The Utilization Premium*

Fotis Grigoris and Gill Segal

May 2022

Abstract

We study the interaction of flexible capital utilization and depreciation for expected returns and investment of firms. Empirically, an investment strategy that buys (sells) equities with low (high) utilization rates earns 5% p.a. Utilization predicts excess returns beyond other production-based variables. We reconcile this novel utilization premium quantitatively using a production model. The model suggests that flexible utilization is important for matching the cross-sectional distribution of investment and stock prices jointly. A model without flexible utilization yields many counterfactuals that flexible utilization addresses by making depreciation fluctuate endogenously. Overall, utilization tightens the link between firms' production and valuations.

JEL classification: G12, E23, E32

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

*Grigoris: Kelley School of Business, Indiana University, Hodge Hall, Bloomington, IN 47405, U.S.A. (e-mail: fgrigori@iu.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 by two anonymous referees, an anonymous associate editor, the decision editor (Lukas Schmid), and suggestions by Max Croce, Ric Colacito, Winston Dou (discussant), Eric Ghysels, Erik Loualiche (discussant), Christian Lundblad, Jun Li (discussant), Marcelo Ochoa, Dimitris Papanikolaou, Nick Roussanov, Jincheng Tong (discussant), Miao Ben Zhang (discussant), Harold Zhang, and seminars participants at the 2019 MFA annual meeting, 2019 Northern Finance Association meeting, 2019 Fall UT Dallas Finance Conference, 2020 RAPS Winter Conference, 2020 Australasian Finance & Banking Conference, 2021 Eastern Finance Association annual meeting, and the University of North Carolina at Chapel Hill. All errors are our own.

Capacity utilization measures the extent to which a business uses its production potential. Flexible capacity utilization lets firms scale their production by choosing how much of their machinery to operate. For instance, instead of decreasing production by selling machines, the firm can choose to keep some machines idle. While existing studies show the ability of aggregate-level utilization to predict the business cycle, the extent to which granular-level (i.e., firm or industry level) utilizations quantitatively affect risk and investment remains largely unexplored. In this paper we examine this relationship empirically and theoretically, and show that it is not only sizable but also bears key implications for explaining the joint distribution of cross-sectional real quantities and prices.

First, we show that lower utilization is associated with a substantially higher risk premium in the cross-section of equities. Second, we construct a production economy that highlights the role of flexible utilization for the intertemporal choice of capital. The model serves two goals: (1) it explains the relation between utilization and risk premia quantitatively, and (2) it shows that flexible utilization is crucial for production models with real options to jointly target cross-sectional investment moments and risk premia. When utilization changes from fixed to flexible, the dispersion and skewness of investment rates rise. This increases the dispersion of firms' exposures to aggregate productivity. Thus, a model with flexible utilization generates large cross-sectional variation in expected returns while relying on lower exogenous capital adjustment costs parameters.

Empirically, we start by establishing two novel facts. Using capacity utilization data for a cross-section of industries, we document that firms that belong to low capacity utilization industries earn an average annual return that is 5.7% higher than the annual return earned by firms that belong to high capacity utilization industries. We term this return spread the *Utilization Premium*. We then show that there exists a monotonically decreasing relation between utilization rates and aggregate productivity exposures: the low utilization portfolio has a higher aggregate productivity beta than the high utilization portfolio, consistent with the premium. Moreover, we find that exposures to aggregate productivity are time-varying and depend negatively on the utilization rate. The utilization premium is distinct from related production-based spreads. For instance, Fama and MacBeth (1973) regressions show that utilization's explanatory power for risk premia is incremental to key characteristics, such as investment, hiring, book-to-market, and productivity.[1]

To rationalize our findings, we incorporate the realistic feature of flexible utilization decisions into a quantitative production model. The model yields two general implications. The first im-

[1]Notably, while the baseline utilization premium is based on industry-level return data, the premium does not simply reflect cross-sector heterogeneity. The premium also exists *within* economic sectors (e.g., among durable manufacturers only). Moreover, we construct proxies for firm-level utilization rates using Compustat data and show that the premium persists when we sort *firms* in portfolios based on these firm-level utilization proxies.

plication relates to the utilization premium. The model is able to quantitative replicate the novel facts. In particular, the magnitude of the utilization premium in the model matches the data.

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, making selling machines a real option. The key ingredient is a variable capacity utilization rate that controls the extent to which installed capital is utilized. Increasing the utilization rate is costly, as it makes capital depreciate faster.

If the capacity utilization rate is fixed, then firms can only reduce the cyclicality of their payouts via investment decisions. The risk of each firm is determined entirely by the interaction between aggregate productivity and these capital adjustment costs.[2] With flexible utilization, firms have an additional mechanism to decrease the cyclicality of productivity shocks on their payouts.

To illustrate how flexible utilization is tied to a firm's risk, consider an economy with only convex and symmetric capital adjustment costs. A firm operating in a low productivity state has the incentive to reduce its capital, thereby exposing itself to potentially large adjustment costs. Simultaneously, the firm has an incentive to lower its utilization rate. 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] Similarly, 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 seeking to make large changes to their capital under adjustment costs. Simultaneously altering utilization partially hedges their (dis)investment policies.

Two mechanisms break the symmetry between high and low utilization firms. First, a fixed adjustment cost causes disinvestment to become a real option. Less productive firms *substitute* disinvestment by lowering utilization. Instead of selling capital, firms temporarily downscale by under-utilizing their machines. As the friction for selling capital is higher for these firms, they are riskier. This real option also decouples firms' investment policies from their utilization policies, yielding a distinction between the utilization premium and investment-based spreads.[4] Second, the model features a countercyclical market price of risk. Thus, firms' whose valuations covary more

[2] Firms that disinvest (invest) the most in low (high) aggregate productivity states are required to pay large capital adjustments costs, making them riskier in these states.

[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.

[4] In the Online Appendix we extend the model to include priced investment efficiency shocks. These shocks affect the value premium but do not affect the utilization premium, leading to a starker distinction between the two spreads.

with productivity during bad states (i.e., low utilization firms) command a larger premium.

The second general implication of the model is that flexible utilization plays a pivotal role for simultaneously matching investment and asset-pricing moments in the presence of real investment options. With fixed utilization, the model-implied cross-sectional skewness and volatility of investment are too low, the time-series skewness of investment is counterfactually negative, and risk premia are too small. The broad intuition behind these model misses is that under a real options setup there are many periods of investment inaction in which firms do not alter their capital stocks. This inaction distorts investment's moments. Since a mass of "waiting" firms all choose the same investment policy (equals to constant depreciation rate), the entire distribution of investment rates and risk exposures becomes too compressed. Flexible utilization overcomes these model misses via two channels. First, it induces the depreciation rate to become firm-specific and time-varying. With varying depreciation, periods of investment inaction are eliminated. The distribution of investment and risk exposures also becomes more dispersed, in line with the data. Second, flexible utilization substitutes the need to rely on a high degree of exogenous adjustment costs to jointly target investment and asset-pricing moments.

