with reducing capacity rather than expanding capacity, both for investment moments and countercyclical risk premia (e.g., Zhang (2005)).[7] It is worth noting that our adjustment cost specification in equation (6) is parsimonious compared to other specifications in the literature (e.g., Cooper and Haltiwanger (2006b), Belo and Lin (2012)) as it only features two free parameters, ϕ and f.

Firms face a perfectly elastic supply of labor at a given equilibrium real wage rate as per Belo et al. (2014). We follow Jones and Tuzel (2013) and Imrohoroglu and Tuzel (2014) by assuming that the wage rate is positive and increasing in the level of aggregate productivity. Specifically, the wage rate, W_t, is given by:

$$W_t = \exp(\omega x_t), \tag{7}$$

where $\omega \in (0, 1)$ measures the sensitivity of wages to aggregate productivity.

Stochastic discount factor (SDF). In line with Berk et al. (1999) and Zhang (2005) we do not explicitly model the consumer's problem. Instead, we assume that the pricing kernel of the household is given by:

$$\ln(M_{t+1}) = \ln(\beta) - \gamma_t \varepsilon^x_{t+1} - \frac{1}{2}\gamma_t^2 \sigma_x^2, \text{ where } \ln(\gamma_t) = \gamma_0 + \gamma_1 x_t. \tag{8}$$

Here, $0 < \beta < 1$, $\gamma_0 > 0$, and $\gamma_1 < 0$ are constants. This form of the SDF is drawn from Jones and Tuzel (2013) and Imrohoroglu and Tuzel (2014), and is as an adaptation of the SDF assumed by Zhang (2005). Two key features of this SDF are worth noting. First, the volatility of the SDF is time-varying and driven by γ_t. This volatility increases during economic contractions, and results in a countercyclical price of risk.[8] Second, the $-\frac{1}{2}\gamma_t^2\sigma_x^2$ term in the SDF implies that the risk-free rate is constant and equal to $-\ln(\beta)$ in each period. Thus, γ_0 and γ_1 only affect the market risk premium.

Firm value, risk, and expected returns. Firms are all-equity financed. The dividend to the shareholders in period t is given by:

$$D_{i,t} = Y_{i,t} - I_{i,t} - G_{i,t} - L_{i,t}W_t. \tag{9}$$

logic.

[7] Untabulated analyses confirm that our results remain materially unchanged when we add a fixed cost for investing to the adjustment cost function. Specifically, we set $G = \frac{\phi}{2}\left(\frac{I}{K} - \delta(u)\right)^2 K + f^- \mathbf{1}\left\{\left(\frac{I}{K} - \delta(u)\right) < 0\right\}K + f^+ \mathbf{1}\left\{\left(\frac{I}{K} - \delta(u)\right) > 0\right\}K$, with $f^- > f^+$. While this alternative function also captures the notion of costly reversibility, it includes an extra degree of freedom: f^+. Our goal is to demonstrate that utilization can produce a close fit to the data without relying on a high-dimensional adjustment cost function. Consequently, we use equation (6) without loss of generality.

[8] An economic mechanism that could lead to a countercyclical price of risk is, for example, time-varying risk aversion as in Campbell and Cochrane (1999).

In each period, each firm chooses $\{I_{i,t}, L_{i,t}, u_{i,t}\}$ to maximize firm value:

$$V_{i,t} = \max_{\{I_{i,t}, L_{i,t}, u_{i,t}\}} D_{i,t} + \mathrm{E}_t\left[\sum_{j=1}^{\infty} M_{t,t+j} D_{i,t+j}\right], \tag{10}$$

subject to equations (1) – (9). In the equation above, $M_{t,t+j}$ represents the SDF between times t and $t+j$, and $V_{i,t}$ denotes the cum-divided value of firm i at time t. Finally, the gross stock return of firm i is given by:

$$R^S_{i,t+1} = \frac{V_{i,t+1}}{V_{i,t} - D_{i,t}} \tag{11}$$

Equilibrium. The first-order conditions for the firm's value maximization problem lead to the following pricing equation:

$$1 = \mathrm{E}_t\left[M_{t,t+1} R^I_{i,t+1}\right], \tag{12}$$

where $R^I_{i,t+1}$ denotes the returns to investment. This Euler equation determines each firm's optimal choice of $I_{i,t}$ by ensuring that the marginal cost of investment equals the discounted marginal benefit of purchasing new capital. Labor allocation, $L_{i,t}$, is a static decision set such that the marginal product of labor equals the wage. Moreover, in (partial) equilibrium, (i) firms' investment and utilization policies maximize program (10) given the SDF, and (ii) firms' valuation and return on investment satisfy equations (10) and (12) given their optimal policies.

1.2 Calibration and solution method

The model is calibrated and solved at the annual frequency. Table 1 presents the set of parameter values used in the model's solution. The first set of parameters governs the dynamics of the exogenous shocks that firms face and also controls the SDF. The second set of parameters controls the production of firms.

Stochastic processes and SDF. We base our annualized values of ρ_x and σ_x, the parameters governing the aggregate productivity process, on the quarterly estimates of these parameters reported by King and Rebelo (1999). We fix ρ_x at 0.922 and σ_x at 0.014. We set ρ_z to 0.60 and σ_z to 0.30 to match the unconditional volatility of firm-level productivity reported by Imrohoroglu and Tuzel (2014). The long-run average level of idiosyncratic productivity ($\bar{z}$) is a scaling variable set so that the long-run amount of firm-level capital in the economy is one. This implies that $\bar{z} = -0.163$.

We choose β, γ_0, and γ_1, the parameters governing the SDF, by matching the average annual real risk-free rate, and the average annual volatility and excess returns of the value-weighted market portfolio, respectively. We set the discount factor, β, to 0.988 to produce an average real risk-free rate of 1.2% per annum. γ_0 and γ_1 are set to 3.375 and -8.8, respectively. These parameter result in a value-weighted equity premium of 5.39% per annum and a market return volatility of 20.89% per annum.

Technology. We fix α_K and α_L at 0.333 and 0.667, respectively. We set θ, the parameter governing the degree of returns to scale in the production function, to 0.95 since slightly decreasing returns to scale are important to keep firm size bounded. δ_k, the average capital depreciation rate, is set to 8% per annum and δ_u, the incremental depreciation rate, is chosen such that utilization is equal to one in the model's deterministic steady state. λ, the parameter governing the elasticity of depreciation to utilization, is chosen to match the average volatility of the aggregate utilization rate. Setting λ to 3 produces an average annual volatility of utilization of 4.15% per annum in the model compared to an annual volatility of 4.09% per annum in the data. We set ω, the wage sensitivity to aggregate productivity, to 0.20. This value is comparable to both Imrohoroglu and Tuzel (2014) and Jones and Tuzel (2013), and is consistent with the empirical correlation between real GDP growth and wage growth.

We calibrate ϕ, the degree of convex capital adjustment costs, to match the volatility of investment in the data. Setting this parameter to 1.5 results in a model-implied annual volatility of investment of 0.14. This is identical to the empirical volatility of firm-level investment during our sample period. Finally, we set f, the parameter governing the fixed cost of disinvestment, to 0.028 to match the first-order autocorrelation of firm-level investment. The value of this correlation is 0.58 (0.52) in the model (data).

We solve the model numerically using the value function iteration method, as described in Section OA.5 of the Online Appendix. We compute model-implied moments by simulating an economy featuring $1,000$ firms for $40,000$ periods (years).

2 The utilization premium: theory and evidence

This section examines the relation between capacity utilization and cross-sectional risk premia both through the lens of the production model presented in Section 1 and empirically. Section 2.1 lays out the novel asset-pricing predictions of the model. Section 2.2 tests these predictions using industry-level utilization data reported by the FRB. Section 2.3 shows that the relation between utilization and asset prices is neither confounded by the risk premia associated with other production-based characteristics, such as book-to-market, nor driven by industry-specific effects, such as the durability premium. Section 2.3 also develops a novel empirical proxy for firm-level utilization rates, and present firm-level evidence in support of the model.

2.1 Model predictions for the cross-section of returns

2.1.1 Investment and return moments: model versus data

Table 2 compares the fit of the model to the data along dimensions related to distribution of firm-level investment rates, the aggregate utilization rate, and asset-pricing quantities. The main takeaway from this table is that our benchmark model simultaneously produces a realistic distribution of firm-level investment rates and sizable risk premia. Below, we illustrate our model's fit to the data along each dimension.[9]

Time-series of investment rates. Panel A of Table 2 shows that the model-implied volatility and first-order autocorrelation of firm-level investment rates are 14% and 0.58, respectively. These figures are very close to their empirical counterparts since the two capital adjustment cost parameters are set to fit these moments. Additionally, the model also produces realistic estimates for two untargeted moments: the time-series skewness of investment rates (0.66 in the model versus 0.67 in the data) and the second-order autocorrelation of investment (0.38 in the model versus 0.26 in the data).

Cross-section of investment rates. Our model produces a realistic cross-sectional distribution of investment rates *without* targeting any dispersion-related moments of investment. Panel A of Table 2 shows that the dispersion of investment rates is 0.11 in the model versus 0.16 in the data. Similarly, the inter-decile range of investment is 0.22 in the model and 0.32 in the data. We also document that our model produces positively skewed firm-level investment rates that are consistent with the data. The skewness of investment is 1.19 in the model versus 1.89 in the data.

Aggregate capacity utilization rate. Finally, Panel A of Table 2 shows the volatility of the aggregate utilization rate is just over 4% in both the model and the data. This close fit is achieved by calibrating the sensitivity of utilization to depreciation to match this volatility. The model also produces a realistic, and fairly persistent,

[9]Section 2.2.1 provides details about the data used to construct the empirical estimates for the table.

autocorrelation of utilization (0.7 in the data versus 0.9 in the model).

Aggregate asset-pricing moments. Panel B of Table 2 indicates the model-implied annual risk-free rate and equity premium are 1.2% and 5.4%, respectively. The volatility of excess market returns in the model is 20.9%. These three moments are calibrated to match the data. The model also produces a slightly negative autocorrelation of excess market returns that is close to its empirical counterpart of -0.05. Here, model-implied returns are multiplied by 5/3 to account for financial leverage.

Cross-sectional risk premia. Lastly, Panel B of Table 2 demonstrates that our model is quantitatively reliable in regards to cross-sectional risk premia. This is evidenced by the fact that the model produces book-to-market and investment spreads that align with the data. The value premium in the model is 3.76% per annum, whereas this spread is 3.71% per annum in the data. The model-implied investment premium of 4.27% per annum is also close to its empirical magnitude of 3.7% per annum. The success of the model along this dimension is achieved without calibrating any model parameters to match either of the aforementioned spreads.

2.1.2 Model-implied capacity utilization spread

In this section we examine the relation between capacity utilization and risk premia in our model. Using model simulated data, we sort the cross-section of firms into portfolios on the basis of realized capacity utilization rates. Specifically, at each point in time t, the low (high) capacity utilization portfolio includes all firms whose utilization rates were at or below (above) the 10^{th} (90^{th}) percentile of the cross-sectional distribution of utilization rates at time $t-1$. We define the *capacity utilization spread* as the return differential between these low and high capacity utilization portfolios.

Table 3 shows the average returns associated with each utilization-sorted portfolio, on both a value- and equal-weighted basis, as well as the average capacity utilization spread in our simulated economy.[10] The model-implied relation between capacity utilization and average stock returns is both negative and monotonic. In the model, low

[10] To be consistent with our empirical results, discussed in Section 2.2, the main results in this table are obtained by simulating our economy 500 times and reporting the average of each moment across these simulations. Each simulation features a cross-section of 4,000 firms, roughly matching the number of firms underlying our empirical analyses, and spans 50 periods (years), approximating our sample period that ranges from 1967 to 2015. The table also reports population moments from one simulation of 1,000 firms over 40,000 periods (years) in square brackets.

utilization firms earn a larger risk premium. On a value-weighted basis, the portfolio of low utilization firms in the model earns an average return of 9.86% per annum, whereas the high utilization portfolio earns an average return of 4.64% per annum. Thus, the model predicts that the value-weighted utilization spread is about 5.2% per annum. Given the model features a continuum of firms, the equal-weighted returns are similar.

The utilization spread is risk-based. Table 3 reports the exposure of each value-weighted utilization portfolio to aggregate productivity, and the spread between these productivity betas. The model-implied exposure of each portfolio to aggregative productivity is decreasing with the portfolio's utilization rate, and the pattern in these productivity betas is consistent with the pattern of expected returns.

Finally, we examine whether the unconditional CAPM explains the utilization spread. CAPM alphas, obtained via projections in the model, are shown in Table 3. The CAPM alpha associated with the value-weighted utilization spread is approximately 4.2% per annum, with the 90% confidence interval ranging from 2.77% to 5.98%. Despite featuring only a single aggregate shock, the unconditional CAPM cannot fully account for the utilization spread.[11]

Taken together, Table 3 implies three key predictions. First, average portfolio returns are decreasing in the capacity utilization rate. Quantitatively, the utilization spread is about 5.2% per annum. Second, the exposures of the utilization portfolios to aggregate productivity are decreasing with the utilization rate. The spread in productivity betas is close to 0.20. Third, the unconditional CAPM alpha associated with the utilization spread is non-zero. Section 2.1.3 discusses the intuition behind these predictions, while Section 2.2 tests these novel predictions in the data.

Double sort on book-to-market. Despite featuring only one aggregate shock, our model predicts a distinction between the utilization premium and the well-known value premium. To show this, we perform a conditional double-sort analysis using simulated model data. We first sort the cross-section into portfolios based on book-to-market. Then, within each book-to-market bucket, we sort firms based on utilization. Table OA.1.1 in the Online Appendix reports the results of this conditional portfolio double sort. While the two spreads are related, the table shows that the utilization

[11]The CAPM fails because the CAPM beta of each capacity utilization portfolio varies over time in a way that is correlated with aggregate productivity. In particular, the low capacity utilization portfolio tends to have a high beta during periods of poor aggregate productivity when its constituent firms tend to scale down their output by lowering utilization and reducing their capital stocks.

spread is positive and economically significant *within* each book-to-market portfolio. In Section OA.4.1 of the Online Appendix, we perform the same double sort in the data and obtain a quantitatively similar result. Section OA.1.1 of the Online Appendix discuss the rationale behind this result.

2.1.3 Economic rationale for the capacity utilization spread

The mechanism relating capacity utilization to risk premia in the model hinges on three ingredients: (1) a quadratic capital adjustment cost ($\phi > 0$), (2) a fixed cost of disinvestment ($f > 0$), and (3) a countercyclical market price of risk ($\gamma_1 < 0$). Firms in the model are risky because they can neither costlessly (nor fully) adjust their capital stock $K_{i,t}$ in response to productivity shocks. However, a flexible utilization rate, $u_{i,t}$, provides firms with a mechanism to reduce these capital frictions, and the cyclicality of their dividends. Consequently, the utilization rate is inherently tied a firm's risk and its expected returns. We illustrate this logic by shutting down ingredients (1)–(3) in our economy, as well as utilization, and explaining the role of each ingredient in turn.

In a frictionless economy, all firms facing economic downturns have an incentive to reduce investment in response to the lower expected marginal product of capital in future periods. This is because productivity shocks in the model are persistent. At the same time, each firm also has the incentive to lower its utilization rate. The marginal output gained by increasing utilization, $MPU_{i,t} = \theta\alpha_k Y_{i,t} u_{i,t}^{-1}$, in bad times does not compensate the firm for the extra capital depreciation cost incurred, $\delta'(u_{i,t})$. Consequently, firms respond to economic downturns by reducing investment and utilization. In contrast, periods featuring high aggregate productivity result in the opposite policies. Thus, investment and utilization comove positively (i.e, $corr(u_{i,t}, i_{i,t}) > 0$).

If we reintroduce ingredient (1), the quadratic capital adjustment cost, to the economy while fixing utilization ($\lambda \to \infty$), then the risk of each firm is determined entirely by the interaction between firm-level investment and capital adjustment costs, as implied by Tobin's q (as $\beta_{i,t} = \frac{\partial V_{i,t}}{\partial \varepsilon_{x,t}} \approx \frac{\partial}{\partial \varepsilon_{x,t}} q_{i,t}(i_{i,t})\dot{K}_i$ and $q'(i_{i,t}) > 0$). Consequently, firms that make large (dis)investments in capital are required to pay large capital adjustments costs, and are unable to fully absorb the impact of productivity shocks on their dividends. This means that the valuations of investing (disinvesting) firms covary more with aggregate productivity in good (bad) times. Thus, both high and low investment rate firms are risky (i.e., $\beta_{i,t} \uparrow$ if either $i_{i,t} \uparrow$ and $x_t \uparrow$, or $i_{i,t} \downarrow$ and $x_t \downarrow$).

When the capacity utilization rate in the economy becomes variable, each firm has an additional mechanism to mitigate the effects of productivity shocks on the cyclicality of its dividends. Consider a firm facing an economic downturn. While this firm still has the incentive to reduce its capital stock, thereby exposing itself to potentially large quadratic capital adjustment costs, the firm also has the incentive to lower its utilization rate for two reasons. First, by lowering its utilization rate, the firm can reduce its capital depreciation rate. This conserves capital for more productive states in the future (i.e., $u_{i,t} \downarrow \Rightarrow \delta(u_{i,t}) \downarrow \Rightarrow K_{i,t+1} \uparrow$). Second, lowering the utilization rate also decreases the firm's natural rate of investment (the investment required to maintain its current capital). By lowering its natural rate of investment, the firm can pay a lower quadratic adjustment cost to disinvest capital. This is because adjustment costs are proportion to the distance between the investment rate, $i_{i,t}$, and the depreciation rate, $\delta(u_{i,t})$ (i.e., $\phi(i_{i,t} - \delta(u_{i,t}))^2 \downarrow$ if $\delta(u_{i,t}) \downarrow$ whenever $i_{i,t} \downarrow$).

The incentives above create complementarity between a firm's need to disinvest and low utilization. Therefore, during economic downturns, low utilization firms are risky because they face large downscaling costs which they attempt to vent by lowering utilization (i.e., $\beta_{i,t} \uparrow$ if $u_{i,t} \downarrow$ and $x_t \downarrow$ by $corr(u_{i,t}, i_{i,t}) > 0$). The converse holds true for high utilization firms during periods of high aggregate productivity. Hence, both very high and very low utilization firms are risky, depending on the state of the world.

We break the risk symmetry between high and low utilization firms by introducing ingredients (2) and (3). First, when we enrich the model with ingredient (2), the fixed cost of disinvestment, we introduce a higher friction to disinvest. Reducing capital becomes a costly real option. Consequently, firms facing a moderate drop in productivity do not disinvest immediately. Instead, they "wait and see" if productivity improves before exercising the costly disinvestment options. Simultaneously, these waiting firm substitute disinvestment by lowering utilization (further). These waiting firms opt to temporarily scale down their production by reducing the utilization of their installed machines instead of selling installed capital. Since the friction in the market for selling capital are even greater for low utilization firms, their betas in bad states of the world exceeds the betas of high utilization firms in good states (i.e., $\beta_{U_L,X_L} > \beta_{U_H,X_H}$, where $X_L(X_H)$ is low (high) productivity, and $U_L(U_H)$ is a low (high) utilization firm).