Related literature. The paper contributes to the literatures on the role of capacity utilization in RBC models, costly reversibility, and production-based asset pricing. The implications of capacity utilization for asset prices has received considerably less attention than the role of flexible utilization for explaining the business cycle in the macroeconomic literature. Of the small set of papers that study capacity utilization in the context of asset pricing, most focus on aggregate moments. Garlappi and Song (2017) include capacity utilization in a production-based asset pricing model and show that varying utilization is important for the market price of risk of investment-specific technology (IST) shocks. Da, Huang, and Yun (2017) use industrial electricity usage as a proxy for utilization and find that higher electricity usage predicts lower stock market returns in the future. This latter result is broadly consistent with our utilization premium, but pertains to the time-series of market returns rather than the cross-section of equities that we study.

The model in Cooper, Wu, and Gerard (2005) focuses on explaining the value premium and also includes capacity utilization. Although the authors find a *qualitatively* negative relation between utilization and stock returns using OLS regressions, we emphasize the *quantitative* relation between utilization and risk. We do this both theoretically, via a calibrated model, and empirically, by establishing a novel spread. The utilization spread is distinct from the value premium, and a host of other production-based characteristics. Our analysis also illustrates the importance of flexible utilization for the joint distribution of investment rates and asset prices.

The studies of Zhang (2005), Carlson, Fisher, and Giammarino (2004) and Cooper (2006), among others, explain cross-sectional spreads by assuming that capital is partially irreversible. Recent literature considers whether this irreversibility produces realistic distributions of investment rates and risk premia jointly. While Clementi and Palazzo (2019) present evidence that investment may not be as irreversible as these models suggest, Bai, Li, Xue, and Zhang (2019) show that few firms disinvest capital. The importance of utilization for jointly matching investment and prices extends beyond these two studies, as we consider additional frictions generated by costly real investment options. In the presence of real options, flexible utilization is key for producing a realistic distribution of investment rates and sizable risk premia. Thus, real options are quantitatively important for disentangling the utilization premium from the value premium.

More broadly, our paper is related to asset-pricing studies that connect production economies to expected returns (e.g., Belo and Lin (2012), Jones and Tuzel (2013), Kuehn and Schmid (2014), Belo, Li, Lin, and Zhao (2017), Kilic (2017), Tuzel and Zhang (2017), Ai, Li, Li, and Schlag (2019), Dou, Ji, Reibstein, and Wu (2019), Loualiche et al. (2019), Kogan, Li, and Zhang (2019)). While Belo, Lin, and Bazdresch (2014) find that firms with low hiring rates earn high returns, hiring rates are indistinguishable between low and high utilization industries. Likewise, differences in neither intangible capital (e.g., Eisfeldt and Papanikolaou (2013)), capital overhang (e.g., Aretz and Pope (2018)), nor financing costs (e.g., Belo, Lin, and Yang (2018)) explain the utilization premium. Imrohoroglu and Tuzel (2014) examine firm-level total factor productivity (TFP) and show that low TFP firms earn a productivity premium. While utilization is a component of TFP, we show empirically that most of the productivity premium stems from the technology and markup components of TFP. Controlling for utilization (TFP), the productivity (utilization) premium persists.

1 Empirical evidence

1.1 Data

Capacity utilization. We obtain industry-level 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 (i) quantify how effectively different industries are utilizing factors of production, and (ii) to assess inflationary pressures (e.g., Corrado and Mattey (1997)). 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:

$$CU_{i,t} = \frac{IP_{i,t}}{Capacity_{i,t}}. \tag{1}$$

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.

Our benchmark cross-section encompasses 45 industries, featuring a mix of durable manufacturers (18 industries), nondurable manufacturers (17 industries), and mining and utilities (10 industries).[5] The time period of our benchmark analysis ranges from January 1967 to December 2015.[6] 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 varies over time. We provide further details on the sample composition in Section OA.2 of the Online Appendix.

Returns data. Stock return data are taken from CRSP and accounting data are taken from the CRSP/Compustat Merged Fundamentals Annual file. We obtain returns for portfolios sorted on key characteristics 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.[7] Variable definitions are provided in Section OA.1 of the Online Appendix.

1.2 Portfolio formation

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 March utilization data and the June sort date ensures that this strategy is tradeable, as all data used are publicly available by the portfolio formation dates.[8] Each portfolio is then held

[5] While the FRB's utilization data mainly covers manufacturing industries, we stress that: (1) this sector continues to play an important role in the aggregate economy, as these industries constitute the industrial production index and are key for long-term growth. For instance, Andreou, Gagliardini, Ghysels, and Rubin (2019) show that while the manufacturing sector has diminished over time, the sector continues to explain about 61% of GDP growth. Thus, our sample encompasses an important segment of the economy; (2) the industries covered by the FRB generally reflect good producers that utilize capital, ensuring a tight link between the production model in Section 2 and the data.

[6] The start date is based on the availability of capacity utilization data by the FRB. Notably, Table OA.3.2 in the Online Appendix shows that our results are strengthened in the second half of the sample period.

[7] We thank Kenneth French, Lu Zhang, and Selale Tuzel for making this data available to us.

[8] 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 $t+1$. Since 1967,

from July of year t to the end of June of year $t+1$, at which time all portfolios are rebalanced.

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. We focus on these breakpoints to increase the power of our asset-pricing tests. This is useful because our ability to detect a relation between utilization and future stock returns is already limited by the cross-section of industries for which the FRB reports utilization data. However, since each portfolio contains multiple industries, each of which is comprised of many firms, this choice of breakpoints produces three well-diversified portfolios. We discuss the composition of the portfolios and their characteristics in Section 1.5.[9]

1.3 Fact I: Utilization portfolios and expected returns

Table 1 reports the annual value- and equal-weighted returns of portfolios sorted on capacity utilization. We document an economically and statistically significant spread between the returns of the low and high utilization portfolios. We define the *Utilization Premium* as the average return differential between the low and high utilization portfolios. The table also shows that portfolio returns are monotonically decreasing in the average rate of capacity utilization.

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 extreme utilization portfolios are 5.67% and 5.44% per annum, respectively. Each spread is significant at the 5% level.[10]

1.4 Fact II: Utilization portfolios and productivity exposures

Fixed exposures. We check whether the monotonic pattern between utilization rates and expected return is a result of differential exposures to fundamental macroeconomic risk, as captured by aggregate productivity betas. We consider the following projection

$$Ret^e_{i,t} = \beta_{0,i} + \beta_{1,i}\text{Agg-Prod}_t + \varepsilon_{i,t}, \tag{2}$$

where $Ret^e_{i,t}$ is the value-weighted excess return of portfolio i, Agg-Prod$_t$ is a proxy for aggregate

March utilization rates have been publicly available by April 17th at the latest.

[9]Table OA.3.16 in the Online Appendix reports the portfolio transition matrix and shows that the probability of transitions out of the extreme portfolios is relatively high (about 25%). This demonstrates the importance of the conditional portfolio rebalancing procedure, as industries change their relative utilization ranking over time.

[10]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.