The second mechanism that breaks the symmetry is ingredient (3), the countercyclical market price of risk. Since the market price of risk is higher in low aggregate productivity states (i.e., $\gamma_t'(x_t) < 0$), the firms whose returns covary more with economic conditions during bad times command a larger risk premium. As discussed above, low utilization firms are riskier (have higher betas) during economic downturns. Since these states feature a higher market price of risk, low utilization firms earn a risk premium (i.e., if $x_t \downarrow$ and $u_{i,t} \downarrow$, then $E[R^e_{i,t+1}] \approx \beta_{i,t}\gamma(x_t) \uparrow$ because $\beta_{i,t} \uparrow$ and $\gamma(x_t) \uparrow$). In contrast, high utilization firms earn a much lower premium (in both good and bad states). High utilization firms have greater exposures ($\beta_{i,t}$) to aggregate risk only in good times. Since the market price of risk is very small in these periods, the risk premium of high utilization firms is also small. Overall, this logic implies that low utilization firms should earn larger risk premia, consistent with Table 3.

Combined, ingredients (2) and (3) yield a monotonic relation between utilization and risk premia. The real option channel impacts (mostly) firms with moderate idiosyncratic productivity that find it optimal to leave their capital unaltered in bad states. As a result, medium utilization firms are considerably riskier than high utilization firms. The countercylical price of risk impacts (mostly) firms with lower idiosyncratic productivity that have a high beta. As a result, low utilization firms are riskier than both medium or high utilization firms. Moreover, a very low price of risk in good states implies that the risk associated with high utilization firms is not translated into a higher risk premium in good times.

Section OA.1 of the Online Appendix provides further details regarding this intuition for the utilization spread, and our model in general. Section OA.1.2 illustrates the intuition above by perturbing the model parameters related to the capital adjustment costs and the market price of risk. Section OA.1.3 discusses the assumptions of our model in more detail. Finally, Section OA.1.4 outlines how extending the model to feature investment-specific technology shocks could interact with the utilization spread.

2.2 Empirical evidence from the cross-section of returns

In this section we empirically examine the key novel prediction of our model: the existence of a capacity utilization premium in the cross-section of equity returns. We base our empirical analyses on publicly available capacity utilization data that is published by the FRB and discussed in Section 2.2.1. Sections 2.2.3 and 2.2.4 contain the main empirical tests of our predictions, while Section 2.3 empirically investigates whether the utilization premium is distinct from other premia related to production-based characteristics. We report extensive robustness tests confirming the novel relation between capacity utilization and risk premia in Section OA.4 of the Online Appendix.

2.2.1 Data

Capacity utilization data. We obtain industry-level capacity utilization data from the FRB's monthly report on Industrial Production and Capacity Utilization (report G.17) that releases publicly available estimates of capacity utilization for a cross-section of industries that cover the manufacturing and mining sectors, as well as utilities. The FRB uses this data to quantify how effectively different industries are utilizing factors of production and to assess inflationary pressures (e.g., Corrado and Mattey (1997) and Board of Govenors of the Federal Reserve System (2014)). A major advantage of this FRB data is that it provides a measure of utilization that is available at a much higher frequency than estimates elicited from low-frequency accounting data. The capacity utilization rate ($CU_{i,t}$) of industry i at time t is given by:

$$CU_{i,t} = \frac{IP_{i,t}}{Capacity_{i,t}}. \tag{13}$$

Here, $IP_{i,t}$ is the actual output of the industry, measured by seasonally-adjusted industrial production, and $Capacity_{i,t}$ is the FRB's estimate of the industry's sustainable maximal output at time t. The capacity estimate for most industries is derived from the Quarterly Survey of Plant Capacity Utilization conducted by the U.S. Census Bureau.

In the benchmark case, our cross-section encompasses 45 industries that feature a mix of durable manufacturers (18 industries), nondurable manufacturers (17 industries) and mining and utilities (10 industries). Due to data availability, the time period of our benchmark analysis ranges from from January 1967 to December 2015. The average utilization rate across all industries is roughly 80%. The unconditional moments of the mean, variance and autocorrelation of the utilization rate are similar across different sectors. However, the relative ranking of industries in terms of utilization rates varies substantially over time. We provide further details on the sample composition, including summary statistics, in Section OA.2 of the Online Appendix.

Returns data. Monthly stock return data are taken from the Center for Research in Security Prices (CRSP), and accounting data are taken from the CRSP/Compustat Merged Fundamentals Annual file. We obtain returns for portfolios sorted on key characteristics, such as size and book-to-market, as well as asset pricing factors related to the Fama and French (1993, 2015) three- and five-factor models, and the Carhart (1997) four-factor model, from the data library of Kenneth French. Data related to the Hou, Xue, and Zhang (2015) q-factor model are provided by Lu Zhang and firm-level TFP data are from the website of Selale Tuzel.[12] The definitions of the accounting ratios used in this paper are provided in Section OA.3 of the Online Appendix.

2.2.2 Portfolio formation

To examine the predicted relation between capacity utilization and stock returns in the data, we form portfolios by sorting the cross-section of industries on the basis of each industry's utilization rate. Specifically, at the end of each June from 1967 to 2015 we sort industries into portfolios based on their level of utilization in March of the same year. The three month lag between the release of March utilization data and the June sort date ensures that this strategy is tradeable, as all data used to form portfolios are publicly available by the portfolio formation dates.[13] Each portfolio is then held from July of year t to the end of June of year $t+1$, at which time all portfolios are rebalanced. Annual rebalancing allows us to capture conditional variation in utilization rates.

We form three portfolios on each June sorting date. The low (high) capacity utilization portfolio includes all industries whose utilization rates are at or below (above) the 10^{th} (90^{th}) percentile of the cross-sectional distribution of utilization rates in March of the same year. The medium utilization portfolio includes the remaining industries whose utilization rates fall between these two breakpoints. We focus on these relatively extreme breakpoints to increase the power of our asset-pricing tests. Choosing these breakpoints is useful because our ability to detect a relation between utilization and future stock returns is already attenuated by the relatively small cross-section of industries for which the FRB reports utilization data. It worth stressing, however, that since each portfolio contains multiple industries, each of which is comprised of many individual firms, this choice of breakpoints results in three well-diversified portfolios. We discuss the composition of these portfolios and their characteristics in Section 2.3.1.

[12] We thank Kenneth French, Lu Zhang, and Selale Tuzel for making this data available to us. The firm-level TFP data can be obtained from `http://www-bcf.usc.edu/~tuzel/`.

[13] A three month lag between the portfolio formation month and the month in which utilization rates are measured is conservative since the utilization data for month t are released approximately 15 days into month $t+1$. For example, March utilization rates are typically available by April 15th. Since 1967, the latest the March data has become publicly available is April 17th. A full list of release dates for the G.17 report is available at `https://www.federalreserve.gov/releases/g17/release_dates.htm`.

Overall, this empirical portfolio formation procedure mimics our sorting procedure in the model. Moreover, in Section OA.4.5 of the Online Appendix we show that our results are robust to methodological variations of the portfolio formation procedure (e.g., breakpoints choice, industry choice, etc).[14]

2.2.3 Capacity utilization portfolios: expected return spread

Table 4 reports the annual value- and equal-weighted returns of portfolios sorted on capacity utilization using the procedure described above. Consistent with our model's key prediction, we empirically verify that portfolio returns are monotonically decreasing in the average utilization rate, and that there is an economically and statistically significant spread between returns of the low and high utilization portfolios.

Specifically, the portfolio of industries that utilize a low amount of their productive capacity earns a value-weighted (equal-weighted) average return of 13.64% (10.62%) per annum, whereas the portfolio of industries that utilize a large degree of their capacity earns a value-weighted (equal-weighted) average return of return of 7.96% (5.18%) per annum. The value- and equal-weighted spreads between the returns of the low and the high utilization portfolios are 5.67% and 5.44% per annum, respectively. Each spread is statistically significant at the 5% level.[15] Comparing the theoretical quantitative predictions for the utilization spread shown in Table 3 to the empirical results shown in Table 4, the magnitude of the spread is almost identical (about 5%). Moreover, all portfolio returns in the data, especially those computed on an equal-weighted basis, are close to the predicted returns in the model.

The model results in Table 3 also show that the unconditional CAPM fails to explain the utilization spread, producing alphas of approximately 4.2% per annum. The last row of Table 4 tests this second prediction by examining whether the unconditional CAPM also fails to explain the utilization spread in the data. Empirically, the alphas are obtained by regressing the monthly returns of the spread on the excess returns of the market portfolio, and are annualized by multiplying the resulting regression intercepts

[14]Table OA.4.15 in the Online Appendix reports the transition matrix of industry movements between these three utilization portfolios. The matrix indicates that although capacity utilization is a fairly persistent measure, transitions between the portfolios are still relatively frequent. The former fact demonstrates that the utilization rate can affect the long-horizon risk premium of an industry while the latter fact emphasize the importance of the conditional portfolio rebalancing procedure.

[15]Furthermore, the Sharpe ratio of the value-weighted (equal-weighted) spread is 0.32 (0.35). This is comparable to the Sharpe ratio earned by investing in the value premium over the same period.

by 12. In line with the prediction, the alpha associated with each weighting scheme is positive, slightly smaller in magnitude than the average annual return, and statistically significant. In particular, the value-weighted (equal-weighted) alpha is 4.54% (4.74%) per annum in the data and close to its model-implied counterpart in Table 3.

Table OA.4.16 of the Online Appendix extends this analysis by exploring whether the value-weighted utilization spread is subsumed by four other common empirical asset-pricing models: the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama and French (2015) five-factor model, and the Hou et al. (2015) q-factor model. The annualized alpha resulting from each model is positive and statistically significant at close to the 10% level, or better. This demonstrates that the utilization spread contains some time-series variation that is orthogonal to variation in factor-mimicking portfolios based on alternative investment-related characteristics, such as book-to-market ratios, investment, and profitability.

Recalling our economic rationale for the capacity utilization premium in Section 2.1.3, we note that the failure of both the CAPM and other factor models to explain the capacity utilization spread stems from two reasons. First, the market portfolio is a noisy proxy for aggregate productivity. Second, our risk-based explanation for the spread relies on time-varying exposures to risk, whereas the empirical specifications of these models are unconditional and assume that risk exposures are constant over time.

2.2.4 Capacity utilization portfolios: productivity exposures

Our model's third prediction is that the exposures of the utilization-sorted portfolios to aggregate productivity are decreasing with the utilization rate. In this section we examine this prediction by considering the following projection:

$$Ret^e_{i,t} = \beta_{0,i} + \beta_{1,i}\text{Agg-Prod}_t + \varepsilon_{i,t}, \tag{14}$$

where $Ret^e_{i,t}$ is the value-weighted excess return of the portfolio of interest, Agg-Prod_t is a proxy for aggregate productivity, and $\beta_{1,i}$ captures the exposure of portfolio i to aggregate productivity. To implement this analysis we consider three different proxies for aggregate productivity: the market return,[16] the TFP of the final good sector from Basu et al. (2006), and labor productivity from the Bureau of Labor Statistics (BLS).

Table 5 reports the results and shows the utilization portfolios are indeed differ-

[16]In a single shock model the market return is driven by productivity shocks.

entially exposed to fluctuations in aggregate productivity. The exposure to aggregate productivity falls monotonically from the low to the high utilization portfolio, regardless of the productivity proxy. Importantly, differences in the productivity betas of the extreme utilization portfolios are not only positive, but are also statistically significant at the 5% level. The pattern of these empirical betas is consistent with the pattern of the model-implied betas reported in Table 3. Furthermore, when aggregate productivity is measured using market returns, the magnitude and the spread in the productivity betas are almost identical in both the model and data. This empirical result highlights the importance of aggregate productivity for explaining the utilization spread and is consistent with the economic rationale for the spread developed in Section 2.1.3.[17]

2.3 Distinguishing the utilization premium from related spreads

Section 2.2 documents an economically large and statistically significant utilization spread in the cross-section of returns. Here, we investigate the extent to which this utilization premium is separate from other production-based spreads. Section 2.3.1 reports the characteristics of the utilization-sorted portfolios and indicates that the value and investment effects interact with the utilization spread. Despite these interactions, the results of Fama and MacBeth (1973) regressions in Section 2.3.2 show that the utilization premium is distinct from both of these, and a host of other, cross-sectional effects. We validate the results of these regressions by conducting an extensive set of portfolio double sorts. The double sorts show that the utilization premium is empirically distinct from the value and investment premia, the capacity overhang effect of Aretz and Pope (2018), and the productivity premium of Imrohoroglu and Tuzel (2014). This double sort analysis is reported in Section OA.4 of the Online Appendix.

We also demonstrate that the utilization premium is distinct from cross-sectoral effects, such as the durability spread of Gomes, Kogan, and Yogo (2009). Section 2.3.3 constructs a proxy for (unobservable) firm-level utilization rates using *industry-demeaned* utilization data. Sorting firms into portfolios based on this proxy shows that the utilization premium remains significant at the firm level. Section 2.3.4 shows that the utilization spread also persists when portfolios are sorted on the growth rate

[17]The studies of Zhang (2005), Belo and Lin (2012), and Imrohoroglu and Tuzel (2014), among others, also suggest that heterogeneity in exposures to aggregate productivity can explain the production-related spreads, such as the value and investment premia.

of utilization. These sorts on the growth rate of utilization eliminate any fixed differences in the level of utilization across industries. Finally, Section OA.4 in the Online Appendix reports portfolio sorts *within* economic sectors. The results indicate that an economically large utilization spread persists within the durable manufacturing sector.

2.3.1 Capacity utilization portfolios: characteristics

This section examines the constituents and the characteristics of the utilization-sorted portfolios to empirically determine whether capacity utilization is related to other production-based characteristics.

Portfolio constituents. Panel A of Table 6 reports the average number of firms and industries that constitute each utilization portfolio. By construction, the high and low utilization portfolios each contain approximately 10% of the 45 industries that comprise our sample. Although the number of industries falling into these extreme portfolios is small, these industries are comprised of roughly 960 firms. This means that the low and high utilization portfolios collectively contain about 18% of all firms in our merged CRSP-Compustat sample. This is approximately the same proportion of firms that these portfolios would have contained if the sorts were conducted using firm-level, rather than industry-level, data. Importantly, the fact that each portfolio contains hundreds of firms also implies that these portfolios are well-diversified. Despite a difference in the number of firms underlying each portfolio, there is no statistically significant difference in average firm size between the extreme utilization portfolios, as discussed below.[18] Panel A also shows that the average utilization rate is, by construction, monotonically increasing from the low to the high utilization portfolio.

To shed light on the industries underlying each portfolio, Table 7 reports the five industries that populate the extreme utilization portfolios most often. For each industry, the table also reports the sector to which the industry belongs, and the proportion of years the industry is sorted into the portfolio. The key takeaway from this table is that there is a large degree of sectoral variation associated with the industries that typically populate these portfolios. Panel A shows that leather producers, aerospace manufacturers, and industries that provide supporting services to miners frequently reside in the low utilization portfolio. Panel B shows that the high utilization portfolio

[18] In untabulated results we also show that there is no statistically significant difference in the Herfindahl-Hirschman index between the extreme portfolios.

typically contains both mining industries and utilities, and nondurable manufactures.

Taken together, these results provide suggestive evidence that the utilization premium is not driven by any one sector in particular. Section OA.4.4 of the Online Appendix provides more rigorous evidence that the utilization spread is mostly a within-sector, rather than a cross-section, phenomenon. Specifically, we conduct portfolio sorts *within* economic sectors. This analysis demonstrates that the utilization spread is orthogonal to the durability spread of Gomes et al. (2009): the utilization premium exists *within* the durable sector (and when excluding the mining and utilities sectors).

Portfolio characteristics. Panel B of Table 6 reports the average industry-level characteristics of each capacity utilization portfolio. There is no statistically significant difference between the low and high portfolios in terms of size, probability, as measured by either ROA or gross profitability, or asset growth. Consistent with the fact that TFP and the hiring rate are positively correlated with capacity utilization (see Table OA.2.5 of the Online Appendix), the low portfolio has both lower industry-level TFP and hiring rates than the high portfolio. However, these differences are small and statistically insignificant. Additionally, while existing studies show that low productivity and low hiring rate firms command a larger risk premium (see Imrohoroglu and Tuzel (2014) and Belo et al. (2014)), the evidence in Panel B shows that the capacity utilization premium is materially independent of these effects. The only three characteristics that are significantly different between the two extreme portfolios are the book-to-market ratios, investment rates, and idiosyncratic return volatilities (IVOL). However, as Ang, Hodrick, Xing, and Zhang (2006) show that high IVOL firms earn low expected returns, this difference in IVOL cannot account for the capacity utilization premium.

Panel B raises the concerns that since low (high) capacity utilization industries also tend to be value (growth) industries with low (high) investment rates, the utilization spread may be driven by the value or the investment premium. Each of these potentially confounding effects are well-established in the context of the asset pricing literature. For instance, Fama and French (1993) demonstrate the ability of book-to-market to predict future stock returns, while Titman, Wei, and Xie (2004) and Xing (2008) show that low investment rates are associated with high future returns. To establish a degree of independence between the utilization spread and the value and the investment premia, we implement a Fama and MacBeth (1973) regression analysis. We show that relation between utilization and risk premia remains negative, economically sizable, and statistically significant after controlling for book-to-market, investment, and a host of other production-based characteristics. Section OA.4 of the Online Appendix provides a double sort analysis that corroborates the results of these regressions.

2.3.2 Firm-level Fama-Macbeth regression analysis

The previous section shows that differences in book-to-market and investment are statistically different between the extreme capacity utilization portfolios. To address this finding, this section performs firm-level Fama and MacBeth (1973) regressions and shows that capacity utilization has predictive power for risk premia that is incremental to value, investment, and multiple other investment-related characteristics. These regressions are implemented as follows. In each year t we run a separate cross-sectional regression in which the dependent variable is a firm's annual excess return from the start of July in year t to the end of June in year $t+1$, and the independent variables are a vector of the firm's characteristics, $\boldsymbol{X}_t$, measured at the end of June in year t. These cross-sectional regressions are specified as follows:

$$R_{i,t\to t+1} = \beta_{0,t} + \boldsymbol{\beta}_t'\boldsymbol{X}_{i,t} + \varepsilon_{i,t\to t+1} \quad \forall t \in \{1967, \ldots, 2014\}. \tag{15}$$

The characteristics we consider are capacity utilization, TFP, hiring, investment over physical capital, capacity overhang, the natural logarithms of size and book-to-market, and the lagged annual return. A capacity utilization rate is assigned to each firm following the matching procedure described in Section OA.4.1 of the Online Appendix, and all control variables are divided by their unconditional standard deviation to aid comparisons between different regressions. After running these cross-sectional regressions we compute the time-series average of each estimated slope coefficient to assess the relation between a particular characteristic and future stock returns, while holding all other characteristics constant. The results of this analysis are reported in Table 8.

Columns 1 to 8 show the results of including each characteristic in a univariate regression. The average loading of capacity utilization is negative and statistically significant at the 5% level. The loadings on TFP, hiring and investment rates, capacity overhang, and size are also negative and significant, while the coefficient associated with book-to-market is positive and significant. The relation between lagged annual returns and future annual returns is statistically insignificant, indicating that returns at the annual horizon are not predictable. The signs of these variables are consistent with the documented spreads associated with each characteristic of interest.[19]

Columns 9 to 15 show that the coefficient on utilization remains negative and significant at the 5% level when we include one or more investment-related characteristics in the regressions. Furthermore, each of the companion characteristics we consider also remains negative and significant. This provides additional evidence that the relation between utilization and stock returns is somewhat orthogonal to the known relations between returns and each of TFP, hiring, investment, and capacity overhang.

Lastly, Column 16 considers a regression including all eight characteristics simultaneously. The loading on utilization remains negative and statistically significant at the 5% level, even after augmenting the set of control variables used in Column 15 to include size, book-to-market, and past returns. This demonstrates the robustness of relation between utilization and risk premia. Compared to the results in Column 15, the remaining results in Column 16 are largely similar with the exception of the loading on TFP, which flips sign from negative to positive. However, this change in sign does not compromise the validity of the TFP spread as a number of the investment-related characteristics included in this specification are relatively highly correlated.

Section OA.4 of the Online Appendix validates this regression analysis by conducting portfolio double sorts. The sorts confirm the distinction between the utilization premium, and the value, investment, and overhang premia. In Section OA.4.3 we decompose firms-level TFP into its components and compare the utilization premium to the productivity premium of Imrohoroglu and Tuzel (2014). We show the productivity premium is driven by two underlying and distinct components: the utilization premium from Section 2.2.3, and a spread based on time-varying technology and markups.

2.3.3 Empirical proxy for firm-level capacity utilization rates

Our empirical results in Section 2.2.3 are based on industry-level utilization data from the FRB. We opt to use this data for the benchmark analysis, rather than estimate

[19]In particular, Imrohoroglu and Tuzel (2014) show that low TFP predicts high stock returns, Belo et al. (2014) find that low hiring is associated with high stock returns, Titman et al. (2004) and Xing (2008) document the relation between low investment rates and high stock returns, Aretz and Pope (2018) find that higher capacity overhang predicts lower stock returns, and Fama and French (1993) discuss how both low market capitalization and high market-to-book ratios predict high stock returns. While the estimates associated with lagged excess returns are not statistically significant, the sign of these point estimates is in line with long-term reversals (De Bondt and Thaler, 1985).

unobserved firm-level utilization rates using Compustat data, due to the FRB data's transparency, coverage, and high frequency. However, to tighten the link between the empirical results and the aggregation level of the model, which is based on firm-level production units, this section presents firm-level evidence for the utilization spread.

We develop a proxy for the unobservable firm-level capacity utilization rates. Our approach projects industry-level utilization rates onto five production-related characteristics that are motivated by our model. We use either raw utilization rates or utilization rates that are demeaned at the industry-level to construct the proxy. The latter approach ensures that the utilization proxy is not driven by fixed differences in average utilization rates across industries.

The proxy for the firm-level utilization rate is obtained as follows. First, the raw or industry-demeaned utilization rate of industry j at time t ($CU_{j,t}$) is projected on the five industry-level characteristics. We use the following regression:

$$CU_{j,t} = \beta_{j,0}+\beta_{j,1}\ln(ME)_{j,t}+\beta_{j,2}\ln(B/M)_{j,t}+\beta_{j,3}\left(I/K\right)_{j,t}+\beta_{j,4}IVOL_{j,t}+\beta_{j,5}TFP_{j,t}+\varepsilon_{j,t} \tag{16}$$

By estimating the projection above separately for each industry, the relation between the characteristics and utilization is industry-specific. Second, each firm is assigned to a unique industry by the matching procedure described in Section OA.4.1 of the Online Appendix. Third, the proxy for the utilization rate of a firm i assigned to industry j at time t ($\hat{CU}_{i,j,t}$) is computed using the slope coefficients from the projection above:

$$\hat{CU}_{i,j,t} = \hat{\beta}_{j,0}+\hat{\beta}_{j,1}\ln(ME)_{i,j,t}+\hat{\beta}_{i,j,2}\ln(B/M)_{i,j,t}+\hat{\beta}_{j,3}\left(I/K\right)_{i,j,t}+\hat{\beta}_{j,4}IVOL_{i,j,t}+\hat{\beta}_{i,j,5}TFP_{i,j,t} \tag{17}$$

Combining firm-level characteristics with industry-specific slope coefficients allows the utilization proxy to vary between firms in the same industry.

We use the proxy for the firm-level utilization rate to sort firms into portfolios as per Section 2.2.2, and report the results in Table 9. The table shows an economically large and statistically significant utilization spread also exists in the cross-section of firm-level returns. Similar results are obtained using both raw or industry-demeaned utilization rates. The relation between utilization and stock returns also remains monotonically decreasing in either case. Taken together, the results show the utilization spread (1) persists at the firm-level, in line with our model's level of aggregation, and (2) is not driven by fixed differences in utilization rates between industries.

2.3.4 Removing industry-specific fixed effects

Our main analysis sorts industries into portfolios based on the level of each industry's utilization rate. Our choice to sort on the level of utilization, rather than its growth or other moments, is motivated by the fact that the level of capacity utilization is the economic primitive of interest in our model.

While each industry's utilization rate fluctuates considerably over time, the average level of utilization shows some (typically statistically insignificant) differences across industries. This raises a concern that the utilization premium may be influenced by ex-ante heterogeneity between industries. The previous sections alleviate this concern in three ways. First, industries from different sectors are commonly sorted into the extreme-utilization portfolios (recall Table 7). Second, the utilization spread exists *within* the durable manufacturing sector (see Section OA.4.4 of the Online Appendix). Third, sorting firms into portfolios based on a proxy computed using industry-demeaned utilization rates also yields a sizable utilization spread (see Section 2.3.3). This section provides additional evidence that the utilization spread is not driven by industry-specific differences in average utilization rates.

Here, we sort industries into portfolios based on the year-on-year *growth* rate, rather than the *level*, of utilization. Using the growth rate removes any (potential) difference in the average level of utilization across industries. The portfolio formation procedure follows that in Section 2.2.2, apart from the use of growth rates. The results are reported in Table 10. The table shows that the value-weighted (equal-weighted) utilization spread is 4.80% (5.74%) per annum and is significant at the 5% level. Portfolio returns are also monotonically decreasing in the utilization growth rate.

3 Utilization's role: fixing investment dynamics and substituting adjustment costs

In this section we highlight the roles of flexible utilization for jointly targeting cross-sectional risk premia and investment moments, and serving as a substitute for greater adjustment cost parameters. We first show the failures of the model without utilization to target asset-pricing and production moments, and highlight the tensions that gives rise to these model misses. We then explain how utilization provides a solution to these tensions, and also show that these tensions cannot *jointly* be addressed by altering the calibration of the model with fixed utilization. Lastly, we demonstrate that flexible utilization permits us to target key moments with a lower degree of adjustment costs vis-à-vis a model with fixed utilization.

3.1 A fixed utilization model: the failures

The benchmark model's success in jointly fitting (1) the volatility and skewness of investment, both across time and across firms, and (2) risk premia, crucially hinges on variable capacity utilization rates. To illustrate this point, row (1) of Table 11 shows model-implied moments in an economy without flexible utilization (i.e., $\lambda \to \infty$).

With fixed utilization, the distribution of investment rates exhibits far less variability and asymmetry compared to the data both in the time-series and the cross-section. The time-series skewness of firm-level investment turns to -0.09, at odds with its empirical sign and magnitude of 0.67. Investment's time-series volatility drops to only 11%, and its autocorrelation becomes slightly too high. Cross-sectional moments also become severely distorted. The dispersion of investment rates is about a half of its empirical counterpart (7% in the model vs 16% in the data). The cross-sectional skewness of investment rates is merely 0.11 in the model, whereas it is much higher in the data (about 1.9). The model-implied value and investment spreads are also about 1% per annum smaller in this model than the data.

The model with fixed utilization fails to capture both investment's moments and risk premia because (1) the fixed adjustment cost makes disinvestment a real option, and (2) without flexible utilization, a firm's *only* way to respond to a negative productivity shock is by exercising this option to disinvest. In general however, a drop in productivity triggers two opposite forces. On the one hand, firms desire to sell capital as its expected marginal product falls. On the other hand, firms desire to "wait and see" if productivity will recover before exercising their real options to sell capital.

If a productivity drop at time t is not extremely severe, then the "wait and see" effect tends to dominates. This leads to periods of investment-policy inaction in which many firms do not alter their stocks of capital over time. During these periods, each waiting firm j sets its investment rate, $i_{j,\tau}$, to the constant depreciation rate of δ_k for all $\tau \in [t, t+\hat{t})$, where $\hat{t}$ is the ending time of the endogenous inaction period. Since a mass of waiting firms are clustered around the center of investment's distribution (i.e., around δ_k), investment's dispersion and cross-sectional skewness are both lowered.

Furthermore, if productivity remains persistently low, then at time $t + \hat{t}$ waiting firms pass a tipping point in which they are overly burdened with unproductive capital and choose to disinvest this capital sharply. This implies that $i_{j,t+\hat{t}} << \delta_k$.[20] Thus, these periods of inaction are often followed by negative investment spikes, producing the negative skewness of firm-level investment that is inconsistent with the data.[21]

The distorted distribution of firm-level investment rates in the model with fixed utilization also has an adverse impact on risk premia. Because investment's distribution features too little dispersion and asymmetry, there is too little heterogeneity between firms in this economy. Thus, sorting firms into portfolios based on investment (or valuation ratios) implies that both the top and bottom quintiles (tails) of Tobin's Q contain fewer extreme outcomes compared to the benchmark model with flexible utilization. As differences in cross-sectional risk premia are fundamentally driven by heterogeneity in economic activity, investment-related spreads get smaller when utilization is fixed.

3.2 Flexible utilization: a solution

Our benchmark model with flexible utilization overcomes the counterfactuals outlined in Section 3.1 by making the depreciation rate endogenously stochastic. This improved model fit is highlighted in row (6) of Table 11. When firms can choose utilization, they have an additional mechanism with which to scale down production in response to adverse productivity shocks, even as they "wait and see" if productivity recovers. Put differently, firms in this economy can respond to moderate drops in productivity by utilizing their existing machines less intensively rather than selling machines. As underutilized capital depreciates slower, more capital is preserved for more productive future periods (i.e., $\delta(u_{j,\tau}) < \delta_k$ if $u_{j,\tau} < 1 \ \forall \tau \in [t, t + \hat{t})$).

[20]Put differently, when firms are close to the disinvestment threshold, then the investment policy is locally concave and the expected value of investment becomes negative.

[21]Importantly, the counterfactuals in the model with fixed utilization cannot be remedied simply via the aggregation level. As many real options operate at the plant level, one can argue that a collection of production units in the model comprise one firm. Aggregation of many units into a single firm does indeed smooth model-implied investment rates by shrinking periods of investment inaction, and lowering the size of disinvestment jumps. However, these model-implied moments remain unaligned with the data. We verify this in untabulated simulation by aggregating 30 production units into a firm. The resulting model-implied volatility of investment is smaller than the data. While the skewness of investment turns positive, this quantity is close to zero (remaining significantly lower than the data).

Lower utilization also reduces the natural rate of investment needed to maintain the current capital stock. Thus, even as firms wait to sell their capital, the investment required to maintain their existing machinery, $i_{j,\tau} = \delta(u_{j,\tau})$, becomes endogenously stochastic. This time-varying depreciation, micro founded by utilization, eliminates the counterfactual long periods of constant investment. As a result, the time-series volatility of firm-level investment rises, and the cross-sectional dispersion of investment increases. To see the latter, note that each firm's utilization rate is positively correlated with its idiosyncratic productivity shocks. Since these shocks are heterogeneous between waiting firms, $u_{j,\tau} \neq u_{k,\tau} \Rightarrow i_{j,\tau} \neq i_{k,\tau}$ for firms j and k.

Moreover, the positive correlation between productivity and utilization also implies that firms opt to raise utilization in times of higher productivity. Utilizing capital more intensively in good times increases both depreciation and the natural rate of investment (i.e., $\delta(u_{j,\tau}) > \delta_k$), and means that larger investments are required to expand capacity for future periods. To see this, suppose that at time τ a firm wants to expand its capacity by $\delta_k K$ units. With fixed utilization, the required investment rate is $i_\tau = 2\delta_k$. However, with flexible utilization, the required investment rate increases to $i_\tau = \delta(u_{j,\tau}) + \delta_k > 2\delta_k$. Since investment becomes even more procyclical, its time-series and cross-sectional skewness rise and turn positive (consistent with the data).

The increases in the skewness and dispersion of investment rates also boost risk premia in the model with flexible utilization, as seen by comparing the value premium between rows (6) and (1) of Table 11. The fact that the cross-section of investment rates is almost 12 times as skewed in the model with flexible utilization has implications for asset prices. This is because the magnitude of the value premium is (partially) determined by the economic fundamentals of the firms that fall into the extreme book-to-market portfolios. With fixed utilization, the portfolio of growth firms (bottom 20% of book-to-market) includes both firms with very high and moderately elevated investment rates (as investment comoves with Tobin's Q, which is inversely related to book-to-market). With flexible utilization, the right tail of investment's distribution becomes thicker, and the portfolio of growth firms includes firms with only very high investment rates. Since the growth options of these extreme-investment firms increase in value exactly when the price of risk is high (bad states), their skewed investment behavior provides an excellent hedge against bad states. This decreases the risk premium of growth firms, and increases the magnitude of the book-to-market spread.

3.3 Flexible utilization as substitute for adjustment costs

Without flexible utilization, the problem of matching investment's moments with the data is not simply alleviated by recalibrating the model. In this section we show that flexible utilization can reduce both the amount and the magnitude of the exogenous adjustment parameters required to target investment and risk premia jointly. Below, we illustrate this role of utilization by perturbing the capital adjustment cost parameters in a model without utilization.

Rows (2) and (3) of Table 11 show salient investment's moments when we alter the fixed cost of disinvestment after utilization is shut down. Row (2) shows that while lowering the fixed cost causes investment's skewness to move closer to the data, this quantity remains too small. Additionally, this improved fit comes at the expenses of further decreasing (increasing) the magnitude of the value premium (autocorrelation of investment) compared to its empirical counterpart.[22] In addition, lowering the fixed cost has virtually no impact on the cross-sectional dispersion of investment which is still too small. Row (3) shows that raising the fixed cost causes the model's mismatch to become even more severe. The time-series and cross-sectional skewness of investment become counterfactually negative, and investment's volatility falls. Despite the higher fixed cost, the value premium is still smaller than the data.

In rows (4) we lower the quadratic adjustment cost while keeping utilization fixed. The lower friction in row (4) can help turn the time-series skewness of investment to a positive value, but the cross-sectional skewness of investment is still too small and falls further from the data. Fewer frictions also cause risk premia to fall. For instance, the value premium, which is already too low in the model with fixed utilization, falls to almost half of its empirical magnitude.

Naturally, the diminished value premium in the model without utilization can be boosted by increasing the quadratic capital adjustment cost. With higher adjustment costs, shocks are absorbed in asset prices rather than investment quantities. We demonstrate this in row (5), where we search for an adjustment cost parameter ϕ to match the value premium in the model with fixed utilization. Our structural search suggests that ϕ needs to be around 3.00 to match this spread. While this parameter value is

[22]In untabulated results we confirm that even if f is set to zero, the cross-sectional skewness of investment is only a half of its empirical magnitude.

broadly consistent with the existing literature, this value is twice as high as the value of the same parameter in our benchmark model with flexible utilization.

While row (5) shows that risk premia in the model with fixed utilization can be boosted by increasing ϕ, this simultaneously distorts the distribution of investment rates. Investment's dispersion becomes a quarter of its magnitude in the data. Furthermore, unlike the data, the time-series skewness of investment becomes even more counterfactually negative under this alternative calibration. The counterfactual distributions of investment in rows (3) and (5) complement the evidence in Clementi and Palazzo (2019). They show that production models that target cross-sectional risk premia often rely on excessive capital adjustment costs (particularly for disinvestment). While boosting returns, these frictions imply that the model-implied investment exhibits too few depressed periods and too much inaction compared to the data.

Below, we highlight how flexible utilization may remedy the concerns raised by Clementi and Palazzo (2019). As we show in row (6) of Table 11, flexible utilization allows our baseline model to feature capital adjustment costs that are smaller than those in the model without utilization. These smaller costs are still sufficient to simultaneously produce sizable risk premia and realistic investment dynamics, including many periods of depressed investment. This happens because of three key mechanisms.

The first mechanism is related to the impact of lower utilization on convex adjustment costs. For a given quadratic adjustment cost parameter, flexible utilization implies more observed disinvestment while keeping the amount of friction (risk) the same. To see this, suppose that a firm desires to drop its capital stock by $\delta_k K$. With fixed utilization, the firm chooses an investment rate of $i = 0$, and the quadratic cost will be proportional to δ_k^2. However, with flexible utilization, a drop in productivity triggers a drop in utilization, which in turn lowers the firm's depreciation to $\delta(u_{i,t}) < \delta_k$. To shed $\delta_k K$ capital, the investment rate is set to the lower rate of $i = -\delta_k + \delta(u_{i,t}) < 0$. The quadratic cost will be unaltered, and remain proportional to $(-\delta_k + \delta(u_{i,t}) - \delta(u_{i,t}))^2 = \delta_k^2$. Thus, in a model with flexible utilization, one may see more disinvestment without compromising on the frictions that induce risk premia.

The second mechanism is the time-varying depreciation rate driven by changes in utilization. Because of the frictions caused by the disinvestment option and the "wait and see" policy, the fixed utilization model features too many periods of stale investment and too few periods of very low investment compared to the data. Consequently, the dispersion and skewness of investment in the model with fixed utilization are too low for the reasons discussed in Section 3.1 (recall row (1) of Table 11). However, as explained in Section 3.2, a drop in utilization lowers depreciation, leading to many periods in which the investment of waiting firms oscillates in a region below its unconditional level (i.e., below the average depreciation rate), consistent with the data.

The third mechanism involves the changes in the cross-sectional distribution of investment rates that are caused by flexible utilization. As we outline in Section 3.2, utilization makes the cross-sectional distribution of investment more skewed. By featuring more extreme observations in the tails of investment's distribution, one can obtain quantitatively large return spreads with moderated values of adjustment costs.

Note that the model without utilization could also match some of the aforementioned moments given a more elaborate adjustment cost function. For example, extra adjustment costs may include piecewise quadratic and linear terms (e.g., Bloom (2009) and Belo and Lin (2012)). However, these additional costs are driven by more exogenous calibration parameters. Furthermore, while featuring a higher degree of frictions may help target risk premia, these same frictions may further inhibit the dispersion-related moments of investment. By contrast, capacity utilization relies on only one additional parameter, the elasticity of depreciation to changes in utilization, that is calibrated to the volatility of utilization rather than an investment-related moment.

Finally, we check the sensitivity of the benchmark model with flexible utilization to other model parameter in Panel B of Table 11. As the elasticity of depreciation to utilization (λ) rises, utilization's volatility falls since increasing utilization causes capital to depreciate even faster. This means that, as shown in row (8), utilization induces smaller cyclical fluctuations in the natural rate of investment, and makes its skewness less positive. Rows (9) and (10) show that when the quadratic adjustment cost rises, the volatility of investment falls. Since capital is less volatile, the marginal product of utilization is less volatile and smaller fluctuations in utilization are needed to scale output up (down) in good (bad) economic times.

Overall, flexible utilization not only requires fewer degrees of freedom than featuring more complex adjustment costs, but also offers a way to endogenize the implications of higher adjustment cost parameters using a micro-founded margin.