Table 1: **Capacity utilization and stock returns: Data**

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)	

The table reports annual returns of portfolios sorted on capacity utilization, as well as the spread between the returns of the Low (L) and the High (H) 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 returns. Returns are annualized by multiplying the average monthly return by 12. Parentheses report Newey and West (1987) t-statistics. The sample is from July 1967 to December 2015.

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, utilization-adjusted TFP growth from Fernald (2012), and labor productivity from the Bureau of Labor Statistics (BLS). As the latter two proxies are only available quarterly, we aggregate monthly returns to the quarterly frequency when estimating equation (2). Note that when we proxy for productivity using excess market returns, β_0 corresponds to the CAPM alpha.

Table 2 reports the results. The exposure of the low utilization portfolio to aggregate productivity is higher than the exposure of the high utilization portfolio, regardless of the productivity proxy. The differences in the productivity betas of the low and high utilization portfolios are statistically significant at roughly the 5% level. The table also shows that the intercepts from projecting the utilization spread on each productivity proxy are generally insignificant. In particular, the CAPM alpha is insignificant at the 5% level (t-statistic of 1.78).[11]

Time-varying exposures. The CAPM alpha is insignificant at the 5% level, but significant at the 10% level. A marginally significant CAPM alpha can arise due to time-varying exposures to aggregate productivity. We verify this conjecture in Section OA.3.2 of the Online Appendix. We show that a conditional single-factor model *fully* absorbs the utilization premium (using the Lewellen and Nagel (2006) methodology and non-linear model specifications). We also establish a negative relation between time-varying exposures and utilization rates.

Why do exposures to aggregate productivity negatively correlate with utilization? Is the spread

[11] β_0 can be interpreted as test of the CAPM only when we proxy for aggregate productivity using the market return, but it is not restricted to zero under the null hypothesis of the CAPM when using TFP or labor productivity (as these are non-tradable factors). Nonetheless, the fact that β_0 is insignificant in these two cases complements the CAPM alpha, and emphasizes that statistically, the premium is absorbed by variations in productivity. Moreover, for all proxies used, $\beta_{1,i}$ is still informative about the degree to which utilization is related to aggregate productivity.

Table 2: **Exposure of CU-sorted portfolios to aggregate productivity proxies**

	Market returns		Util.-adjusted TFP		Labor productivity	
Portfolio	β	$t(\beta)$	β	$t(\beta)$	β	$t(\beta)$
Low (L)	1.33	(10.09)	1.17	(2.67)	0.95	(1.87)
Medium	1.23	(12.15)	0.75	(2.18)	0.60	(1.41)
High (H)	1.07	(7.68)	0.77	(1.93)	0.60	(1.28)
Spread (L-H)	0.26	(3.50)	0.40	(1.93)	0.34	(1.97)
Intercept	4.36	(1.78)	4.60	(1.72)	3.34	(1.18)

The table reports the exposures of portfolios sorted on capacity utilization to three different aggregate productivity proxies. The 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 portfolio i, Agg-Proxy_t is a proxy of aggregate productivity, and β_1 is the exposure of interest. Agg-Proxy_t is either (i) excess market returns, (ii) utilization-adjusted TFP growth from Fernald (2012), or (iii) labor productivity growth from the BLS. Monthly returns are aggregated to the quarterly frequency so that each regression is estimated using quarterly data. Newey and West (1987) t-statistics are reported in parentheses, and "Intercept" refers to the annualized value of β_0. The sample spans July 1967 to December 2015.

in these exposures sufficient to quantitatively explain the utilization premium? To answer these question, we develop a model in Section 2 which endogenizes firms' exposures to aggregate productivity, based on their utilization policies. Given the evidence above, the model features a single source of aggregate risk, but we relax this assumption in Section 3.3.

1.5 Capacity utilization portfolios: Characteristics

Portfolio constituents. Panel A of Table 3 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 in our sample. Although the number of industries falling into these extreme portfolios is small, these industries are comprised of roughly 960 firms. Thus, the low and high utilization portfolios collectively contain about 18% of all firms in our merged CRSP/Compustat sample, and the extreme portfolios are well diversified. The average utilization rate is monotonically increasing from the low to the high utilization portfolio.

To shed light on the industries underlying each portfolio, Table 4 reports the five industries that populate the extreme utilization portfolios most often. For each industry, the table reports the sector to which the industry belongs and the proportion of years the industry is sorted into the portfolio. There is a large degree of sectoral variation associated with the industries that 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 often contains mining industries, utilities, and nondurable manufactures.[12] These results provide suggestive evidence that the utilization premium is not

[12]While oil extraction frequently appears in the high utilization portfolio, Section OA.3.7 in the Online Appendix

Table 3: **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 $(Ind.)$	5	35	4		
		Panel B: Portfolio characteristics			
ME ($b)	1.02	0.81	1.08	-0.06	(-0.59)
BE / ME	1.34	1.19	1.13	0.21	(1.81)
ROA	0.02	0.01	0.01	0.00	(0.27)
GP / Assets	0.35	0.34	0.30	0.05	(1.27)
Asset Growth (%)	10.01	11.25	13.04	-3.03	(-1.08)
Inventory Growth (%)	10.29	10.59	12.62	-2.34	(-0.37)
I / K	0.06	0.06	0.09	-0.03	(-2.03)
IVOL (%)	3.18	2.89	2.86	0.32	(1.96)
TFP	0.76	0.73	0.78	-0.02	(-0.51)
Hire Rate (%)	3.95	3.48	5.80	-1.85	(-0.68)
R&D / ME	0.06	0.05	0.05	0.01	(0.97)
OC / AT	1.31	1.21	0.73	0.57	(1.12)
Leverage	0.26	0.24	0.26	-0.01	(-0.76)
Debt growth (%)	3.64	2.80	5.32	-1.68	(-1.38)
Equity issue. (%)	4.78	5.19	5.38	-0.60	(-1.46)

The table shows both the composition and the 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 $N(Ind.)$ 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.1 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.

driven by any one sector in particular. Section 1.7.3 provides more rigorous evidence that the utilization spread is mostly a within-sector, rather than a cross-sector phenomenon.

Portfolio characteristics. Panel B of Table 3 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 ROA or gross profitability), asset growth, inventory growth, external financing frictions (as measured by leverage, debt growth, and equity issuance rates of Belo et al. (2018)), or intangible capital (as measured by R&D/ME or organizational capital as in Lin (2012) and Eisfeldt and Papanikolaou (2013), respectively). The low portfolio has both lower industry-level TFP and hiring rates than the high portfolio. However,

demonstrates that the utilization premium remains positive and significant if we exclude the entire mining sector.

Table 4: **Most frequent industry constituents of capacity utilization portfolios**

Industry	Sector	Freq.	Industry	Sector	Freq.
Panel A: Low capacity utilization portfolio			Panel B: High capacity utilization portfolio		
Leather and allied product	ND	42.27	Oil and gas extraction	MU	79.31
Aerospace and transport eq.	D	41.24	Plastics material & resin	ND	54.02
Support activities for mining	MU	37.93	Electric power transm.	MU	35.05
Automobile and motorvehicle	D	29.89	Mining	MU	34.02
Motor vehicles and parts	D	28.87	Petroleum and coal	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 utilization portfolio. Panel A (Panel B) shows the results for the low (high) capacity utilization portfolio, and Freq. 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).

these differences are small and statistically insignificant.