4 Conclusion

We introduce flexible capacity utilization into a production framework and examine the model's implications for both cross-sectional risk premia and firm-level investment dynamics. Incorporating capacity utilization into the model yields novel theoretical implications regarding the relation between capacity utilization and future stock returns. Flexible utilization also allows our model to match numerous asset-pricing and investment-related moments to the data simultaneously, and to endogenize the implications of exogenously specified adjustment costs via a micro-founded margin.

Our model yields two key novel predictions for asset prices: (1) low capacity utilization firms should earn an average excess return that is 5% per annum greater than that of high capacity utilization firms; and (2) lower utilization firms should have higher exposures to aggregate productivity.

The economic rationale for the utilization premium relates to utilization's ability to offset disinvestment risk. In the model, downscaling capital in the presence of capital adjustment costs increases firms' exposures to aggregate risk. Lowering utilization allows firms to hedge this risk in multiple ways. First, lower capacity utilization implies that firms use their installed machines less intensively, and causes the capital depreciation rate to decrease. This lower depreciation conserves more capital for future periods that are more productive. Second, the decrease in the depreciation rate is particularly useful for firms that disinvest. This is because a decrease in depreciation drops the natural rate of investment and, consequently, reduces the convex adjustment costs required to downscale. Moreover, when selling machines involves paying a fixed cost, firms can substitute selling capital (and exercising their costly real options to disinvest) by lowering utilization. Overall, low utilization firms are riskier because a low capacity utilization rate is indicative of a firm that faces high frictions in the market for selling capital (and tries to partially alleviate these frictions through utilization).

Using industry-level utilization data from the FRB, we confirm the predicted sign and magnitude of the novel capacity utilization spread in the data. We also confirm that the pattern in the exposures of the utilization-sorted portfolios to aggregate productivity is in line with our model's second prediction. This relation between capacity utilization and risk premia is distinct from other production-related spreads, such as those related to investment and book-to-market.

Although we use industry-level portfolios to establish the capacity utilization spread, the spread does not reflect ex-ante heterogeneity in utilization rates between different industries. For instance, we show that the utilization spread also persists *within* groups of similar industries, such as a subsample of durable goods manufacturers. When we construct a proxy for firm-level utilization rates using Compustat data we obtain a similar spread. Furthermore, the negative relation between utilization and excess returns continues to holds if we use the growth rate, rather than the level, of industry-level capacity utilization to form portfolios.

In regards to real quantities, we find that flexible utilization is crucial to jointly match investment's volatility, dispersion, skewness (in both the time-series and cross-section), and the value premium to the data. When we fix utilization in our model, the distribution of firm-level investment rates becomes distorted. Specifically, the cross-section of investment's distribution becomes too compressed, and the time-series skewness of firm-level investment rates becomes counterfactually negative. Risk premia in this model are also too small unless capital adjustment frictions are increased, further distorting investment's distribution.

We show that these model misses can be addressed jointly in a model with flexible utilization rates. By inducing a time-varying depreciation rate, flexible utilization increases the dispersion and asymmetry in investment's distribution. Matching these properties of investment's distribution to the data is key for production-based asset pricing. When model-implied investment rates are as skewed as their empirical counterparts, the model can match cross-sectional risk premia with lower adjustment costs. With flexible utilization we are able to match the value premium using only half of the adjustment costs required under fixed utilization. More generally, we show that for a given quadratic adjustment cost parameter, flexible utilization implies more observed disinvestment, while keeping the amount of friction (risk) the same.

In all, the empirical and theoretical results in this study demonstrate the economic importance of varying utilization for business cycles moment and risk premia. The role of flexible utilization in improving investment dynamics and relaxing adjustment costs provides a new hope for one-factor production-based asset-pricing models.

References

Abel, A. B., Eberly, J. C., 1996. Optimal investment with costly reversibility. The Review of Economic Studies 63, 581–593.

Andreou, E., Gagliardini, P., Ghysels, E., Rubin, M., 2019. Inference in group factor models with an application to mixed frequency data. Econometrica .

Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. Journal of Finance 61, 259–299.

Aretz, K., Pope, P. F., 2018. Real options models of the firm, capacity overhang, and the cross section of stock returns. The Journal of Finance 73, 1363–1415.

Bansak, C., Morin, N., Starr, M., 2007. Technology, capital spending, and capacity utilization. Economic Inquiry 45, 631–645.

Bansal, R., Yaron, A., 2004. Risks for the long run: A potential resolution of asset pricing puzzles. Journal of Finance 59, 1481–1509.

Basu, S., Fernald, J. G., Kimball, M. S., 2006. Are technology improvements contractionary? American Economic Review 96, 1418–1448.

Belo, F., Lin, X., 2012. The inventory growth spread. The Review of Financial Studies 25, 278–313.

Belo, F., Lin, X., Bazdresch, S., 2014. Labor hiring, investment, and stock return predictability in the cross section. Journal of Political Economy 122, 129–177.

Berk, J. B., Green, R. C., Naik, V., 1999. Optimal investment, growth options, and security returns. The Journal of Finance 54, 1553–1607.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.

Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I., Terry, S. J., 2018. Really uncertain business cycles. Econometrica 86, 1031–1065.

Board of Govenors of the Federal Reserve System, 2014. History of the capacity and capacity utilization index.

Board of Govenors of the Federal Reserve System, 2017a. Capacity utilization explanatory notes.

Board of Govenors of the Federal Reserve System, 2017b. Industry structure of capacity and capacity utilization: Classification, value-added proportions, and descriptions of series.

Burnside, C., Eichenbaum, M., 1996. Factor-hoarding and the propagation of business-cycle shocks. American Economic Review 86, 1154–74.

Burnside, C., Eichenbaum, M., Rebelo, S., 1995. Capital utilization and returns to scale. University of Chicago Press, vol. 10, pp. 67–124.

Campbell, J., Cochrane, J., 1999. Force of habit: A consumption-based explanation of aggregate stock market behavior. Journal of Political Economy 107, 205–251.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52, 57–82.

Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: Implications for the cross-section of returns. The Journal of Finance 59, 2577–2603.

Clementi, G. L., Palazzo, B., 2019. Investment and the cross-section of equity returns. The Journal of Finance 74, 281–321.

Comin, D., Gertler, M., 2006. Medium-term business cycles. American Economic Review 96, 523–551.

Cooper, I., 2006. Asset pricing implications of nonconvex adjustment costs and irreversibility of investment. The Journal of Finance 61, 139–170.

Cooper, I., Wu, G., Gerard, B., 2005. Investment irreversibility, real activity and the value premium. Working Paper .

Cooper, M. J., Gulen, H., Schill, M. J., 2008. Asset growth and the cross-section of stock returns. The Journal of Finance 63, 1609–1651.

Cooper, R. W., Haltiwanger, J., 2006a. On the nature of capital adjustment costs. Review of Economic Studies 73, 611–633.

Cooper, R. W., Haltiwanger, J. C., 2006b. On the nature of capital adjustment costs. The Review of Economic Studies 73, 611–633.

Corrado, C., Mattey, J., 1997. Capacity utilization. Journal of Economic Perspectives 11, 151–167.

Da, Z., Huang, D., Yun, H., 2017. Industrial electricity usage and stock returns. Journal of Financial and Quantitative Analysis 52, 37–69.

Daniel, K., Titman, S., 2006. Market reactions to tangible and intangible information. The Journal of Finance 61, 1605–1643.

De Bondt, W. F. M., Thaler, R., 1985. Does the stock market overreact? Journal of Finance 40, 793–805.

Dixit, A., Pindyck, R., 1994. Investment under Uncertainty. Princeton University Press.

Fama, E., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1 – 22.

Fama, E. F., MacBeth, J. D., 1973. Risk, Return, and Equilibrium: Empirical Tests. Journal of Political Economy 81, 607–636.

Garlappi, L., Song, Z., 2017a. Can investment shocks explain the cross section of equity returns? Management Science 63, 3829–3848.

Garlappi, L., Song, Z., 2017b. Capital utilization, market power, and the pricing of investment shocks. Journal of Financial Economics 126, 447 – 470.

Gilbert, C., Morin, N. J., Raddock, R. D., 2000. Industrial production and capacity utilization: recent developments and the 1999 annual revision. Federal Reserve Bulletin pp. 20–33.

Gomes, J. F., Kogan, L., Yogo, M., 2009. Durability of output and expected stock returns. Journal of Political Economy 117, 941–986.

Greenwood, J., Hercowitz, Z., Huffman, G., 1988. Investment, capacity utilization, and the real business cycle. American Economic Review 78, 402–17.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: An investment approach. The Review of Financial Studies 28, 650.

Imrohoroglu, A., Tuzel, S., 2014. Firm-level productivity, risk, and return. Management Science 60, 2073–2090.

Jaimovich, N., Rebelo, S., 2009. Can news about the future drive the business cycle? American Economic Review 99, 1097–1118.

Jones, C. S., Tuzel, S., 2013. Inventory investment and the cost of capital. Journal of Financial Economics 107, 557 – 579.

King, R., Rebelo, S., 1999. Resuscitating real business cycles. In: *Handbook of Macroeconomics*, Elsevier, vol. 1, Part B, chap. 14, pp. 927–1007, first ed.

Kydland, F., Prescott, E., 1988. The workweek of capital and its cyclical implications. Journal of Monetary Economics 21, 343–360.

Li, J., 2018. Explaining momentum and value simultaneously. Management Science 64, 4239–4260.

Lucas, R., 1970. Capacity, overtime, and empirical production functions. American Economic Review 60, 23–27.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–08.

Novy-Marx, R., 2013. The other side of value: The gross profitability premium. Journal of Financial Economics 108, 1–28.

Papanikolaou, D., 2011. Investment shocks and asset prices. Journal of Political Economy 119, 639 – 685.

Rouwenhorst, G., 1995. Asset pricing implications of equilibrium business cycle models. In: *Frontiers of Business Cycle Research*, Princeton University Press, p. 294–330.

Stambaugh, R. F., Yuan, Y., 2017. Mispricing factors. The Review of Financial Studies 30, 1270–1315.

Tauchen, G., Hussey, R., 1991. Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. Econometrica 59, 371–396.

Titman, S., Wei, K. C. J., Xie, F., 2004. Capital investments and stock returns. Journal of Financial and Quantitative Analysis 39, 677–700.

Xing, Y., 2008. Interpreting the value effect through the q-theory: An empirical investigation. The Review of Financial Studies 21, 1767–1795.

Zhang, L., 2005. The value premium. The Journal of Finance 60, 67–103.

Tables and Figures

Table 1: **Model calibration**

Symbol	Description	Value
Stochastic processes		
ρ_x	Persistence of aggregate productivity	0.922
σ_x	Conditional volatility of aggregate productivity	0.014
$\bar{z}$	Long-run average of idiosyncratic productivity	-0.163
ρ_z	Persistence of idiosyncratic productivity	0.600
σ_z	Conditional volatility of idiosyncratic productivity	0.300
β	Time discount factor	0.988
γ_0	Constant price of risk	3.375
γ_1	Time-varying price of risk	-8.800
Technology		
α_k	Capital share	0.333
α_l	Labor share	0.667
θ	Returns to scale of production	0.950
δ_k	Fixed capital depreciation rate	0.080
δ_u	Slope of depreciation rate	0.092
λ	Elasticity of marginal depreciation	3.000
ω	Sensitivity of wages to aggregate productivity	0.200
ϕ	Adjustment cost parameter	1.500
f	Fixed cost parameter	0.028

This table reports the calibrated parameter values of the production-based asset pricing model described in Section 1.1.

Table 2: **Model-implied moments**

Variable	Data	Model
Panel A: Real quantities		
Volatility of firm-level investment rates (time-series)	0.14	0.14
Dispersion of firm-level investment rates (cross-sectional)	0.16	0.11
AC(1) of firm-level investment rates	0.52	0.58
AC(2) of firm-level investment rates	0.26	0.38
Skewness of firm-level investment rates (time-series)	0.67	0.66
Skewness of firm-level investment rates (cross-sectional)	1.89	1.19
Inter-decile range of investment rates	0.32	0.22
Volatility of aggregate capacity utilization level	4.09	4.15
AC(1) of aggregate capacity utilization level	0.65	0.92
Panel B: Asset prices		
Real risk-free rate	1.19	1.21
Excess market return	6.28	5.39
Volatility of excess market return	17.20	20.89
AC(1) of excess market return	-0.05	-0.00
Book-to-market spread	3.71	3.76
Investment spread	3.70	4.27

The table shows model-implied moments, obtained by simulating 1,000 firms for 40,000 periods (years), alongside their empirical counterparts, computed using data from 1967 to 2015. Panel A displays moments associated with firm-level investment rates and aggregate capacity utilization rates, while Panel B reports asset-pricing moments related to the risk-free rate, equity premium, and the book-to-market and investment spreads. In each panel AC(1) and AC(2) refer to the first- and second-order autocorrelation of the given variable.

Table 3: **Capacity utilization and stock returns: model**

Portfolio	Value-weighted Mean		Value-weighted β		Equal-weighted Mean	
Low (L)	9.86	[9.53]	0.34	[1.19]	10.08	[9.74]
	(6.21, 14.50)		(0.21, 0.51)		(6.37, 14.66)	
Medium	7.11	[6.83]	0.32	[1.10]	7.54	[7.24]
	(3.99, 11.04)		(0.20, 0.47)		(4.28, 11.59)	
High (H)	4.64	[4.41]	0.28	[0.98]	4.82	[4.57]
	(1.98, 8.07)		(0.18, 0.41)		(2.11, 8.26)	
Spread	5.21	[5.12]	0.06	[0.20]	5.26	[5.16]
(L-H)	(4.14, 6.54)		(0.02, 0.11)		(4.20, 6.50)	
α_{CAPM}	4.17	[3.89]			4.21	[3.93]
(L-H)	(2.77, 5.98)				(2.81, 5.96)	

The table shows the model-implied annual value- and equal-weighted returns of portfolios sorted on capacity utilization, as well as the exposure of each model-implied capacity utilization portfolio to aggregate productivity (β) and the CAPM α (α_{CAPM}). To compute β in the model the volatility of aggregate productivity in the model is scaled to match the volatility of market returns in the data. Model-implied moments are based on 500 simulations of 4,000 firms for 50 periods (years). The population moments reported in the table are obtained by averaging all model-implied moments across the Monte Carlo simulations of the economy. In each simulation, and in each year, firms are sorted into the high (low) utilization portfolio if their level of capacity utilization is above (below) the 90^{th} (10^{th}) percentile of the cross-sectional distribution of capacity utilization rates in the previous period. Parenthesis report the 90% confidence related to each moment and square brackets report each population moment, obtained from one simulation of 1,000 firms for 40,000 periods (years).

Table 4: **Capacity utilization and stock returns: empirical**

Portfolio	Value-weighted Mean	Value-weighted SD	Equal-weighted Mean	Equal-weighted SD
Low (L)	13.64	21.23	10.62	21.14
Medium	10.49	16.70	8.20	17.63
High (H)	7.96	20.22	5.18	20.39
Spread	5.67	17.71	5.44	15.51
(L-H)	(2.31)		(2.47)	
α_{CAPM}	4.54		4.74	
(L-H)	(1.94)		(2.20)	

The table reports annual returns of portfolios sorted on capacity utilization, as well as the level and the CAPM alpha (α_{CAPM}) of the spread between the returns of the Low (L) and the High (H) capacity utilization portfolio. Both value- and equal-weighted portfolio returns are reported. The Mean refers to the average annual return, and SD denotes the standard deviation of annual returns. Returns are annualized by multiplying the average monthly return by 12. Parentheses report Newey and West (1987) robust t-statistics. All portfolios are formed at the end of each June and are rebalanced annually. The sample is from July 1967 to December 2015.

Table 5: **Exposure of CU-sorted portfolios to aggregate productivity proxies**

Portfolio	Market Return β	Market Return $t(\beta)$	TFP β	TFP $t(\beta)$	Labor Productivity β	Labor Productivity $t(\beta)$
Low (L)	1.24	(11.74)	0.90	(2.67)	0.66	(2.04)
Medium	1.15	(14.91)	0.51	(2.06)	0.31	(1.28)
High (H)	1.05	(11.59)	0.50	(1.70)	0.31	(1.10)
Spread	0.19	(2.96)	0.40	(2.03)	0.34	(1.97)
(L-H)						

The table reports the exposures of portfolios sorted on capacity utilization to three different aggregate productivity proxies. The exposure regression is: $Ret^e_{i,t} = \beta_0 + \beta_1 \text{Agg-Proxy}_t + \varepsilon_{i,t}$, where $Ret^e_{i,t}$ is the value-weighted excess return of a given portfolio, Agg-Proxy is a variable measuring aggregate productivity, and β is the exposure of interest. Agg-Proxy is measured by the excess return of the market portfolio, or quarterly growth rate of consumption-sector TFP computed by Basu, Fernald and Kimball (2012), and labor productivity growth from the BLS. Monthly returns are aggregated to the quarterly frequency, when appropriate, so that each regression is estimated using quarterly data. The t-statistics associated with each exposure, $\beta(t)$, are computed using Newey and West (1987) standard errors and are reported in parentheses. The sample is from July 1967 to December 2015.

Table 6: **Characteristics of capacity utilization sorted portfolios**

	Low (L)	Medium	High (H)	Diff(L-H)	t(Diff)
Panel A: Portfolio constituents					
CU (%)	67	79	91		
N (Stocks)	617	4423	348		
N (Industries)	5	35	4		
Panel B: Portfolio characteristics					
ME ($b)	0.24	0.23	0.36	-0.11	(-1.03)
BE / ME	1.27	1.15	1.08	0.19	(2.09)
ROA	0.04	0.04	0.03	0.00	(0.71)
GP / Assets	0.34	0.33	0.29	0.05	(1.20)
Asset Growth (%)	5.56	6.68	8.30	-2.73	(-1.36)
I / K	0.06	0.06	0.09	-0.03	(-2.05)
IVOL (%)	2.70	2.44	2.39	0.30	(2.20)
TFP	0.78	0.74	0.81	-0.03	(-0.81)
Hire Rate (%)	2.49	2.04	3.13	-0.63	(-0.26)

The table shows both the composition and characteristics of capacity utilization sorted portfolios. Panel A reports the composition of each portfolio, while Panel B reports industry-level characteristics, averaged across all industries that are assigned to a particular portfolio. All data is annual and is recorded at the end of each June from 1967 to 2015. In Panel A, CU denotes the capacity utilization rate, while N (Stocks) and N (Industries) refer to the average number of individual firms and industries comprising each portfolio, respectively. In Panel B, all statistics are computed as the time-series average of each portfolio's simple firm-level average of a certain characteristic. Details on the construction of each variable are provided in Section OA.3 of the Online Appendix. The column Diff(L-H) refers to the difference between the average characteristics of the low and high capacity utilization portfolios, and t(Diff) is the Newey and West (1987) t-statistic associated with this difference.

Table 7: **Most frequent industry constituents of capacity utilization portfolios**

Industry	Sector	Freq(Extreme)
Panel A: Low capacity utilization portfolio		
Leather and allied product	ND	42.27
Aerospace and miscellaneous transportation eq.	D	41.24
Support activities for mining	MU	37.93
Automobile and light duty motor vehicle	D	29.89
Motor vehicles and parts	D	28.87
Panel B: High capacity utilization portfolio		
Oil and gas extraction	MU	79.31
Plastics material and resin	ND	54.02
Electric power generation, transmission, and distribution	MU	35.05
Mining	MU	34.02
Petroleum and coal products	ND	24.74

The table reports the name of each of the five industries that most frequently populate either the low or the high capacity utilization portfolio. For these industries, the table also reports the frequency, measured as percentage of years over the entire sample period with which each industry is sorted into a particular capacity utilization portfolio. Panel A (Panel B) shows the results for the the low (high) capacity utilization portfolio, and Freq(Extreme) refers to the percentage of years that each industry of interest belongs to the low (high) capacity utilization portfolio. The Sector column reports how each industry is classified into one of three broad categories: durable goods manufacturing (D), nondurable goods manufacturing (ND), or mining or utilities (MU).

Table 8: **Fama-Macbeth regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
CU	-1.66								-1.51	-1.64	-1.53	-1.93	-1.51	-1.42	-1.64	-1.54
	(-2.31)								(-2.13)	(-2.27)	(-2.11)	(-2.44)	(-2.12)	(-2.01)	(-2.07)	(-2.08)
TFP		-1.96							-1.91				-1.66	-1.29	-1.25	0.92
		(-2.28)							(-2.23)				(-1.94)	(-1.46)	(-1.59)	(2.03)
HIRE			-4.56							-4.46			-3.84	-2.99	-7.22	-6.25
			(-3.94)							(-3.91)			(-3.45)	(-2.71)	(-4.24)	(-3.77)
I/K				-3.73							-3.85			-2.12	-2.66	-2.22
				(-3.52)							(-3.78)			(-1.96)	(-2.17)	(-2.29)
OVER					-3.14							-3.18			-3.52	-3.36
					(-3.18)							(-3.14)			(-3.60)	(-3.67)
ln(ME)						-3.74										-2.27
						(-2.93)										(-1.83)
ln(B/M)							3.38									3.59
							(3.86)									(3.97)
RET_{t-1}								-0.23								-1.07
								(-0.16)								(-0.72)
R^2	0.006	0.008	0.005	0.011	0.007	0.026	0.016	0.014	0.014	0.011	0.017	0.015	0.019	0.027	0.037	0.079

The table reports the results of Fama-Macbeth regressions in which future excess returns are regressed on current characteristics. In each year t we run a separate cross-sectional regression in which the dependent variable is a firm's annual excess return from the start of July in year t to the end of June in year $t+1$, and the independent variables are a vector of the firm's characteristics, $\boldsymbol{X}_t$ measured at the end of June in year t:

$$R_{i,t\to t+1} = \beta_0 + \boldsymbol{\beta}_t' \boldsymbol{X}_{i,t} + \varepsilon_{i,t\to t+1} \quad \forall t \in \{1967, \ldots, 2014\}.$$

The characteristics considered are capacity utilization (CU), total factor productivity (TFP), the hiring rate (HIRE), the natural investment rate (I/K), capacity overhang (OVER), the natural logarithm of the market value of equity (ln(ME)), the natural logarithm of the book-to-market (ln(B/M)) ratio, and lagged annual return (RET_{t-1}). After running these cross-sectional regressions we compute the time-series average of each element of the vectors the estimated slope coefficients, $\{\hat{\boldsymbol{\beta}}_t\}_{t=1967}^{2014}$. Each column reports the average slope coefficients for the characteristics of interest. Parenthesis report Newey and West (1987) t-statistics. Columns 1 to 8 of the table show the results when each characteristic is included in a seperate univariate regression. Columns 9 to 15 show the results when a subset of characteristics are used in multivariate regressions. In Column 16 all characteristics of interest are included in the cross-sectional regressions simultaneously. Each control variable is standardized by dividing it by its unconditional standard deviation. The table also report the time-series average of the R^2 obtained from each set of cross-sectional regressions. The first regression is run in 1967 and the last regression is run in 2014, when the TFP data becomes unavailable.

Table 9: **Capacity utilization spread: proxy for firm-level utilization rates**

	Utilization		De-meaned utilization	
Portfolio	Mean	SD	Mean	SD
Low (L)	12.65	19.92	11.92	17.67
Medium	11.98	15.93	11.80	16.50
High (H)	7.66	21.58	6.77	22.08
Spread	4.98	15.39	5.14	15.48
(L-H)	(2.32)		(2.02)	

The table reports the annual value-weighted returns of portfolios sorted on the basis of estimated firm-level capacity utilization rates, as well as the spread between the low (L) and high (H) utilization portfolios. The firm-level proxy for utilization is constructed following the procedure outlined in Section 2.3.3. The table reports the average value-weighted return (Mean) and standard deviation (SD) of each portfolio's returns, and all portfolios are formed by following the procedure described in Section 2.2.2. t-statistics, reported in parentheses, are computed using Newey and West (1987) standard errors. The sample period is between July 1967 to December 2015.

Table 10: **Capacity utilization spread: sorting on growth rates**

	Value-weighted		Equal-weighted	
Portfolio	Mean	SD	Mean	SD
Low (L)	14.49	21.41	11.53	21.92
Medium	10.05	16.63	7.78	17.59
High (H)	9.69	20.59	5.79	20.93
Spread	4.80	16.93	5.74	16.41
(L-H)	(2.00)		(2.45)	

The table reports the annual returns of three portfolios sorted on the basis of capacity utilization growth, as well as the spread between the low (L) and high (H) utilization growth portfolios. The construction of the portfolios is identical to the benchmark analysis, except that portfolios are sorted on the basis of the growth rate of utilization rather than the level of utilization. The growth rate of utilization is measured between March of years t and $t-1$. Mean refers to the average annual return and SD denotes the standard deviation of annual raw returns, and the parentheses report t-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1968 to 2015 and are rebalanced annually, with portfolio returns spanning July 1968 to December 2015

Table 11: **Model-implied moments across alternative calibrations of the model**

Row	Model	Time-series $\sigma_{TS}(ik)$	Time-series $S_{TS}(ik)$	Time-series $\rho_1(ik)$	Cross-sectional $\sigma_{CS}(ik)$	Cross-sectional $S_{CS}(ik)$	Risk premia $E\left[R^{bm}\right]$	Risk premia $E\left[R^{ik}\right]$
	Data							
		0.14	0.67	0.52	0.16	1.89	3.71	3.70
	Panel A: Model without capacity utilization							
	Baseline without utilization							
(1)		0.11	-0.09	0.64	0.07	0.11	2.93	2.30
	Different fixed cost							
(2)	Low FC ($f = 0.01$)	0.12	0.25	0.67	0.07	0.26	2.77	2.52
(3)	High FC ($f = 0.06$)	0.10	-0.75	0.58	0.06	-0.11	3.20	1.96
	Different ϕ							
(4)	Very low ($\phi = 0.75$)	0.19	0.51	0.60	0.12	0.06	2.03	2.04
(5)	Very high ($\phi = 3.00$)	0.06	-0.67	0.65	0.04	0.13	3.72	2.40
	Panel B: Model with capacity utilization							
	Baseline							
(6)		0.14	0.66	0.58	0.11	1.19	3.76	4.27
	Different λ							
(7)	Low ($\lambda = 2.90$)	0.14	0.69	0.58	0.11	1.22	3.78	4.30
(8)	High ($\lambda = 3.10$)	0.14	0.63	0.58	0.11	1.16	3.74	4.24
	Different ϕ							
(9)	Low ($\phi = 1.40$)	0.15	0.65	0.58	0.11	1.15	3.66	4.20
(10)	High ($\phi = 1.60$)	0.14	0.67	0.58	0.10	1.23	3.85	4.34

The table reports model-implied population moments related to the time-series and cross-section of investment rates, as well as risk premia, under various calibrations. The table reports the time-series volatility ($\sigma_{TS}(ik)$), skewness ($S_{TS}(ik)$), and first-order autocorrelation ($\rho_1(ik)$) of firm-level investment rates, as well as the cross-sectional dispersion ($\sigma_{CS}(ik)$) and skewness ($S_{CS}(ik)$) of investment rates. In addition, the table also reports the value premium ($E\left[R^{bm}\right]$) and investment premium ($E\left[R^{ik}\right]$) obtained by sorting the cross-section of model-implied returns association with each calibration on book-to-market ratios and investment rates, respectively. These risk premia are expressed as an annualized percentage. Each alternative calibration is identical to the benchmark calibration in all ways except for altering the specified parameter of interests. The parameters altered are the fixed cost of disinvestment (f), the quadratic capital adjustment cost (ϕ), and the elasticity of marginal depreciation (λ). All moments are based on a simulations of 1,000 firms over 40,0000 periods (years). Finally, the top row of the table also reports the empirical counterpart of each moment.

A Online appendix

OA.1 Additional theoretical results

OA.1.1 Model-implied double sort on book-to-market

This section reports the results of a conditional double sort of model-implied stock returns on book-to-market ratios and capacity utilization rates. The portfolio formation procedure follows the discussion in Section 2.1.2 . The results of the analysis are reported in Table OA.1.1 and show that the utilization premium also exists *within* book-to-market portfolios. The section also discusses the rationale for why our model, which features a single aggregate shock, produces a spread along these two separate dimensions.

Table OA.1.1: **Double-sort in the model**

	Low CU	Medium CU	High CU	Spread (L-H)
Low B/M	12.20	11.60	9.29	2.91
Medium B/M	12.30	11.19	8.44	3.86
High B/M	11.83	9.99	9.66	2.17

The table shows the model-implied equal-weighted returns obtained from a conditional double-sort procedure in which the control variable (i.e., the first dimension sorting variable) is the book-to-marke ratio and the second dimension sort variable is the capacity utilization rate. The portfolios are constructed as follows. First, in each period firms are sorted into three portfolios based on the cross-section of book-to-market ratios from period $t-1$ using the 20^{th} and 80^{th} percentiles of the cross-sectional distribution of book-to-market ratios. Next, within each book-to-market portfolio, firms are further sorted into three additional portfolios on the basis of capacity utilization in period $t-1$ using the 20^{th} and 80^{th} percentiles of the cross-sectional distribution of capacity utilization rates. This procedure produces nine portfolios that are held for one period, and are then rebalanced. The table also shows the capacity utilization spread associated with each book-to-market portfolio. Here, model implied moments are based on one simulation of the model that features 1,000 firms and 40,000 periods (years.)

As shown in Table OA.1.1, the model can produce the utilization premium *within* book-to-market portfolios. There are two reasons why our single-shock model is capable of simultaneously generating a spread along these two separate dimensions. First, despite the comovement between investment, utilization and book-to-market in the model (all relate to Tobin's q), the correlation between the latter two margins is less than perfect. Our model features a real option that induces "wait and see" periods of investment inaction. In these periods utilization and investment do not comove, as utilization substitutes exercising the costly option to disinvest capital. Second, while both utilization and book-to-market are linked to the same aggregate shock, these relations to the aggregate shock are non linear. This occurs, for example, because of time-varying betas and non-linear policy functions. Overall, the model produces enough dispersion in firm-level risk to conduct this double sort.

OA.1.2 Sensitivity of the model for risk premia

In this section we illustrate the intuition for the utilization spread, discussed in Section 2.1.3, by showing the sensitivity of the spread to ingredients (1)–(3) of our model (the quadratic capital adjustment cost, fixed cost of disinvestment, and countercyclical market price of risk, respectively). The results of this analysis are presented in Table OA.1.2, which reports the mean and CAPM alpha of the value-weighted utilization spread, alongside the mean and volatility of the equity risk premium in the model.

The results in rows (2) and (3) show that when the extent of the first friction, the quadratic capital adjustment costs, is perturbed, the magnitudes of the utilization spread and the CAPM alpha change. As this friction is increased in row (3), the magnitude of both the utilization spread and the CAPM alpha increase. With higher adjustment costs, firms can less readily alter the level of their capital stocks, and low utilization implies an even greater underlying capital risk.

Row (4) considers an economy is which the second ingredient, the fixed cost of capital disinvestment, is removed but the remaining two frictions are held constant. The utilization spread still exists, although its magnitude is decreased by almost 1% per annum. The decrease in the utilization spread reflects how removing the fixed cost of disinvestment better allows firms to shed their capital stock instead of substituting disinvestment with temporary declines in utilization. However, the fact that the utilization spread remains sizable indicates that firms still cannot fully absorb productivity shocks into their capital stock.

Finally, rows (5), (6), and (7) consider the role of the third ingredient, the countercyclical market price of risk. In particular, row (6) illustrates how a more countercyclical market price of risk translates into a higher equity risk premium and volatility of aggregate market returns, as well as an increased utilization spread. This occurs because the asymmetry between good and bad aggregate productivity is widened. Row (7) indicates that both the equity risk premium and capacity utilization spread are severely diminished with an acyclical market price of risk.

Table OA.1.2: **Model-implied capacity utilization spread across alternative calibrations of the model**

Row	Model	$E\left[R^M\right]$	$\sigma\left(R^M\right)$	$E\left[R^{CU}\right]$	α_{CAPM}
	Baseline				
(1)		5.39	20.89	5.12	3.89
	Different ϕ				
(2)	Low ($\phi = 1.40$)	5.39	20.85	5.06	3.82
(3)	High ($\phi = 1.60$)	5.39	20.94	5.18	3.95
	No fixed cost				
(4)		5.72	20.39	4.53	3.30
	Different γ_1				
(5)	Low ($\gamma_1 = -8.60$)	5.26	20.48	5.06	3.82
(6)	High ($\gamma_1 = -9.00$)	5.52	21.31	5.18	3.95
(7)	Acyclical ($\gamma_1 = 0$)	1.88	9.94	2.68	1.47

The table reports model-implied population moments under various calibrations. The table reports the equity premium ($E[R^M]$), the volatility of the market return ($\sigma\left(R^M\right)$), the level of the capacity utilization spread ($E[R^{CU}]$), and the CAPM alpha of the utilization spread (α_{CAPM}). Each moment is reported as an annual percentage. and each alternative calibration is identical to the benchmark calibration in all ways except for altering the specified parameter of interests. The parameters altered are the fixed cost of disinvestment (f), the quadratic capital adjustment cost (ϕ), and the cyclicality of the market price of risk (γ_1). All moments are based on a simulations of 1,000 firms over 40,0000 periods (years).

OA.1.3 Discussion of the model's assumptions

The model assumes a countercyclical market price of risk to break the symmetry between high and low utilization firms in the presence of symmetric convex adjustment costs. While we do not micro found the cyclicality, it can arise in a general equilibrium setup by assuming habits preferences or time-varying volatility that is countercyclical (e.g., Campbell and Cochrane (1999) and Bansal and Yaron (2004)).

Additionally, the model only features a real option to disinvest. While we could, in principle, also include a fixed cost for expanding capacity to the model, thereby making investment a real option, we refrain for doing so to keep the dimensionality of the model's parameters low. Since the prior literature emphasizes that the adjustment costs of disinvestment are larger than those of investment (e.g., Zhang (2005)), our model captures this notion in a parsimonious manner. Importantly, Section 3 shows that our model with flexible utilization produces a sizable dispersion in risk premia without relying on large adjustment frictions that distort firm-level investment dynamics (Clementi and Palazzo, 2019).

For parsimony, our framework relies on exposures to a single priced state variable: productivity. Despite having only a single aggregate shock, the model-implied CAPM alpha is non-zero, and the correlation between utilization and investment (or book-to-market) is positive but smaller than one. The former happens because of cyclicality in risk exposures, and the latter happens as utilization can serve as a substitute for disinvestment in downturns. While not necessary for our purpose, featuring additional sources of aggregate risk could further enhance the failure of CAPM, or reduce the model-implied correlation between the utilization premium and other spreads related to intensive-margin characteristics. We qualitatively discuss such an extension in Section OA.1.4 of the Online Appendix by outlining how investment-specific technology shocks, such as those considered by Papanikolaou (2011) and Garlappi and Song (2017b), can interact with the utilization premium, but we leave the quantitative explorations of these richer frameworks to future research.

OA.1.4 Theoretical extensions: intuition regarding the role of IST shocks

In this section we consider how adding investment-specific technology (IST) shocks, such as those studied by Papanikolaou (2011) and Garlappi and Song (2017b), to the economy we outline in Section 1 can interact with the utilization spread.

A positive IST shock implies that the price of firms' capital inputs decreases. As a result, the replacement value of existing capital drops. Under perfect competition this reduces the value of all firms, and firms have a negative exposure to IST shocks. In the presence of real disinvestment options, low utilization firms substitute selling machines by utilizing existing capital less. This substitution makes low utilization firms riskier than high utilization firms, reflecting the frictions in the secondary market for capital (recall Section 2.1.3).

In general, a positive IST shock should encourage firms to raise utilization because capital is now worth less and there is less incentive to preserve it for future periods.

However, firms that are simultaneously subject to a negative idiosyncratic productivity shock still desire to lower utilization to avoid paying the fixed cost of disinvestment. A positive IST shock should enhance this substitution effect, since this shock discourages these firms from selling capital as the sale price declines. Thus, the exposures of low utilization firms to IST shocks should be even more negative than the exposures of high utilization firms. Put differently, in bad states of the world, and without IST shocks, low utilization firms are typically already burdened with unproductive capital. When IST shocks are introduced to the economy, these shocks reduce the value of this unproductive capital further, and hurt firm valuation.

The extent to which these differential exposures to IST shocks enhances or diminishes the utilization premium crucially hinges on the market price of the IST shocks. If the market price of IST shocks is negative (positive), then IST shocks complement (counteract) the baseline results and magnify (reduce) the capacity utilization spread. Given the disagreement regarding the market price of these technology shocks (see, e.g., Papanikolaou (2011), Garlappi and Song (2017a), and Li (2018)), the exact relation between the utilization premium and IST shocks is theoretically ambiguous.

OA.2 Capacity utilization data and summary statistics

The public report on industrial capacity utilization covers 57 industries. These industries are defined at different levels of aggregation ranging from two- to six-digit North American Industry Classification System (NAICS) codes. Specifically, 12 of the industries are crude aggregates that span *multiple* two-digit NAICS codes. For example, one of these 12 aggregates includes the average capacity utilization rate of all manufacturers in the U.S. We remove these 12 crude aggregates from our benchmark sample for two reasons. First, these aggregates do not provide new information as they are spanned by more granularly defined sub-industries that are also included in the sample. Second, these aggregates represents a considerable proportion of total market value and would consequently dominate the returns of the value-weighted portfolios we form in Section 2.2. Removing these 12 crude aggregates leaves us with a benchmark cross-section of 45 industries that features a mix of durable manufacturers, nondurable manufacturers, and miners and utilities.[23]

As the 45 industries included in the benchmark sample are defined from the relatively coarse two-digit NAICS code level to the most granular six-digit NAICS code

[23] A list of these 45 industries, along with each industry's sectoral affiliation, is provided in Table OA.2.3 of the Online Appendix.

level, the benchmark cross-section includes a number of overlapping industries.[24] For instance, the capacity utilization rate of food manufacturers is included in the utilization rate of two industries reported by the FRB: "Food," as well as "Food, beverage, and tobacco." Since removing overlapping industries from our benchmark sample would significantly reduce the number of cross-sectional assets, lower statistical power, and make certain asset pricing tests, such as portfolio double sorts, infeasible, we deal with this overlap in three ways. First, we ensure that our empirical results are valid using both value- and equal-weighted portfolio returns. Second, we verify that our results are not driven by any particular industry that dominates the sample. Third, we also remove the industries that overlap with others and conduct our baseline empirical tests in a sub-sample of 24 distinct industries, each of which corresponds to a unique three-digit NAICS code. Following the example above, this set of distinct industries includes both "Food"' and "Beverage and Tobacco" manufacturers, but excludes the composite index that covers both groups of manufacturers.[25]

Monthly capacity utilization data for 32 industries is available beginning in January 1967, and data for an additional 25 industries becomes available in January 1972.[26] The capacity utilization data we collect ends in December 2015. Consequently, we set the time frame of our analysis from January 1967 to December 2015. As a robustness check we verify that our results also hold when we only consider the most recent half of the sample period, when capacity utilization data is available for the entire cross-section of 45 industries.[27]

OA.2.1 Summary statistics

Below, we describe the properties of the aggregate capacity utilization rate and report summary statistics related to the cross-section of industry-level utilization rates.