There are differences between the two extreme portfolios in the book-to-market ratios, investment rates, and idiosyncratic return volatilities (IVOL). The latter difference in IVOL cannot account for the capacity utilization premium, as Ang, Hodrick, Xing, and Zhang (2006) show that high IVOL firms earn low expected returns, but low utilization firms have higher IVOL. However, the former differences raise a concern 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.[13] We address this concern in the next subsection.

1.6 Fama-MacBeth analysis

To establish a degree of independence between the utilization spread and the value and investment premia, we conduct a Fama and MacBeth (1973) analysis. We show that the relation between utilization and risk premia remains negative, economically large, and statistically significant after controlling for book-to-market ratios, investment rates, and a host of other production-based characteristics. These regressions are implemented as follows. In each year t we run a cross-sectional regression in which the dependent variable is a firm's annual excess return from July in year t to 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. The cross-sectional regression specification is:

$$R_{i,t\to t+1} = \beta_{0,t} + \boldsymbol{\beta}_t' \boldsymbol{X}_{i,t} + \varepsilon_{i,t\to t+1} \tag{3}$$

[13] We note that even though the low and high utilization portfolios differ in their investment rates, there is no statistically significant difference in their total asset growth.

The characteristics we consider are capacity utilization, TFP, hiring, investment over physical capital, capacity overhang, organization capital, the natural logarithms of size and book-to-market, and the lagged annual return. A utilization rate is assigned to each firm following the procedure described in Section OA.3.3 of the Online Appendix, and each control variable is divided by its unconditional standard deviation to aid comparisons between regressions. After running these cross-sectional regressions we compute the time-series average of each estimated slope coefficient to assess the relation between utilization and future stock returns, while holding all other characteristics constant.[14] The results are reported in Table 5 and show that utilization predicts negative and significantly (at the 5% level or better) in all specifications. This indicates that utilization's ability to predict risk premia is incremental to a host of other production-related margins, such as investment-related characteristics. We provide a more detailed discussion of the full implications of Table 5 in Section OA.3.4 of the Online Appendix.

1.7 Empirical robustness

1.7.1 The Utilization Premium under alternative portfolio formations

Section OA.3.1 of the Online Appendix shows that the utilization premium is robust to numerous variations to the portfolio formation procedure. For example, Table OA.3.1. shows that assigning industries to quintile portfolios, or three portfolios based on the $30^{th}/70^{th}$ percentiles of cross-sectional distribution of utilization rates, results in spreads that are close to 5% per annum. Moreover, Table OA.3.2. shows that the magnitude of the utilization premium rises to over 9% per annum in the most recent half of the sample period. We also re-examine the risk exposures to aggregate productivity using quintile portfolios in Table OA.3.8. These exposures to aggregate productivity generally fall with the utilization rate using both linear and non-linear models.

1.7.2 Distinction of the Utilization Premium

Section OA.3.4 of the Online Appendix complements the Fama-Macbeth evidence in four ways. First, Table OA.3.19 in the Online Appendix confirms that the negative relation between utilization and future excess returns is not driven by small-cap firms. We re-estimate projection (3) *after* removing all firms with a market capitalization below the cross-sectional median. The slope coefficient on the utilization rate remains negative and significant at better than the 1% level. Second, Table OA.3.18 shows the results of estimating projection (3) when both returns and characteristics are

[14] We estimate regression (3) at the annual frequency since, unlike the utilization rate, many characteristics of interest (e.g., firm-level productivity and hiring rates) are only available at the annual frequency. For robustness, Table OA.3.18 in the Online Appendix shows the results based on monthly characteristics and returns. While utilization varies every month, characteristics that are only available annually are held constant throughout the year.

Table 5: **Fama-MacBeth regressions**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CU	-1.66	-1.51	-1.64	-1.53	-1.93	-1.41	-1.51	-1.42	-1.64	-1.33
	(-2.31)	(-2.13)	(-2.27)	(-2.11)	(-2.44)	(-2.06)	(-2.12)	(-2.01)	(-2.07)	(-2.01)
TFP		-1.91					-1.66	-1.29	-1.25	1.33
		(-2.23)					(-1.94)	(-1.46)	(-1.59)	(2.81)
HIRE			-4.46				-3.84	-2.99	-7.22	-5.54
			(-3.91)				(-3.45)	(-2.71)	(-4.24)	(-3.64)
I/K				-3.85				-2.12	-2.66	-2.32
				(-3.78)				(-1.96)	(-2.17)	(-2.47)
OVER					-3.18				-3.52	-2.75
					(-3.14)				(-3.60)	(-2.81)
ln(B/M)						2.55				3.87
						(2.74)				(4.11)
ln(ME)										-2.00
										(-1.70)
OC / AT										2.15
										(1.79)
RET_{t-1}										-1.17
										(-0.79)
R^2	0.006	0.014	0.011	0.017	0.015	0.022	0.019	0.027	0.037	0.089

The table reports the results of Fama-Macbeth regressions in which future annual excess returns are regressed on current characteristics. In each year t from 1967 to 2014 (when the TFP data becomes unavailable) we run the cross-sectional regression denoted by equation (3) and report the time-series average of the resulting slope coefficients alongside the associated Newey and West (1987) t-statistics (in parentheses). 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 ratio of organization capital to assets (OC / AT), 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}). Finally, each characteristic is standardized by dividing it by its unconditional standard deviation.

aggregated to the industry-level. We perform this industry-level projection both at an annual and a frequency. In both cases, we obtain the same conclusion as our benchmark. Third, we conduct portfolio double sorts that confirm the distinction between the utilization premium, and the value, investment, organizational capital, and overhang premia. Finally, we simply project the utilization premium on the returns of one of a host of other production-related spreads (e.g., the value and the investment premia) and show that none of these other spreads subsume the utilization premium.

In Section OA.3.6. we decompose firm-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 1.3, and a spread based on time-varying technology (and markups). Overall, the utilization premium is distinct from other known production-based spreads.

1.7.3 Utilization premium: Within-sector and firm-level evidence

Section 1.3 shows the existence of a capacity utilization premium based on cross-sectional data from the FRB. While the FRB database is only available at the industry level, the Online Appendix employs several robustness checks to alleviate potential concerns related to this aggregation level.

First, Section OA.3.7 shows that the utilization spread not only exists *across* sectors but also exists *within* sectors. The spread remains large and significant (i) within the subsample of durable manufacturers (about 5.8% per annum) and when excluding mining industries and utilities; and (ii) when we sort industries into portfolios based on the *growth* rate of utilization, which removes differences in the average level of utilization across industries. The results are materially unchanged.

Second, Section OA.3.8 constructs a firm-level proxy for utilization using Compustat data for robustness. Constructing the proxy entails two steps. First, for each industry j, we project its (demeaned) utilization rate onto its industry-level characteristics. Second, for firm i in industry j, we combine the slope coefficients of industry j with the firm-level characteristics of firm i to construct a proxy for firm i's utilization rate. The resulting proxy varies across firms within each industry. Sorting firms into portfolios based on this proxy yields a negative relation between utilization and expected return, and a firm-level utilization premium of about 5% per annum.

2 The model

We construct a quantitative production-based asset-pricing model with *two* goals: (1) explaining Facts I and II from Section 1, and (2) highlighting the merit of variable utilization rate for fitting the joint distribution of risk premia and investment rates.

Our model deviates from other single-shock production-based models (e.g., Berk, Green, and Naik (1999) and Zhang (2005)) in two important ways: First, we introduce the choice of flexible capacity utilization rates to the model. To the best of our knowledge, while flexible utilization is widespread in the macroeconomic literature, this feature is largely overlooked by quantitative production-based models in cross-sectional asset-pricing.[15] Second, our model also departs from other setups by including a fixed cost of disinvestment that makes selling machines a real option.

2.1 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 aggregate ($X_t \equiv \exp(x_t)$) and

[15] A notable exception is Garlappi and Song (2017) who examine a different set of questions related to IST exposure.

idiosyncratic ($Z_{i,t} \equiv \exp(z_{i,t})$) productivity shocks. The production function for firm i at time t is

$$Y_{i,t} = \exp(x_t + z_{i,t})\big(u_{i,t}K_{i,t}\big)^{\theta\alpha_K}\left(L_{i,t}\right)^{\theta\alpha_L}, \tag{4}$$

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 and controls the intensity with which the firm utilizes its capital. The presence of $u_{i,t}$ in equation (4) provides firms with the flexibility to scale production in response to productivity shocks while keeping the capital stock fixed.[16]

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

$$K_{i,t+1} = \big(1 - \delta(u_{i,t})\big)K_{i,t} + I_{i,t}. \tag{5}$$

$I_{i,t}$ represents gross investment and $\delta(u_{i,t})$ is the depreciation rate of the firm's capital stock that depends on the degree to which capital is utilized. We assume that $\delta'(u_{i,t}) > 0$ to reflect the intuition that capital that is utilized more intensively depreciates at a faster rate.

Productivity. Aggregate productivity (x_t) follows as a stationary AR(1) process:

$$x_{t+1} = \rho_x x_t + \varepsilon^x_{t+1}, \tag{6}$$

where $\varepsilon^x_{t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_x^2\right)$. 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:

$$z_{i,t+1} = \bar{z}\left(1 - \rho_z\right) + \rho_z z_{i,t} + \varepsilon^z_{i,t+1}, \tag{7}$$

where $\varepsilon^z_{i,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_z^2\right)$. 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 (2017) and specify the depreciation function as

$$\delta(u_{i,t}) = \delta_k + \delta_u\left[\frac{u_{i,t}^{1+\lambda} - 1}{1+\lambda}\right]. \tag{8}$$

Here, δ_k represents the depreciation rate when $u_{i,t} = 1$ and δ_u measures the extra 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.

[16] This type of production function featuring utilization is similar to those in Basu, Fernald, and Kimball (2006), Jaimovich and Rebelo (2009), and Garlappi and Song (2017). The fact that utilization scales capital is consistent with the FRB's definition of capacity that mainly reflects changes in capital rather than labor (e.g., Morin and Stevens (2005)). Note that while utilization in the production function is *explicitly* related to capital, the equilibrium choice of labor will *implicitly* (and endogenously) depend on utilization (see equation (14)).

Capital adjustment costs are given by the following function:

$$G_{i,t} \equiv G\left(I_{,t}, K_{i,t}, u_{i,t}\right) = \frac{\phi}{2}\left(\frac{I_{i,t}}{K_{i,t}} - \delta\left(u_{i,t}\right)\right)^2 K_{i,t} + f\mathbf{1}_{\left\{\left(\frac{I_{i,t}}{K_{i,t}} - \delta(u_{i,t})\right)<0\right\}} K_{i,t}, \tag{9}$$

where $\phi > 0$, $f > 0$, and $\mathbf{1}_{\{.\}}$ is an indicator function equal to one when a firm reduces capacity. The adjustment cost function features two components: the standard neoclassical convex cost governed by ϕ and a fixed cost of disinvestment governed by f. This fixed cost reflects frictions in the secondary market for capital, such as the cost of matching with a counterparty (buyer). Structural estimations of adjustment cost functions highlight the existence of non-convex adjustment costs (e.g., Cooper and Haltiwanger (2006)). In our setup, the fixed cost makes disinvestment a real option. This decouples a firm's utilization policy from its investment policy in equilibrium: lower utilization leads lower investment in downturns, as observed empirically. Consequently, this cost also disentangles the utilization premium from the value premium. We discuss the role of f at length in Section OA.4.3 of the Online Appendix. Note that our adjustment cost specification in equation (9) is parsimonious compared to other specifications in the literature (e.g., Cooper and Haltiwanger (2006), 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, W_t, is increasing in the level of aggregate productivity:

$$W_t = \exp(\omega x_t), \tag{10}$$

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 is:

$$\ln\left(M_{t+1}\right) = \ln\left(\beta\right) - \gamma_t \varepsilon^x_{t+1} - \frac{1}{2}\gamma_t^2 \sigma_x^2, \text{ where } \ln\left(\gamma_t\right) = \gamma_0 + \gamma_1 x_t. \tag{11}$$

Here, $0 < \beta < 1$, $\gamma_0 > 0$, and $\gamma_1 < 0$. This SDF is consistent with Jones and Tuzel (2013) and Zhang (2005). The volatility of the SDF is time-varying and driven by γ_t. It increases during economic contractions, and results in a countercyclical price of risk. The term $-\frac{1}{2}\gamma_t^2\sigma_x^2$ in the SDF implies that the risk-free rate is constant. 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 of firm i in period t is given by $D_{i,t} = Y_{i,t} - I_{i,t} - G_{i,t} - L_{i,t}W_t$. In each period, each firm chooses $\{I_{i,t}, L_{i,t}, u_{i,t}\}$ to maximize firm value:

$$V_{i,t} = \max_{\{I_{i,t}, L_{i,t}, u_{i,t}\}} D_{i,t} + \mathrm{E}_t\left[\sum_{j=1}^{\infty} M_{t,t+j} D_{i,t+j}\right], \tag{12}$$

subject to equations (4) – (11). Here, $M_{t,t+j}$ represents the SDF between times t and $t + j$, and

$V_{i,t}$ is the cum-divided value of firm i at time t. Finally, the gross stock return of firm i

$$R^S_{i,t+1} = \frac{V_{i,t+1}}{V_{i,t} - D_{i,t}} \tag{13}$$

Equilibrium. Firms' (i) investment, labor, and utilization policies maximize equation (12) given the SDF, and (ii) valuations satisfy equation (12) given their optimal policies.