Figure OA.2.1 shows the annual growth rate of aggregate capacity utilization over

---

[24] In particular, our final sample consists of one sector defined at the two-digit NAICS level, 27 subsectors defined at the three-digit NAICS level, 13 industry groups defined at the four-digit NAICS level, two industries defined at the five-digit NAICS level, and two U.S. industries defined at the six-digit NAICS level. In Section OA.4 we ensure that our results are robust to this heterogeneity in classification levels.

[25] These distinct industries are listed in Table OA.2.3 of the Online Appendix and correspond to the entries for which the overlap indicator is marked as "No."

[26] There are only 11 monthly time-series reported between January 1948 to December 1966. As eleven industries is a very small cross-section, we do not consider the pre-1967 period in our benchmark sample.

[27] The first year in which data is available for each industry in the sample is reported in the rightmost column of Table OA.2.3 in the Online Appendix.

the sample period. The figure shows that capacity utilization fluctuates significantly over time and that the growth rate of aggregate utilization is procylical. The aggregate utilization rate drops during recessions, particularly during the Great Recession. The growth rate of aggregate capacity utilization tends to slightly lead the business cycle, and has often served as an early warning for recessions. In five out of the seven recessions during our sample period the growth rate of utilization begins to drop prior to the start of the NBER defined recession. The growth rate of capacity utilization increases during the technological revolution of the late 1990's, the housing bubble, and the recovery from the Great Recession.

As illustrated by equation (13), the capacity utilization rate is a combination of both industrial production and capacity. The former variable is studied extensively in the macroeconomic and finance literature, and features prominently in the context of asset pricing. For instance, Cooper, Gulen, and Schill (2008) document a premium for firms with lower total asset growth. The growth rate of assets is directly linked to firms' output, and is consequently captured by the FRB's measure of industrial production. To establish the empirical novelty in examining capacity utilization, we examine the extent to which utilization fluctuates independently of industrial production using the following projection:

$$\Delta CU_t = \beta_0 + \beta_1 \Delta IP_t + \varepsilon_t. \tag{18}$$

Here, ΔCU_t (ΔIP_t) is annual growth rate of aggregate capacity utilization (industrial production), and the residual ε_t captures the component of capacity utilization that is orthogonal to industrial production. Figure OA.2.1 also displays this orthogonal component over the sample period. The dynamics of this orthogonal component do not appear to reflect the dynamics of a white noise process. ε_t is smoother than utilization growth, and changes in ε_t are largely procylical. Similar to utilization growth, ε_t tends to drop during NBER recessions. In some instances the orthogonal component also deviates significantly from capacity utilization growth. For example, during the technology boom of mid-1990's, the orthogonal component declines whereas capacity utilization increases. The orthogonal component drops due to an acceleration in the growth of capacity that was likely facilitated by the technological advancements of the era (Bansak, Morin, and Starr, 2007).

Table OA.2.4 reports summary statistics for the capacity utilization rates of each sector in our benchmark sample. The average rate of aggregate capacity utilization rate is 79.91%. This figure for the U.S. is close to the average capacity utilization rates of 81.17%, 84.69%, and 82.49% for the Euro Zone, China, and Israel, respectively.[28]

[28]See `https://www.dallasfed.org/institute/oecd` for data on these capacity utilization rates,

The majority of the capacity utilization data in the sample pertains to the manufacturing sector, with an almost even split between the number of durable and nondurable manufacturing industries. The mean annual utilization rate is 77.39% (80.09%) for durable (nondurable) manufacturers. Each of these rates is statistically indistinguishable from the average rate of capacity utilization across all industries in the sample. The fact that the average capacity utilization rate of the manufacturing sector, and of the durable and nondurable manufacturing subsectors, is not statistically different from the U.S. average alleviates the concern that our results are driven by ex-ante heterogeneity between sectors.

Among mining industries and utilities the average capacity utilization rate is 84.13%. This average rate is slightly higher than, and statistically different from, the average rate across all industries. Due to this difference in average capacity utilization rates we verify that our empirical results are robust to excluding mining industries and utilities from our sample. We also verify that our results still hold when we conduct tests using the growth rate of capacity utilization that eliminates differences in levels by construction. The results of both of these tests are reported in Section OA.4.

Table OA.2.4 also reports the volatility and autocorrelation of capacity utilization for the different sectors in our sample. The volatility of the capacity utilization rate is comparable across sectors and ranges from 6.67% per annum for mining to 8.29% per annum for durables. The autocorrelation ranges from 0.52 to 0.61, with an all-industry average of 0.58. These statistics affirm the notion that the level of capacity utilization follows similar dynamics regardless of sector.

Overall, Table OA.2.4 shows that the *unconditional* average rate of capacity utilization is only slightly different between sectors. In particular, most differences between the average utilization rate of a sector and the average aggregate utilization rate are statistically indistinguishable from zero. In contrast to these unconditional differences, the asset pricing tests we conduct rely on *conditional* variation in utilization rates. In other words, our tests exploit the fact that the relative ranking of industries in terms of capacity utilization changes over time. Untabulated results show that if we assume that utilization rates are constant over time, and try to utilize the small unconditional differences in the average rate of capacity utilization between industries to perform the empirical tests, our results cease to hold.

Finally, Table OA.2.5 shows the correlation between capacity utilization and other industry-level production-based characteristics for the average industry in our sample.

as recorded by the Organization for Economic Cooperation and Development (OECD) and made available by the Federal Reserve Bank of Dallas.

The characteristics considered include book-to-market, TFP, the hiring rate, sales-to-assets, and the investment rate. While utilization has a positive correlation with productivity, hiring, sales, and investment, these average correlation are fairly low. For example, the average correlation between the investment rate and utilization is only 0.16. This suggests that varying utilization constitutes a separate degree of freedom for managers to smooth dividends, and that any interaction between utilization and expected returns is likely to be independent of the well-established book-to-market and investment rate effects on risk premia.

Table OA.2.3: **Sample composition and industry specification**

Industry name	Sector	Overlap	Start year
Nonmetallic mineral product	D	No	1967
Primary metal	D	No	1967
Fabricated metal product	D	No	1967
Machinery	D	No	1967
Transportation equipment	D	No	1967
Motor vehicles and parts	D	Yes	1967
Aerospace and miscellaneous transportation eq.	D	Yes	1967
Furniture and related product	D	No	1967
Computers, communications eq., and semiconductors	D	Yes	1967
Wood product	D	No	1972
Iron and steel products	D	Yes	1972
Computer and electronic product	D	No	1972
Computer and peripheral equipment	D	Yes	1972
Communications equipment	D	Yes	1972
Semiconductor and other electronic component	D	Yes	1972
Electrical equipment, appliance, and component	D	No	1972
Automobile and light duty motor vehicle	D	Yes	1972
Miscellaneous	D	No	1972
Food, beverage, and tobacco	ND	Yes	1967
Leather and allied product	ND	No	1967
Paper	ND	No	1967
Petroleum and coal products	ND	No	1967
Chemical	ND	No	1967

Continued on the next page...

Table OA.2.3 – Continued from the previous page

Industry name	Sector	Overlap	Start year
Plastics and rubber products	ND	No	1967
Food	ND	No	1972
Beverage and tobacco product	ND	No	1972
Textile mills	ND	No	1972
Textiles and products	ND	Yes	1972
Textile product mills	ND	No	1972
Apparel	ND	No	1972
Apparel and leather goods	ND	Yes	1972
Printing and related support activities	ND	No	1972
Synthetic rubber	ND	Yes	1972
Plastics material and resin	ND	Yes	1972
Artificial and synthetic fibers and filaments	ND	Yes	1972
Mining	MU	No	1967
Metal ore mining	MU	Yes	1967
Nonmetallic mineral mining and quarrying	MU	Yes	1967
Electric power generation, transmission, and distribution	MU	Yes	1967
Electric and gas utilities	MU	Yes	1967
Natural gas distribution	MU	Yes	1967
Coal mining	MU	Yes	1967
Oil and gas extraction	MU	No	1972
Mining (except oil and gas)	MU	No	1972
Support activities for mining	MU	No	1972

The table lists the industries for which capacity utilization data is available at FRED. This set of industries comprises our benchmark sample. For each industry, the table specifies its name, its sector (D denotes the durable sector, ND denotes the nondurable sector, and MU refers to the mining and utilities sector), whether certain industry constituents overlap with other industries in the sample, and the first year in which the industry appears in the sample. All data ends at December 2015.

Table OA.2.4: **Summary statistics of the capacity utilization rate by sector**

Sector	N	Mean	t(Sector-All)	SD	AC(1)
All industries	45	79.91	–	7.38	0.58
Manufacturing	35	78.70	(-0.75)	7.58	0.57
Durable	18	77.39	(-1.58)	8.29	0.52
Nondurable	17	80.09	(0.12)	6.83	0.61
Mining and utilities	10	84.13	(2.35)	6.67	0.60

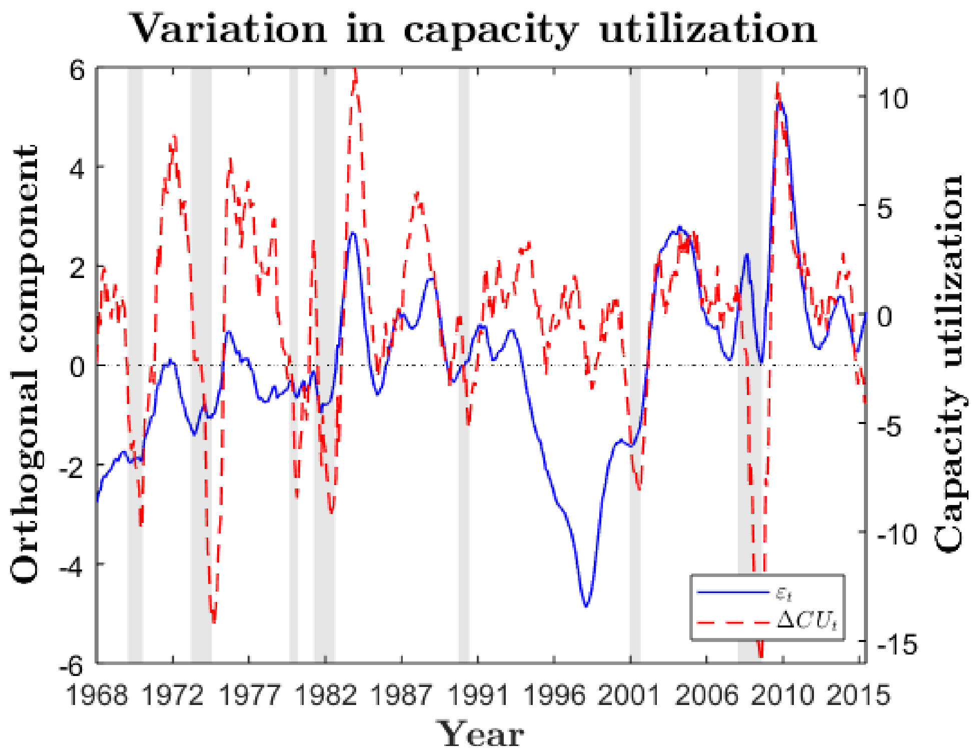
The table reports the mean, standard deviation (SD), and autocorrelation (AC(1)) of the annual capacity utilization rates for each sector in the sample. N represents the number of industries within each sector for which a time-series of capacity utilization data available on FRED. The column t(Sector-All) shows the Newey and West (1987) t-statistic, in parentheses, for the difference in the average capacity utilization rate between the sector denoted in the leftmost column and the average capacity utilization rate across all industries (the top row). The data spans the period 1967 to 2015.

Table OA.2.5: **Average industry-level correlation between production-based characteristics**

	CU	BE / ME	TFP	Hire Rate	Sales / Assets	I / K
CU	1.00	-0.15	0.11	0.13	0.15	0.16
BE / ME		1.00	0.04	-0.26	0.26	-0.00
TFP			1.00	0.23	0.29	0.48
Hire Rate				1.00	0.05	0.53
Sales / Assets					1.00	0.27
I / K						1.00

The table shows the correlation between pairs of industry-level characteristics, averaged over all industries in the sample. The characteristics are the capacity utilization rate (CU), the book-to-market ratio (BE/ME), total factor productivity (TFP), the hiring rate (Hire Rate), the ratio of sales-to assets (Sales / Assets), and the investment rate (I/K). At the end of each June from 1967 to 2015, each industry-level characteristic is constructed as the simple average of the characteristic of interest over all firms that belong to the industry at the point in time. For each industry, we then compute the correlation between industry-level characteristic X and industry-level characteristic Y over the sample period, and report the average value of this correlation across all industries in the sample. The data is annual and runs from 1967 to 2015. Additional details on the construction of each variable are provided in Section OA.3 of the Online Appendix.

Figure OA.2.1: **Capacity utilization: orthogonality from industrial production**



The figure shows the time-series of aggregate capacity utilization growth (dashed red line), as well as the component of capacity utilization growth that is orthogonal to industrial production growth (solid blue line). The orthogonal component of capacity utilization component is obtained from the residuals of the following projection: $\Delta CU_t = \beta_0 + \beta_1 \Delta IP_t + \varepsilon_t$, where CU is aggregate capacity utilization, IP is industrial production, and ε_t is the component of capacity utilization that is orthogonal to industrial production. The horizontal axis shows years and grey shaded regions denote NBER recessions. The right (left) vertical axis represents changes in (the orthogonal component of) capacity utilization. All growth rates are annual, and the sample period ranges from July 1967 to July 2015.

OA.3 Variable description and construction

Asset growth. Asset growth is calculated as the year-on-year annual growth rate of total assets (Compustat Annual item AT) betweens years $t-1$ and t. This definition of asset growth is drawn from Cooper et al. (2008).

Book-to-market. A firm's book-to-market ratio is constructed by following Daniel and Titman (2006). Book equity is defined as shareholders' equity minus the value of preferred stock. If available, shareholders' equity is set equal to stockholders' equity (Compustat Annual item SEQ). If stockholders' equity is missing, then common equity (Compustat Annual item CEQ) plus the par value of preferred stock (Compustat Annual item PSTK) is used instead. If neither of the two previous definitions of stockholders' equity can be constructed, then shareholders' equity is the difference between total assets (Compustat Annual item AT) and total liabilities (Compustat Annual item LT). For the value of preferred stock we use the redemption value (Compustat Annual item PSTKRV), the liquidating value (Compustat Annual item PSTKL), or the carrying value (Compustat Annual item PSTK), in that order of preference. We also add the value of deferred taxes and investment tax credits (Compustat Annual item TXDITC) to, and subtract the value of post retirement benefits (Compustat Annual item PRBA) from, the value of book equity if either variable is available. Finally, the book value of equity in the fiscal year ending in calendar year $t-1$ is divided by the market value of common equity from December of year $t-1$.

Capacity. The capacity estimate measures the maximum amount of output that an industry can produce, assuming the sufficient availability of inputs to production and a realistic work schedule (Board of Govenors of the Federal Reserve System, 2017a). The FRB relies on a variety of sources in order to determine the capacity of each indstry. The primary source of capacity data for manufacturing industries, which make up the bulk of our sample, is currently the Quarterly Survey of Plant Capacity Utilization (QPC). For approximately 20% industries, including a subset of manufacturers, capacity is reported in physical units obtained from government or trade sources, such as the United States Geological Survey. Finally, for a small proportion of industries for which neither of the aforementioned data sources are available, the FRB estimates capacity based on trends through peaks in production. Gilbert, Morin, and Raddock (2000) and Board of Govenors of the Federal Reserve System (2017b) provide a detailed overview of how the FRB measures capacity.

Capacity overhang. We construct a monthly measure of capacity overhang by following the procedure described by Aretz and Pope (2018). In particular we recursively estimate equation (1) of Aretz and Pope (2018) using total assets (Compustat Annual item AT) as our measure of installed capacity.

Gross profitability. Consistent with Novy-Marx (2013), gross profitability is calculated as total revenue (Compustat Annual item REVT) minus the cost of goods sold (Compustat Annual item COGS), divided by total assets (Compustat Annual item AT).

Hiring rate. The hiring rate is computed following Belo et al. (2014). Specifically, the hiring rate in year t is the change in the number of employees (Compustat Annual item EMP) from year $t-1$ to year t, divided by the average number of employees over years $t-1$ and t.

Idiosyncratic volatility. Idiosyncratic volatility is computed in accordance with Ang et al. (2006). At the end of month t, a firm's idiosyncratic volatility over the past month is obtained by regressing its daily excess returns on the daily Fama and French (1993) factors, provided there are at least 15 valid daily returns in the month of interest. Idiosyncratic volatility is then defined as the standard deviation of the residuals from the aforemention regression.

Investment Rate. We follow Stambaugh and Yuan (2017) and compute the investment rate as the change in gross property, plant, and equipment (Compustat Annual item PPEGT) plus the change in inventory (Compustat Annual item INVT) between years $t-1$ and t, divided by the value of total assets (Compustat Annual item AT) in year $t-1$.

Market capitalization. A firm's end of month t market capitalization is computed as the firm's end of month t stock price (CRSP Monthly item PRC) multiplied by the firm's number of shares outstanding (CRSP Monthly item SHROUT).

Natural Investment Rate. Following Belo et al. (2014) the natural rate of investment is computed as capital expenditure (Compustat Annual item CAPX) minus the sales of property, planet, and equipment (Compustat Annual item SPPE) scaled by the average net property, planet, and equipment in years t and $t-1$ (Compustat Annual item PPENT). Missing values of SPPE are set to zero.

Return on assets. Following Imrohoroglu and Tuzel (2014) return on assets (ROA) is computed as net income before extraordinary itmes (Computat Annual item IB) minus preferred dividends (Compustat Annual item DVP), if available, plus deferred income taxes (Compustat Annual item TXDI), if available, all divided by total assets (Compustat Annual item AT).

TechMark. Recalling equation (19), total factor productivity (TFP) is comprised of three distinct components: technology, time-varying markups, and time-varying capacity utilization rates. We isolate the components of TFP related to technology and markups, referred to as "TechMark," as follows. First, we obtain firm-level estimates of the natural logarithm of TFP from Imrohoroglu and Tuzel (2014). We refer to this variable as $\ln(\text{TFP}_{i,t})$. Next, we assign industry-level capacity utilization rates to individual firms by following the matching algorithm described in Section OA.4.1. We take the natural logarithm of these firm firm-level capacity utilization rates, and denote this quantity $\ln(\text{CU}_{i,t})$. Finally, we define the TechMark variable for firm i at time t as $\text{TechMark}_{i,t} = \ln(\text{TFP}_{i,t}) - \ln(\text{CU}_{i,t})$.

Total factor productivity (TFP). The firm-level estimates of TFP are drawn from Imrohoroglu and Tuzel (2014).

OA.4 Additional empirical results

OA.4.1 Independence from value and investment effects

In this section we conditionally sort the sample of industries into portfolios along two dimensions. The first dimension corresponds to either the book-to-market ratio or investment rate, while the second dimension reflects the capacity utilization rate. This methodology allows us to examine the magnitude of the capacity utilization spread while controlling for either the value or the investment premium. We focus on book-to-market ratios and investment rates as these are the only two characteristics in Panel B of Table 6 that are significantly different between the two extreme capacity utilization portfolios and also command a risk premium that is aligned with the utilization spread. Below, we describe the portfolio formation procedure used to undertake this analysis.

Since our cross-section of 45 industries is too narrow to perform double sorts at the industry-level, we perform double sorts at the firm-level. To facilitate this firm-level analysis we need to assign each firm a capacity utilization rate that corresponds to the utilization rate of the industry to which the firm belongs. However, recall from Section OA.2 that our sample is comprised of overlapping industries that are defined with different degrees of granularity. This means that some firms may be matched to more than one industry in our sample. We execute the following matching algorithm to ensure that each firm is matched to the most granularly defined industry to which it belongs.

We start by assigning capacity utilization rates to all firms that belong to a six-digit NAICS code industry for which capacity utilization data is available. We then consider the five-digit NAICS code industries in our sample and identify the constituents of these industries that were not previously assigned a capacity utilization rate. These firms are then assigned a utilization rate corresponding to a five-digit NAICS code industry. This procedure then continues to the four-, three-, and two-digit NAICS code industries, in that order. If a previously unmatched firm belongs to two or more N-digit NAICS code industries, then we assign the firm the utilization rate of its "parent" (N-1)-digit NAICS code industry. Any firms unmatched at the end of this procedure are removed from the sample.

We then compute the book-to-market ratios and the investment rates of firm remaining in the sample using CRSP/Compustat data. Details on the construction of each variable are provided in Section OA.3 of the Online Appendix. We proceed with the bivariate sorting procedure, described below, once all data is computed and assigned to our sample of firms.

At the end of each June from 1967 to 2015 we first sort the cross-section of firms into three portfolios based on either their book-to-market ratios or investment rates. We use the 30^{th} and 70^{th} percentiles of the firm-level cross-sectional distribution of each characteristic to assign each firm to one of three portfolios. We ensure that any accounting data used to form portfolios in this first step has been publicly available for at least four months prior to its use. Next, within each of these three characteristic-sorted portfolios, we further sort firms into three additional portfolios on the basis of capacity utilization. We also use the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of capacity utilization rates in March of the same year to determine portfolio membership in this second step. This process produces nine portfolios that are each held from the beginning of July in year t to the end of June in year $t+1$, at which point in time all portfolios are rebalanced.

Note that the portfolio breakpoints used in the bivariate sorting procedure described above (the 30^{th} and 70^{th} percentiles) differ from those used in our benchmark univariate sorting procedure described in Section 2.2.2 (the 10^{th} and 90^{th} percentiles). This modification is necessary to ensure that there is a sufficient number of firms in each of the nine doubles-sorted portfolios. The cost of these cruder breakpoints is that detecting a relation between capacity utilization and stocks returns after controlling for a potentially confounding characteristic, such as the investment rate, becomes relatively more difficult. It is also important to clarify that even though the second-stage sort is performed at the firm-level, the granularity of the data is still at the industry-level.

Table OA.4.6 reports the results of the bivariate portfolio sorts on the basis of both value- and equal-weighted portfolio returns. The rightmost column of each Panel shows the capacity utilization spread, along with its associated p-value, within portfolios that control for a characteristic of interest. Panels A and B report the results obtained by first controlling for book-to-market ratios, while Panels C and D report the results obtained by first controlling for investment rates. Finally, each Panel of the table also reports the p-value from a joint test on the null hypothesis that the capacity utilization spread across all three characteristic-sorted portfolios is zero.

The results show that after controlling for either book-to-market ratios or investment rate, the capacity utilization spread remains positive in 11 out of 12 cases. The utilization spread is also quantitatively large and statistically significant in most cases. Panel A shows that, keeping book-to-market ratios relatively constant, the equal-weighted capacity utilization spread is significantly different from zero at the 10% level within the low book-to-market portfolio and at the 5% level for both the medium and the high book-to-market portfolios. The joint p-value across the three spread portfolios is under 8%. The value-weighted returns reported in Panel B show that the capacity utilization spread is most pronounced among growth firms. Within this low book-to-market portfolio, the capacity utilization spread exceeds 6% per annum, an average return that is statistically significant at better than the 1% level. While the utilization spread remains positive and significant at the 10% level among medium book-to-market firms, the spread is statistically indistinguishable from zero within the portfolio of value firms. The p-value of 0.016 associated with the joint test in Panel B shows that the three value-weighted utilization spreads are statistically significant after conditioning on book-to-market.

Panels C and D show that, regardless of whether portfolio returns are value-weighted or equal-weighted, the capacity utilization spread typically exceeds 4% per annum within the portfolios of low and medium investment rate firms. In each of these cases the utilization spread is significantly different from zero at better than the 1% level. While the capacity utilization spread is not significant within the high investment rate portfolios, the joint test reported in each Panel is still rejected at the 5% level or better. Panel C (Panel D) shows that, conditioning on investment rates, the three equal-weighted (value-weighted) capacity utilization spreads are jointly and significantly different from zero at the 1% (5%) level.

Overall, the results in Table OA.4.6 suggest that neither the value nor the investment premium is driving the utilization spread. These results complement the Fama and MacBeth (1973) regressions in Section 2.3.2.

Table OA.4.6: **Controlling for book-to-market ratios and investment rates: double-sort analysis**

		Panel A: Capacity Utilization (EW)					Panel B: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
Low (L)	BE/ME	9.40	8.10	6.26	3.14	(p=0.079)	11.63	10.53	5.55	6.08	(p=0.001)
Medium		16.92	13.36	12.82	4.10	(p=0.012)	13.29	10.45	11.01	2.27	(p=0.084)
High (H)		19.89	17.49	16.66	3.23	(p=0.039)	14.47	12.70	14.49	-0.02	(p=0.503)
					Joint test	(p=0.076)				Joint test	(p=0.016)

		Panel C: Capacity Utilization (EW)					Panel D: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
Low (L)	I/K	20.06	17.74	12.62	7.44	($p < 0.001$)	15.52	11.30	10.55	4.97	(p=0.007)
Medium		17.07	13.65	13.08	3.99	(p=0.005)	13.35	10.64	9.07	4.28	(p=0.008)
High (H)		10.60	7.73	9.02	1.59	(p=0.249)	9.96	9.26	7.74	2.23	(p=0.180)
					Joint test	($p < 0.001$)				Joint test	(p=0.045)

The table reports portfolio returns obtained from conditional double-sort procedures, where the controlling variable (i.e., the first dimension sorting variable) is either a firm's book-to-market ratio or investment rate, and the second sorting variable is a firm's rate of capacity utilization. The sorting algorithm is as follows: First, at the end of each June, we sort the cross-section of firms into three portfolios on the basis of either the book-to-market ratio or the investment rate using the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of the characteristic of interest. Second, within each portfolio formed on the basis of the first sorting variable, we further sort firms into three additional portfolios on the basis of capacity utilization, using the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of capacity utilization rates in March of the same year. This process produces nine portfolios that are each held from the beginning of July in year t to the end of June in year $t+1$, at which point in time all portfolios are rebalanced. Portfolio returns are reported for both equal-weighted ("EW", Panels A and C) and value-weighted ("VW", Panels B and D) schemes. The rightmost column of each Panel shows the capacity utilization spread, along with its associated p-value, within portfolios that are first sorted on the controlling variable. These p-values are constructed using Newey and West (1987) standard errors. Each Panel also reports the p-value from a joint test on the null hypothesis that the capacity utilization spread across all three characteristic-sorted portfolios is zero. Panels A and B report the results obtained by first controlling for book-to-market ratios, while Panels C and D report the results obtained by first controlling for investment rates. The sample period is from July 1967 to December 2015.

OA.4.2 Independence from capital overhang

Aretz and Pope (2018) document that firms with higher capital overhang, or firms' whose installed productive capacities exceed their optimal amounts of capacity, have lower expected returns. The authors refer to these firms as possessing "capacity overhang." While Fama and MacBeth (1973) regressions in Section 2.3.2 show that the utilization premium and the overhang spread are empirically distinct, the conceptual similraity between these margins motivatues us to discuss how the notion of capacity utilization materially differs from that of capacity overhang. We also complement the regression analysis by showing that utilization and overhang each have a distinct impact on stock returns using portfolio double sorts.

Recalling equation (13), capacity utilization is defined as the ratio of a firm's actual output to its maximum potential output (its capacity). On the other hand, capacity overhang is the difference between a firm's *installed* capital stock and its *optimal* (value maximizing) level of capital. Intuitively, capacity utilization and capacity overhang are negatively related since a firm that desires to downscale can reduce its output by lowering the utilization of its existing capital. At the same time, the level of the firm's optimal capital stock also drops. If capital adjustments are not frictionless, then these frictions create a wedge between installed and optimal capacity, resulting in capacity overhang. Consequently, capacity utilization tends to decrease at the same time that overhang tends to increase.

The negative correlation between utilization and overhang is neither theoretically perfect nor empirically large in magnitude. Theoretically, the reason for this less than perfect correlation is that low capacity utilization is a result of a *costless and optimal* policy to keep some machines idle.[29] This optimal decision to reduce the utilization of capital does not hinge on any installation frictions or adjustment costs. In contrast, capacity overhang depends crucially on the degree to which investment is irreversible, as influenced by frictions such as convex adjustment costs. While low capacity utilization is optimal in states of low productivity, a non-zero amount of capacity overhang can never represent the first-best outcome for a firm. Consequently, capacity overhang should always be zero in a frictionless economy, whereas capacity utilization may still fluctuate depending on a firm's productivity.

While capacity overhang and capacity utilization are conceptually distinct, the rest of this section examines whether the two effects are also empirically distinct. Since Aretz and Pope (2018) document that high capacity overhang is associated with low expected returns there is no ex-ante reason to believe that the overhang effect is driving the capacity utilization spread. This is because low capacity utilization firms tend to have both high returns and high amounts of capacity overhang. Nonetheless, we perform portfolio double sorts to ensure that the capacity utilization spread is empirically separate from the overhang effect. We show that, controlling for capital adjustment frictions and the degree of irreversibility via the overhang measure of Aretz

[29] Keeping machines idle in bad states is not only costless, but may also benefit the firm by preserving capital for future use in more productive states.

and Pope (2018), the capacity utilization spread survives. Moreover, controlling for the frictionless production decisions represented by capacity utilization, the overhang effect also survives.

To implement this analysis we construct a measure of firm-level capital overhang based on the statistical procedure described by Aretz and Pope (2018), summarized in Section OA.3 of the Online Appendix. Following the discussion on the conceptual relation between capacity utilization and capacity overhang, Table OA.4.7 shows the correlation between overhang and utilization for each industry in our sample.[30] The magnitude of the correlation between the two variables decreases with the degree of aggregation. When we aggregate all firms in our sample, the correlation between capacity utilization and capacity overhang is negative, as expected, and amounts to -0.52. When we compute the correlation between these two variables on an industry-by-industry basis and average these pairwise correlations, the result is a modest average correlation of -0.32. The 95% confidence interval for this cross-sectional correlation shows a high degree of dispersion and ranges from -0.71 to 0.11. Panel C of this table reports that the average firm-level correlation drops to -0.11, and shows that this correlation becomes even more dispersed in the cross-section of firms. These results collectively highlight the fact that while capacity utilization and overhang are conceptually negatively related, the empirical correlation between these two variables is low.

Table OA.4.8 reports the results of performing portfolio double sorts along the dimensions of capacity utilization and capacity overhang using a firm-level analysis as described in Section OA.4.1. Panel A shows the average annual capacity utilization spread within three capacity overhang sorted portfolios when all returns are equal-weighted. The capacity utilization spread is positive and statistically significant within each overhang portfolio. The utilization spread is also jointly significant across all three overhang portfolios. Panel B shows that the results are similar when returns are value-weighted. Panels C and D report that the results are largely similar after changing the order of the sorts. Controlling for capacity utilization, the joint tests in Panels C and D show that the capacity overhang spread is positive and statistically significant on an equal-weighted basis, but is insignificant on a value-weighted basis.

[30]The industry-level capacity overhang measure is obtained by computing the average overhang for all firms that belong to each industry at each point in time. We also note that our sample is only comprised of manufacturing, mining, and utilities firms, whereas the sample of Aretz and Pope (2018) includes the entire Compustat universe, excluding financial firms and utilities.

Table OA.4.7: **Correlation between capacity utilization and capacity overhang**

Panel A: Correlation by Industry		
Industry name	Sector	$\rho_{CU,OVER}$
Food, beverage, and tobacco	ND	-0.741
Printing and related support activities	ND	-0.728
Textile mills	ND	-0.690
Wood product	D	-0.687
Textiles and products	ND	-0.683
Beverage and tobacco product	ND	-0.642
Textile product mills	ND	-0.543
Computer and electronic product	D	-0.537
Food	ND	-0.534
Machinery	D	-0.528
Nonmetallic mineral product	D	-0.502
Support activities for mining	MU	-0.500
Coal mining	MU	-0.472
Computers, communications eq., and semiconductors	D	-0.469
Metal ore mining	MU	-0.450
Communications equipment	D	-0.392
Paper	ND	-0.366
Mining	MU	-0.348
Leather and allied product	ND	-0.345
Transportation equipment	D	-0.342
Mining (except oil and gas)	MU	-0.342
Semiconductor and other electronic component	D	-0.329
Automobile and light duty motor vehicle	D	-0.310
Motor vehicles and parts	D	-0.300
Primary metal	D	-0.254
Artificial and synthetic fibers and filaments	ND	-0.250
Chemical	ND	-0.248

Continued on the next page...

Table OA.4.7 – Continued from the previous page

Panel A: Correlation by Industry

Industry name	Sector	$\rho_{CU,OVER}$
Fabricated metal product	D	-0.214
Electrical equipment, appliance, and component	D	-0.195
Aerospace and miscellaneous transportation eq.	D	-0.183
Computer and peripheral equipment	D	-0.182
Apparel	ND	-0.170
Nonmetallic mineral mining and quarrying	MU	-0.144
Apparel and leather goods	ND	-0.140
Furniture and related product	D	-0.134
Plastics and rubber products	ND	-0.047
Plastics material and resin	ND	0.002
Iron and steel products	D	0.004
Petroleum and coal products	ND	0.071
Miscellaneous	D	0.083
Oil and gas extraction	MU	0.157
Synthetic rubber	ND	0.242

Panel B: Industry-level Summary Statistics

Statistic	Mean	Median	p5	p95
$\rho_{CU,OVER}$	-0.32	-0.34	-0.71	0.11

Pancel C: Firm-level Summary Statistics

$\rho_{CU,OVER}$	-0.11	-0.13	-0.66	0.51

Panel A shows the correlation between industry-level capacity utilization and industry-level capital overhang for each industry in the sample. Overhang at the industry level is computed as the simple average of firm-level overhang rates for all firms that belong to each industry. Panel B reports summary statistics for the industry-level correlations between capacity utilization and capacity overhang that are reported in Panel A. These summary statistics include the cross-sectional mean, median, 5th and 95th percentiles of the distribution of industry-level correlation coefficients. Panel C reports these same summary statistics for firm-level correlations between capacity utilization and capacity overhang.

Table OA.4.8: **Double-sorted portfolios: capacity utilization versus capacity overhang**

		Panel A: Capacity Utilization (EW)					Panel B: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
Low (L)	Overhang	19.27	16.14	15.99	3.28	(p=0.026)	17.03	12.21	11.75	5.28	(p=0.025)
Medium		17.59	14.04	14.54	3.05	(p=0.023)	14.14	10.38	11.21	2.93	(p=0.059)
High (H)		14.10	10.29	8.78	5.32	(p=0.021)	12.84	10.83	7.90	4.94	(p=0.007)
					Joint test	(p=0.061)				Joint test	(p=0.079)
		Panel C: Overhang (EW)					Panel D: Overhang (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
Low (L)	CU	19.00	17.29	13.48	5.52	(p<0.001)	16.03	13.84	12.24	3.79	(p=0.050)
Medium		15.49	14.00	10.98	4.51	(p<0.001)	10.27	10.21	10.24	0.03	(p=0.494)
High (H)		16.31	13.55	9.07	7.23	(p<0.001)	12.29	10.35	8.87	3.42	(p=0.029)
					Joint test	(p<0.001)				Joint test	(p=0.153)

The table reports portfolio returns obtained from conditional double-sort procedures in which one sorting variable is capacity overhang and other sorting variable is capacity utilization. Two cases are considered: in Panels A and B the controlling variable (i.e., the first dimension sorting variable) is overhang, and the variable used in the second-stage sort is capacity utilization. In Panels C and D, the order is flipped: the first (second) stage sorting variable is capacity utilization (overhang). The sorting algorithm is as follows. First, at the end of each June, we sort the cross-section of firms into three portfolios on the basis of the first sorting variable, using the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of the variable of interest. Second, within each portfolio formed on the basis of the first sorting variable, we sort firms into three additional portfolios on the basis of the second sorting variable, using the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of the variable. This process produces nine portfolios that are each held from the beginning of July in year t to the end of June in year $t+1$, at which point in time all portfolios are rebalanced. Both equal-weighted ("EW", Panels A and C) and value-weighted ("VW", Panels B and D) portfolio returns are reported. The rightmost column of each Panel shows the spread on the basis of the second sorting variable, along with the p-value associated with null hypothesis this spread is zero. These p-values are constructed using Newey and West (1987) standard errors. Each Panel also reports the p-value from a joint test on the null hypothesis that the three spreads obtained by forming portfolios in the second stage are jointly equal to zero. The sample period is from July 1967 to December 2015.

OA.4.3 Dissecting the productivity spread

Imrohoroglu and Tuzel (2014) show that low productivity firms earn a high risk premium. While the results of Fama and MacBeth (1973) regressions reported in Table 8 show that productivity cannot explain the utilization premium, this section explores the opposite relation by examining whether capacity utilization can explain the productivity premium. We also examine the explanatory power of the component of productivity that is orthogonal to utilization. These analyses are motivated by the general form of a firm's production function:

$$Y = \underbrace{\text{Technology} \times \text{Markups} \times \text{Utilization}}_{\text{Total factor productivity (TFP)}} \cdot F(K, L), \tag{19}$$

where $F(\cdot)$ is a production function over capital (K) and labor (L). The residual obtained by projecting output on factor-share weighted capital and labor provides an estimate for TFP that can then be decomposed into three elements: technology shocks, time-varying markups, and time-varying capacity utilization. This decomposition of TFP motivates the empirical tests conducted below.