2.2 Optimality conditions

Whenever $f > 0$ in equation (9), disinvestment is a real option, $G(\cdot)$ is not differentiable, and the equilibrium conditions are not admissible in closed form. To develop intuition, we analyze the tractable case in which $f = 0$. We then explain how the optimality conditions change for $f > 0$.

No fixed disinvestment cost ($f = 0$). Labor, $L_{i,t}$, is set such that the marginal product of labor ($MPL_{i,t}$) equals the wage rate: $MPL_{i,t} = W_t$.[17] Together with equation (10) this suggests

$$L_{i,t} = \left[X_t^{1-\omega} Z_{i,t} \left(u_{i,t}K_{i,t}\right)^{\theta\alpha_k}\right]^{(1-\theta(1-\alpha_k))^{-1}}. \tag{14}$$

The investment choice, $I_{i,t}$, is determined using the Euler equation

$$1 = \mathrm{E}_t\Big[M_{t,t+1}R^I_{i,t+1}\Big], \tag{15}$$

where $R^I_{i,t+1}$ denotes the returns to investment that can be expressed as

$$R^I_{i,t+1} = \frac{MPK_{i,t+1} + \big(1 - \delta(u_{i,t+1})\big)q_{i,t+1} + \frac{\phi}{2}\left[\left(\frac{I_{i,t+1}}{K_{i,t+1}}\right)^2 - \Big(\delta(u_{i,t+1})\Big)^2\right]}{q_{i,t+1}} \tag{16}$$

Here, $MPK_{i,t+1}$ is the marginal product of capital at time $t+1$, and Tobin's marginal q is

$$q_{i,t} = 1 + \phi\left(\frac{I_{i,t}}{K_{i,t}} - \delta(u_{i,t})\right). \tag{17}$$

Since $q_{i,t}$ measures the present value of an extra unit of installed capital, equation (15) shows the trade-off between the marginal cost and discounted marginal benefit of buying capital. Using equation (17), the first-order condition for the optimal choice of utilization, $u_{i,t}$, is

$$MPU_{i,t} = \delta'(u_{i,t})K_{i,t}. \tag{18}$$

The left hand side of equation (18) represents the benefit of utilization, as captured by its marginal product, or $MPU_{i,t}$.[18] The right hand side represents the cost of raising utilization. A higher utilization rate increases the capital depreciation rate by $\delta'(u_{i,t})$, and results in an extra $\delta_u u^\lambda_{i,t}K_{i,t}$ units of capital depreciating. Thus, higher utilization implies more output today, but less capital in the future. Combined, equations (14) and (18) yield a closed-form optimal utilization rate

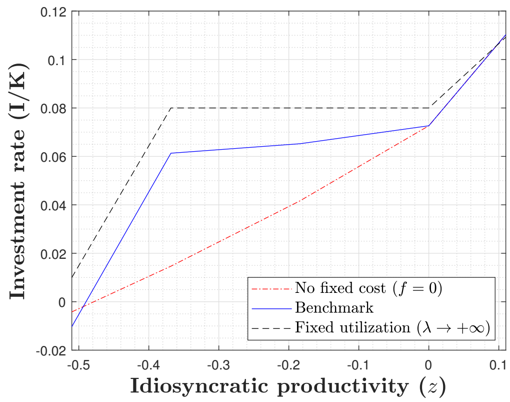
$$u_{i,t} = \left[\delta_u^{-1}\theta\alpha_k X_t^{A_x} Z_{i,t}^{A_z} K_{i,t}^{A_k}\right]^{\left(\frac{1}{\lambda - A_k}\right)}, \tag{19}$$

where $A_x = 1 + (1-\omega)\frac{\theta(1-\alpha_k)}{1-\theta(1-\alpha_k)} > 0$, $A_z = 1 + \frac{\theta(1-\alpha_k)}{1-\theta(1-\alpha_k)} > 0$, and $A_k = \frac{\theta - 1}{1-\theta(1-\alpha_k)} < 0$.

[17] In our setup $MPL_{i,t} \equiv \frac{\partial Y_{i,t}}{\partial L_{i,t}} = \theta\alpha_L X_t Z_{i,t}\big(u_{i,t}K_{i,t}\big)^{\theta\alpha_K}\left(L_{i,t}\right)^{\theta\alpha_L - 1}$.

[18] This marginal product is represented by $\frac{\partial Y_{i,t}}{\partial u_{i,t}} = \theta\alpha_K K_{i,t} X_t Z_{i,t}\left(u_{i,t}K_{i,t}\right)^{\theta\alpha_K - 1}\left(L_{i,t}\right)^{\theta\alpha_L} > 0$.

Figure 1: **Model-implied investment policy**



The figure shows the optimal investment rate policy (I/K) as a function of idiosyncratic productivity (z). Capital and aggregate productivity are set at their stochastic steady-state values, and we focus on z around the region in which the fixed cost of disinvestment apples. We consider the I/K policy under (1) the benchmark model (solid blue line), (2) the model without fixed costs (i.e., $f = 0$) (the dashed red line), and (3) fixed utilization (i.e., $\lambda \to +\infty$).

Given $\lambda - A_k > 0$ and equation (19), we obtain that $\partial u_{i,t}/\partial Z_{i,t} > 0$, $\partial u_{i,t}/\partial X_t > 0$, and $\partial u_{i,t}/\partial K_{i,t} < 0$. When productivity drops, firms seek to drop utilization because the cost of raising utilization (increased depreciation) outweighs the benefit of utilization (increased output). Utilization and capital are negatively related due to decreasing returns to scale. Thus, equation (19) implies that low utilization firms are firms with low idiosyncratic productivity and high capital.

Consequently, firm-level investment and utilization comove positively (i.e., $\rho(\frac{I_{i,t}}{K_{i,t}}, u_{i,t}) > 0$). In particular, when firm-level productivity ($z_{i,t}$) drops, firms want to reduce investment as $MPK_{i,t+1}$ is also smaller in expectation. Simultaneously, a drop in $z_{i,t}$ lowers utilization as explained above.

Fixed disinvestment cost ($f > 0$). When disinvestment is a real option and a firm's productivity drops, there are two opposite forces on optimal investment. On the one hand, the firm wishes to reduce its capital stock as $MPK_{i,t+1}$ is lower in expectation. On the other hand, the firm would like to "wait and see" if productivity improves before making a decision to sell its machines. By waiting, the firm does not incur the fixed cost of disinvestment, $fK_{i,t}$. The balance between these two forces leads to an investment policy whereby firms disinvest if and only if the drop in productivity is sufficiently large. That is, $\exists Z^*(K_{i,t}, X_t)$ such that if $Z_{i,t} < Z^*$, then $K_{i,t+1} < K_{i,t}$, and $K_{i,t+1} = K_{i,t}$ otherwise.[19] In the latter case of keeping the capital stock unaltered, firms set their investment rates to their current depreciation rates, or $I_{i,t}/K_{i,t} = \delta(u_{i,t})$.

To illustrate this trade-off, Figure 1 plots the firm's investment policy under our benchmark calibration (to be described in Section 2.3) when both capital and aggregate productivity are at their

[19]Note that this investment policy breaks the nearly perfect correlation between investment and utilization in the case of $f = 0$. Utilization and investment become substitutes for waiting firms. Thus, $f > 0$ is important for a realistic correlation between $u_{i,t}$ and $I_{i,t}/K_{i,t}$.

Table 6: **Model calibration**

Symbol	Description	Value	Symbol	Description	Value
Stochastic processes			Technology		
ρ_x	Persistence of agg. productivity	0.922	α_k	Capital share	0.333
σ_x	Volatility of agg. productivity	0.014	α_l	Labor share	0.667
$\bar{z}$	Average of idio. productivity	-0.163	θ	Returns to scale	0.950
ρ_z	Persistence of idio. productivity	0.600	δ_k	Fixed dep. rate	0.080
σ_z	Volatility of idio. productivity	0.300	δ_u	Variable dep. rate	0.092
β	Time discount factor	0.988	λ	Elasticity of dep.	3.000
γ_0	Constant price of risk	3.375	ω	Wage sensitivity	0.200
γ_1	Time-varying price of risk	-8.800	ϕ	Adjustment cost	1.500
			f	Fixed cost parameter	0.028

The table reports the parameter values of the production-based asset pricing model in Section 2.1.

steady-state values.[20] The figure shows that with fixed utilization (the case of $\lambda \to \infty$, represented by the dashed black line), all firms that "wait and see" set their investment rates to the common and *constant* rate of δ_k. However, with flexible utilization (solid blue line), the investment rates of waiting firms fluctuate with productivity and over time because $\delta(u_{i,t})$ depends on the stochastic utilization rate. Thus, a flexible utilization rate eliminates periods of investment "inaction."

We also note that utilization *substitutes* selling capital for the purpose of dividend smoothing when $f > 0$. By decreasing utilization in low productivity states, a waiting firm's optimal investment rate of $\delta(u_{i,t})$ drops while maintaining capital, $K_{i,t+1} = K_{i,t}$. This increases dividends in bad states. These model features yield important implications for both risk premia (as we explain in Section 3.1), and real quantities such as investment's dispersion (as we explain in Section 4).

2.3 Calibration and solution method

The model is calibrated at the annual frequency and specified at the firm level. Consistent with this aggregation level, none of the calibration parameters target industry-level quantities. All targeted moments are either firm- or aggregate-level (i.e., an aggregation across all industries) quantities. Importantly, the model does not target the utilization premium. We solve the model numerically using value function iteration as described in Section OA.4.6 of the Online Appendix.

Table 6 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.

[20]We provide the same plot for a wider range of idiosyncratic productivity in Figure OA.8.1 of the Online Appendix. As expected, the figure in the Online Appendix shows that the fixed cost affects the investment policy only when idiosyncratic productivity is negative.

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. This produces a volatility (autocorrelation) of aggregate sales growth rate of 7.5% per annum (0.40) in the model, closely matching the empirical counterpart of 6.6% per annum (0.46). 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).[21] 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, resulting in a value-weighted equity premium of 5.39% per annum and a market return volatility of 20.88% per annum.

Technology. We fix α_K and α_L at 1/3 and 2/3, 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 volatility of the *aggregate* utilization rate. Setting λ to three produces an average annual volatility of aggregate utilization of 4.13% (4.09%) per annum in the model (data). We set ω, the wage sensitivity to aggregate productivity, to 0.20. This value is comparable to both Jones and Tuzel (2013) and Imrohoroglu and Tuzel (2014), 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 to the data. Setting this parameter to 1.5 results in a model-implied annual volatility of investment of 0.14. Finally, we set f, the parameter governing the fixed cost of disinvestment and its lumpiness, 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).

[21]The productivity parameters of Imrohoroglu and Tuzel (2014) are estimated without controlling for time-varying utilization (which is unobserved at the firm-level). Thus, Table OA.4.5 of the Online Appendix shows that our key results are insensitive to perturbing the benchmark values of ρ_x, σ_x, ρ_z, or σ_z.

Furthermore, our choice for σ_x and ρ_x imply that the volatility of (utilization-adjusted) aggregate TFP growth is 1.43% per annum in the model. This is remarkably close to the volatility of utilization-adjusted TFP measure of Fernald (2012), which is 1.48% per annum. In contrast, the volatility of the non-utilization-adjusted TFP measure in the data is quite larger, at 1.79% per annum.

Table 7: **Model-implied moments**

Variable	Data	Model
Panel A: Real quantities		
Volatility of firm-level investment rate (time-series)	0.14	0.14
Volatility of firm-level investment rate (cross-sectional)	0.16	0.11
AC(1) of firm-level investment rate	0.52	0.58
AC(2) of firm-level investment rate	0.26	0.38
Skewness of firm-level investment rate (time-series)	0.67	0.65
Skewness of firm-level investment rate (cross-sectional)	1.89	1.19
Inter-decile range of investment rate	0.32	0.22
Volatility of aggregate capacity utilization level	4.09	4.13
Autocorrelation of aggregate capacity utilization level	0.65	0.92
Volatility of aggregate sales growth	6.58	7.51
Autocorrelation of aggregate sales growth	0.46	0.40
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	20.88
Autocorrelation of excess market return	-0.05	-0.01
Book-to-market spread	3.71	3.77
Investment spread	3.70	4.29

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, aggregate capacity utilization rates, and aggregate sales growth 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.

2.4 Investment and return moments: model versus data

Table 7 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.

Time-series of investment rates. Panel A of Table 7 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. The model also produces realistic estimates for two untargeted moments: the skewness of investment rates (0.65 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 this distribution. Panel A of Table 7 shows 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 (0.32) in the model (data). Our model produces a positively skewed firm-level investment rate of 1.19, consistent with the value of 1.89 in the data.

Aggregate capacity utilization rate. Panel A of Table 7 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 λ to match this volatility. The model also produces a realistic and fairly persistent autocorrelation of utilization (0.7 in the data versus 0.9 in the model).

Aggregate asset-pricing moments. Panel B of Table 7 indicates the risk-free rate and equity premium are 1.2% and 5.4%, respectively. The volatility of excess market returns is 20.9%. The model also produces a slightly negative autocorrelation of excess market returns that is close to the data. Here, model-implied returns are multiplied by 5/3 to account for financial leverage.

Cross-sectional risk premia. Panel B of Table 7 also demonstrates that our model is quantitatively reliable regarding cross-sectional risk premia. The value premium in the model is 3.77% per annum, whereas this spread is 3.71% per annum in the data. The model-implied investment premium of 4.29% per annum is also close to its empirical magnitude of 3.7% per annum without targeting these spreads directly in the calibration.

3 Model implications for the utilization premium

3.1 Model-implied capacity utilization spread

Simulation. We assess the model's ability to produce a monotonically decreasing relation between utilization rates and portfolio returns at both the firm- and industry-level. For the firm-level analysis we simulate a cross-section of four thousand firms. We then sort this cross-section into portfolios on the basis of utilization rates. 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 $t-1$. This procedure is consistent with our empirical portfolio formation procedure described in Section 1.2.