We begin by examining whether the TFP spread exists in our sample of manufacturing firms, mining firms, and utilities. This is necessary because our sample is relatively constrained compared to Imrohoroglu and Tuzel (2014) who examine the TFP spread in the entire Compustat universe, excluding financial and regulated firms. The results of replicating the TFP spread in our subsample of firms are reported in Panel A of Table OA.4.9. The equal-weighted TFP spread amount to 4.22% per annum and is statistically significant.[31]

Since capacity utilization is a fundamental component of TFP, we begin by examining whether the TFP spread survives controlling for capacity utilization. We conduct this analysis using a firm-level dependent double sort as described in Section OA.4.1. In other words, we construct the productivity spread within capacity utilization sorted portfolios. The results are reported in Panel B of Table OA.4.9 and show that the TFP spread is 4.56%, 4.21%, and 1.78% per annum within the portfolio of firms with low, medium, and high rates of capacity utilization, respectively. A joint test on the magnitude of the productivity premium across the three capacity utilization portfolios is statistically significant at the 5% level. This suggests that the productivity premium is distinct from the capacity utilization spread.

Next, we construct a measure for the technology and markup (TechMark) components of TFP by taking the difference between TFP and the capacity utilization rate, as motivated by equation (19).[32] This allows us to isolate the component of TFP that is distinct from capacity utilization and examine the relation between this orthogonal component and stock returns. We sort firms into portfolios based on the TechMark measure at the end of each June and report the results of these univariate sorts in Panel C of Table OA.4.9. The annualized spread between low and high TechMark firms is

[31] The value-weighted TFP spread is positive yet statistically insignificant using our subsample and time frame. For this reason we only focus on equal-weighted returns in this subsection.

[32] Additional details on the construction of this variable are available in Section OA.3 of the Online Appendix.

3.29% and statistically significant.

Taken together, the results above indicate that the TFP premium is driven by two *distinct* underlying spreads: the TechMark and the capacity utilization spreads. Each of these spreads is statistically significant and economically large. We shed light on the contribution of each of these components to the overall productivity spread in Panel D of Table OA.4.9. This panel shows that the correlation between the TFP spread and the capacity utilization (TechMark) spread is 0.39 (0.96). In Table OA.4.10 we project the TFP spread on the utilization spread and find that the adjusted-R^2 is a modest 15%. When the TFP spread is projected on both the utilization spread and the TechMark spread, the adjusted-R^2 increases to 95% and the slope coefficient on the utilization spread remains statistically significant. This means that the majority of the time-series variation in the TFP spread appears to be driven by characteristics related to technology and markups rather than capacity utilization. Overall, this explains why the utilization spread survives controlling for TFP, and vice versa.

Table OA.4.9: **Dissecting the productivity spread**

Panel A: Univariate sorts on TFP

Portfolio	Value-weighted Mean	Value-weighted SD	Equal-weighted Mean	Equal-weighted SD
Low (L)	11.68	23.11	17.00	25.04
Medium	12.25	17.24	15.25	20.23
High (H)	11.16	16.18	12.78	20.73
Spread	0.51	13.79	4.22	12.23
(L-H)	(0.25)		(2.26)	

Panel B: TFP spread controlling for CU

CU	TFP (EW) Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
Low (L)	19.24	17.09	14.68	4.56	(p=0.004)
Medium	16.30	13.81	12.09	4.21	(p=0.007)
High (H)	14.02	15.01	12.24	1.78	(p=0.143)
				Joint test	(p=0.031)

Panel C: Univariate sorts on TechMark

Portfolio	Value-weighted Mean	Value-weighted SD	Equal-weighted Mean	Equal-weighted SD
Low (L)	11.84	22.45	16.89	24.54
Medium	11.64	17.16	15.05	20.28
High (H)	11.66	16.17	13.60	20.91
Spread	0.18	12.90	3.29	11.52
(L-H)	(0.09)		(1.87)	

Panel D: Unconditional correlations

ρ(CU,TFP)	ρ(CU,TechMark)	ρ(TFP,TechMark)
0.39	0.28	0.96

Panel A reports both the annual returns of value- and equal-weighted portfolios formed on total factor productivity (TFP), and the spread between low and high TFP (or productivity) portfolios. Mean (SD) refers to the average (standard deviation) of annual returns, and parentheses report Newey and West (1987) robust t-statistics. Panel B reports equal-weighted portfolio returns obtained from a double sort procedure in which firms are first sorted into three portfolios on the basis of capacity utilization (CU). Within each portfolio, firms are further sorted into three portfolios on the basis of TFP. The rightmost column of the panel show the p-value from a test on the null hypothesis that each TFP spread is zero, as well as a test on null hypothesis that the three spreads are jointly equal to zero. Panel C reports the annual returns of three portfolios sorted on the technology and markups (TechMark) component of TFP. In each of Panels A, B, and C, portfolio breakpoints are based on the 30^{th} and 70^{th} percentiles of the cross-sectional distribution of the characteristic of interest. Panel D shows the pairwise correlations between equal-weighted univariate spreads formed on CU, TechMark, and TFP. The sample period is from July 1967 to June 2015, when the TFP data becomes unavailable. Additional details on the construction of each variable are provided in Section OA.3 of the Online Appendix.

Table OA.4.10: **Projections of the TFP spread on the utilization spread**

	(1)	(2)
β_0	2.50	0.41
	(1.49)	(0.98)
β_{CU}	4.34	1.43
	(8.99)	(6.38)
$\beta_{TFP^{\perp}}$		11.82
		(86.99)
$\bar{R}^2$	0.152	0.945

Panel A reports the slope coefficients from the following regression:

$$Spread_{TFP,t} = \beta_0 + \beta_{CU} Spread_{CU,t} + \varepsilon_{TFP,t},$$

where $Spread_{TFP,t}$ is the productivity spread and $Spread_{CU,t}$ is the utilization spread. Panel B report the coefficients of the following projection:

$$Spread_{TFP,t} = \beta_0 + \beta_{CU} Spread_{CU,t} + \beta_{TFP^{\perp}} Spread_{TechMark,t} + \varepsilon_{TFP,t},$$

where $Spread_{TechMark,t}$ is the Technology/markup spread. Portfolios are formed annually, at the end of each June, following our benchmark portfolio formation procedure, and returns range from July 1967 to June 2015. t-statistics, reported in parentheses, are computed using Newey and West (1987) standard errors.

OA.4.4 Independence from sector-specific effects

Table 7 shows that some durable industries are frequently sorted into the low utilization portfolio, whereas mining industries and utilities frequently exhibit high capacity utilization rates. The former fact raises the concern that the utilization spread may be a manifestation of the durability spread documented by Gomes et al. (2009). That is, the utilization spread may reflect the know fact that durable manufacturers are riskier than nondurable manufacturers. The latter fact raises the concern that the utilization spread is dominated by one particular sector and may reflect ex-ante heterogeneity between different sectors, as opposed to reflecting a risk premium that exists within sectors. We attempt to alleviate both concerns below.

First, it is worth noting that only three (two) of the five industries that are most commonly sorted into the low (high) capacity utilization portfolio are durable (non-durable) manufacturers (see Table 7). Furthermore, the most common industry constituents of the high capacity utilization portfolio are not nondurable manufacturers, as may be expected if the utilization spread were strongly associated with the durability spread. This provides initial evidence that the capacity utilization premium is distinct from the durability spread.

Second, in the left Panel of Table OA.4.11 we examine the capacity utilization spread within a subsample of industries that only includes durable manufacturers. Specifically, we sort the cross-section of 18 durable manufacturers into three portfolios based on the level of capacity utilization by following our benchmark sorting procedure. The capacity utilization spread within this subsample of durable manufacturers amounts to 5.85% per annum, and is statistically significant. Furthermore, the CAPM alpha of the utilization spread within the durables subsector is approximately 6% per annum and is also statistically significant. This demonstrates that the capacity utilization spread is also a within sector phenomenon that is materially unrelated to the ex-ante heterogeneous exposures of durable and nondurable manufacturers to aggregate risk.

Third, we examine the magnitude of the capacity utilization spread when we exclude the only sector that heavily populates the high utilization portfolio: mining and utilities. The mining and utilities sector is also unique in that its average level of capacity utilization over the sample period is statistically different from that of all other industries (see Table OA.2.4). The right Panel of Table OA.4.11 shows the results of sorting all non-mining industries into three portfolios on the basis of capacity utilization. Excluding mining industries and utilities from the sample does not change our baseline results. The capacity utilization spread remains positive, yielding an average return of about 5.3% per annum, and statistically significant.

Table OA.4.11: **Capacity utilization spread: inclusion and exclusion of major sectors**

	Only Durable Sector		Excluding Mining & Utilities Sector	
Portfolio	Mean	SD	Mean	SD
Low (L)	15.08	24.23	14.39	21.63
Medium	10.39	22.11	10.73	17.88
High (H)	9.23	23.77	9.12	20.21
Spread	5.85	19.08	5.27	17.31
(L-H)	(2.13)		(2.12)	
α_{CAPM}	5.97		4.80	
(L-H)	(2.21)		(1.89)	

The table reports the annual returns of portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios when specific sectors are included or excluded from the sample. The left panel shows the results when the sample includes only industries that are classified as durable goods manufacturers. The right panel shows the results when the sample excludes all mining industries and utilities. The table reports the average value-weighted return (Mean) and standard deviation (SD) of each portfolio's returns. t-statistics, reported in parentheses, are computed using Newey and West (1987) standard errors. The sample period is between July 1967 to December 2015.

Taken together, the results of these analyses not only suggest that the capacity utilization spread persists at the firm-level, in line with our model presented in Section 1, but also suggest that the capacity utilization spread is not driven by fixed differences in capacity utilization rates between industries.

OA.4.5 Methodological variations in portfolio formation

In this section we show that the capacity utilization spread is also robust to several implementation choices related to the portfolio formation procedure described in Section 2.2.2.

Variation in breakpoints. In the benchmark analysis we use the 10^{th} and 90^{th} percentiles of the cross-sectional distribution of capacity utilization rates as breakpoints for the low and high utilization portfolios. Here, we modify these breakpoints and sort industries into quintiles instead. This choice of breakpoints doubles the number of industries in each of the extreme capacity utilization portfolios and ensures that the spread is not driven by idiosyncratic factors that may be at work when fewer industries populate the extreme portfolios. The value- and equal-weighted returns of these five portfolios are reported in Table OA.4.12. Despite using coarser breakpoints to form the portfolios, the value-weighted utilization spread is close to 5% per annum and statistically significant. This spread is less than 1% smaller in magnitude than the benchmark spread reported in Table 4. Portfolio returns also tend to decrease as the average utilization rate of each portfolio increases, with the equal-weighted returns monotonically decreasing in average utilization.

Variation in the sample period. While our sample period spans July 1967 to December 2015, we consider the impact of breaking the sample in half and examining the utilization spread in the most recent subsample that starts in July 1991. This subsample analysis is reported in Table OA.4.13 and shows that the magnitude of the spread is larger in the recent subsample than it is over the entire sample period. While the value-weighted spread has a mean return of 5.67% per annum between July 1967 and December 2015, its mean return between July 1991 and December 2015 is 9.09% per annum. As the second half of the sample is populated by two major recessions, the recession of the early 2000s and the Great Recession, this result also shows that the utilization spread is largely countercyclical.

Variation in the sample of industries. Our benchmark results are based on a cross-section of 45 industries. However, as explained in Section OA.2, some of these industries are comprised of firms that belong to multiple industries in the sample. To ensure that our results are not driven by this feature of the data, we repeat our baseline analysis using a subsample of industries whose constituent firms are distinct from one another. These 24 no-overlap industries are listed in Table OA.2.3 and the results of repeating our benchmark portfolio sorts in this subsample of industries are shown in Table OA.4.14. Within this subsample the value-weighted capacity utilization spread is to 6.65% per annum, and even larger than our benchmark spread of 5.7% per annum. Although we lose valuable statistical power by restricting the cross-section of industries, the value-weighted (equal-weighted) utilization spread is still statistically significant at the 5% (10%) level.

Importance of conditional sorting. In untabulated results we demonstrate the importance of the *conditional* portfolio sorting procedure described in Section 2.2.2. Specifically, we consider an alternative procedure in which each industry is permanently assigned to the first portfolio it is sorted into. This *unconditional* portfolio sort leads to a capacity utilization spread that is both economically and statistically insignificant. This result highlights that there is a significant degree of conditional variation in industry-level capacity utilization rates, and that this variation is important for generating the capacity utilization spread.

Table OA.4.12: **Capacity utilization spread: results based on quintile portfolios**

	Value-weighted		Equal-weighted	
Portfolio	Mean	SD	Mean	SD
Low (L)	13.31	18.97	10.23	20.41
2	11.88	17.85	9.74	18.92
Medium	8.78	18.60	7.68	18.57
4	9.25	18.09	7.51	17.52
High (H)	8.44	17.79	5.87	18.64
Spread	4.87	14.07	4.35	11.80
(L-H)	(2.35)		(2.57)	

The table reports annual returns of five portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and the high (H) capacity utilization portfolios. The construction of these portfolios is identical to the benchmark analysis, except that quintile breakpoints used to sort industries into portfolios. Mean refers to the average annual return and SD denotes the standard deviation of annual returns. Parentheses report t-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1967 to 2015 and are rebalanced annually, with portfolio returns ranging from July 1967 to December 2015.

Table OA.4.13: **Capacity utilization spread: results based on the recent subsample**

	Value-weighted		Equal-weighted	
Portfolio	Mean	SD	Mean	SD
Low (L)	15.29	22.74	11.72	21.81
Medium	10.04	16.95	8.24	16.86
High (H)	6.20	20.49	4.39	20.40
Spread	9.09	20.66	7.34	17.53
(L-H)	(2.39)		(2.22)	

The table reports annual returns of portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios. The construction of the portfolios is identical to the benchmark analysis, except that the sample period only includes the recent period from July 1991 to December 2015. Mean refers to the average annual return and SD denotes the standard deviation of annual returns. Parentheses report t-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1991 to 2015 and are rebalanced annually.

Table OA.4.14: **Capacity utilization spread: results based on non-overlapping industries**

	Value-weighted		Equal-weighted	
Portfolio	Mean	SD	Mean	SD
Low (L)	15.43	20.56	11.52	20.25
Medium	10.36	16.70	7.94	18.17
High (H)	8.77	20.24	7.08	20.30
Spread	6.65	18.32	4.45	17.40
(L-H)	(2.51)		(1.77)	

The table reports annual returns of five portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios. The construction of the portfolios is identical to the benchmark analysis, except that the sample of industries is restricted to those industries whose constituents do not belong to multiple industries in the sample (see Table OA.2.3, Column 3, for the list of these non-overlapping industries). Mean refers to the average annual return, SD denotes the standard deviation of annual returns. Parentheses report t-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1967 to 2015 and are rebalanced annually, with portfolio returns spanning July 1967 to December 2015.

OA.4.6 Supplemental tables

Table OA.4.15: **Transition matrix of constituents between capacity utilization portfolios**

Portfolio in	Portfolio in year $t+1$		
year t	Low	Medium	High
Low	0.746	0.254	0.000
Medium	0.033	0.939	0.027
High	0.011	0.232	0.758

The table shows the probability of an industry sorted into portfolio $i \in \{\text{Low, Medium, High}\}$ in year t, where i is the row index, being sorted into portfolio $j \in \{\text{Low, Medium, High}\}$ in year $t+1$, where j is the column index. The transition probabilities are computed using annual capacity utilization data from June 1967 to December 2015. Industries are sorted into portfolios at the end of each June following the portfolio formation procedure described in Section 2.2.2.

Table OA.4.16: **Value-weighted capacity utilization spread and factor models**

	(1)	(2)	(3)	(4)
MKTRF	0.175	0.148	0.189	0.165
	(2.72)	(2.48)	(2.78)	(2.59)
SMB	0.155	0.153	0.069	0.039
	(1.94)	(1.88)	(0.70)	(0.45)
HML	0.074	0.022	-0.075	
	(0.66)	(0.21)	(-0.46)	
UMD		-0.144		
		(-1.76)		
RMW			-0.241	
			(-1.55)	
CMA			0.294	
			(1.17)	
I/A				0.279
				(1.73)
ROE				-0.411
				(-3.01)
α	4.013	5.578	4.235	5.892
	(1.61)	(2.24)	(1.60)	(2.26)
R^2	0.033	0.045	0.047	0.072

The table reports the results of time-series regressions of the value-weighted capacity utilization spread (the portfolio that buys low capacity utilization industries and shorts high capacity utilization industries) on a number of common risk factors. Parameter estimates are obtained by regressing monthly excess returns on each set of monthly risk factors. Each reported α is annualized by multiplying the equivalent monthly coefficient by 12. MKTRF is the excess return of the market portfolio. SMB and HML are the size and value factors of the Fama and French (1993) three-factor model, while MOM is the momentum factor of Carhart (1997). RMW and CMA correspond to the profitability and investment factors of the Fama and French (2015) five-factor model. Finally, I/A and ROE denote the investment and profitability factor in the Hou et al. (2015) q-factor model. t-statistics are computed using Newey and West (1987) standard errors and are reported in parentheses. Returns span July 1967 to December 2015.

Table OA.4.17: **Equal-weighted capacity utilization spread and factor models**

	(1)	(2)	(3)	(4)	(5)
MKTRF	0.121	0.117	0.094	0.131	0.109
	(2.18)	(2.17)	(1.96)	(2.32)	(2.08)
SMB		0.226	0.225	0.161	0.130
		(3.38)	(3.32)	(1.88)	(1.74)
HML		0.229	0.186	0.077	
		(2.43)	(2.19)	(0.54)	
UMD			-0.117		
			(-1.54)		
RMW				-0.187	
				(-1.55)	
CMA				0.275	
				(1.35)	
I/A					0.445
					(3.46)
ROE					-0.382
					(-3.27)
α	4.735	3.329	4.604	3.395	4.629
	(2.20)	(1.54)	(2.07)	(1.52)	(2.07)
R^2	0.013	0.047	0.057	0.059	0.100

The table reports the results of time-series regressions of the equal-weighted capacity utilization spread (the portfolio that buys low capacity utilization industries and shorts high capacity utilization industries) on a number of common risk factors. Parameter estimates are obtained by regressing monthly excess returns on each set of monthly risk factors. Each reported α is annualized by multiplying the equivalent monthly coefficient by 12. MKTRF is the excess return of the market portfolio. SMB and HML are the size and value factors of the Fama and French (1993) three-factor model, while MOM is the momentum factor of Carhart (1997). RMW and CMA correspond to the profitability and investment factors of the Fama and French (2015) five-factor model. Finally, I/A and ROE denote the investment and profitability factor in the Hou et al. (2015) q-factor model. t-statistics are computed using Newey and West (1987) standard errors and are reported in parentheses. Returns span July 1967 to December 2015.

OA.5 Numerical model solution

To solve the model numerically we use value function iteration. The value function and the optimal policies implied by the firm's maximization problem in equation (10) are solved on a grid in a discrete state space. The grid for capital stock, K, features 501 grid points, with the endpoints of the grid chosen to be nonbinding. The aggregate productivity process, x, and the idiosyncratic productivity process, z, are each driven by an independent and identically distributed (i.i.d.) normal distribution. While each of these state variables has continuous support in the model, each variable needs to be transformed into a finite number of states in order to implement the numerical solution algorithm. We use the method of Tauchen and Hussey (1991) to discretized the z process into 11 states. Because the method of Tauchen and Hussey (1991) does not work well for persistent processes, namely those with a persistence parameter greater than 0.90, we use the method of Rouwenhorst (1995) to discretize x into 5 states. Once the discrete state space has been constructed, conditional expectations are computed using matrix multiplication and the firm's maximization problem is solved using a global search routine. All results are robust to choosing finer grids.