For the model-implied industry-level analysis, we simulate industries by aggregating groups of simulated firms. Although the model does not feature an industry-specific productivity shock, we capture the fact that firms in a given industry share a common productivity component by correlating the firm-level productivities of all firms within an industry. Specifically, when simulating a group of M individual firms that comprise an industry, we set the correlation between each pair of firms' $z_{i,t}$ shocks to 0.50. This choice of correlation coefficient is consistent with the fact that the average time-series correlation between the annual returns of a given firm and the return of its industry is 0.46 in the data.[22] To mimic the number of industries and firms per industry in the

[22]The industry-level results are materially unchanged when we perturb the correlation between the $z_{i,t}$ shocks of

data (see Panel A of Table 3), each model-implied industry represents a value-weighted aggregate of 100 firms, and the economy is comprised of 50 simulated industries.

For each industry, the utilization rate is computed as the value-weighted average utilization rates across all constituents firms. We then sort the cross-section of simulated industries into portfolios in an identical fashion to the empirical exercise, as previously described. We compute population and finite-sample moments for both the firm- and industry-level analyses. The population moments are based on a simulation of the economy over 40,000 periods. The finite-sample distribution is obtained from 500 simulations over 50 periods. The number of periods in the short-sample analysis corresponds to the length of the empirical sample.

Expected returns. Table 8 shows the average returns associated with the utilization-sorted portfolios, and the respective capacity utilization spreads in our simulated economy. Panel A corresponds to the (value-weighted) baseline industry-level utilization spread reported in Table 1. Consistent with the data, the model-implied relation between utilization and average stock returns is negative and monotonic. The low utilization portfolio earns a larger risk premium. The median industry-level utilization premium across the finite-sample simulations is close to 4% per annum. The empirical industry-level utilization spread of 5.7% per annum falls within the model-implied confidence interval. Panel B of 8 reports the firm-level utilization spread within the model, corresponding to Table OA.3.15 in the Online Appendix. The relation between utilization rates and average returns remains monotonic and negative at the firm-level. On the basis of finite-sample simulations, low (high) utilization firms earn an average return of 9.60% (4.46%) per annum. The model-implied firm-level utilization premium is about 5.1% per annum, almost identical to the empirical counterpart. We obtain almost identical moments via the population sample.

Risk exposures. Table 8 also reports the exposure of each utilization portfolio to the model-implied aggregate excess market return, and the spread in these exposures between the extreme utilization portfolios. Specifically, the exposures reported in Panel A correspond to the empirical values shown in the left-most column of Table 2. In the model, the market return is an observable proxy for aggregate productivity, as the model features only one aggregate shock.

In line with our second empirical fact, the model-implied exposure of each portfolio to the market (aggregate productivity) decreases with the utilization rate. The spread in betas between the low and high utilization portfolios is 0.15 (0.26) in the model (data).[23] Comparing Panels A

firms in the same industry. Untabulated results show that the model-implied industry-level utilization spread falls within the empirical confidence interval when the correlation is halved (increased) to 0.25 (0.75).

[23]Notably, the empirical difference in the unconditional market betas of the extreme utilization portfolios is not large on average. However, our model can replicate this finding. The reason, discussed in Section 3.2, is that the beta of low utilization firms is larger than that of high utilization firms in bad states (the converse is true in good states).

Table 8: **Capacity utilization and stock returns: model**

Portfolio	Population $E\left[R^{CU}\right]$	Population β	Short sample $E\left[R^{CU}\right]$
Panel A: Industry-level analysis			
Low (L)	8.82	1.32	8.89 [4.71, 15.16]
Medium	6.73	1.24	6.77 [3.40, 12.37]
High (H)	4.93	1.17	4.96 [0.51, 11.00]
Spread (L-H)	3.88	0.15	3.93 [0.67, 7.45]
Panel B: Firm-level analysis			
Low (L)	9.54	1.34	9.60 [5.76, 16.35]
Medium	6.82	1.23	6.86 [3.53, 12.59]
High (H)	4.41	1.10	4.46 [1.40, 9.69]
Spread (L-H)	5.13	0.24	5.14 [3.96, 6.63]

The table reports the average model-implied annual value-weighted returns of portfolios sorted on capacity utilization at both the firm-level and the industry-level. The table also shows the exposure of each capacity utilization portfolio to the aggregate market return (β). As in the empirical analysis, a firm or industry is sorted into the high (low) utilization portfolio if its 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. In Panel B, which reports firm-level moments, population moments are obtained from one simulation of 1,000 firms for 40,0000 periods (years). Short-sample moments are obtained by averaging moments across 500 simulations of 4,000 firms for 50 periods (years). In Panel A, industry-level returns are simulated using the procedure described in Section 3.1. Here, population moments are obtained from one simulation of 50 industries for 40,000 periods (years). Similarly, short sample moments are obtained by averaging moments across 500 simulations of 50 industries for 50 periods (years). To compute β in the model, the volatility of market returns in the model is scaled to match the volatility of market returns in the data. Finally, square brackets associated with the short sample simulations report the 90% confidence interval related to each moment across the 500 Monte Carlo simulations of the economy.

and B of Table 8 shows that aggregation hardly alters the magnitudes of the risk exposures.

Lastly, the empirical evidence in Section 1.4 shows that the unconditional CAPM alpha of the utilization premium is 4.3% p.a., but statistically insignificant at the 5% level. In Table OA.8.1 of the Online Appendix we show that the model can replicate a similar result, with a somewhat lower alpha. We mimic the empirical exercise by considering short-sample simulations of our economy, and use industry-level returns to construct the utilization spread's alpha. In finite samples, the CAPM can explain the model-implied utilization spread. The mean model-implied CAPM alpha of the spread is 3.45% p.a., but statistically insignificant at the 5% level, consistent with the data.

3.2 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.

Moreover, the equity premium is time-varying, in both the model and the data, and comoves with the beta spread.

However, flexible utilization, $u_{i,t}$, provides firms with a mechanism to reduce these capital frictions. Utilization also directly impacts the cyclicality of firms' output. Consequently, the utilization rate is inherently tied a firm's risk. We illustrate this logic by shutting down ingredients (2) and (3) in our economy, as well as utilization, and explaining the role of each ingredient in turn.

Quadratic adjustment costs only. Assume the only frictions present are quadratic capital adjustment cost, and utilization is fixed ($\lambda \to \infty$). In our model, $\theta = 0.95$, which is approximately constant returns to scale. Therefore, a sufficient statistic for the ex-dividend firm value is Tobin's q (as per equation (17) with $\delta(u) = \delta_k$ since $\lambda \to \infty$). The risk of each firm is determined by the interaction between firm-level investment and capital adjustment costs, as implied by Tobin's q. The ex-dividend firm's productivity beta, $\beta_{i,t}$, can be written as $\beta_{i,t} = \frac{\partial V_{i,t}}{\partial \varepsilon_{x,t}} \approx \frac{\partial}{\partial \varepsilon_{x,t}} q_{i,t}(i_{i,t})\dot{K}_i$. As $q'(i_{i,t}) > 0$, the valuation of investing (disinvesting) firms rises (drops), and thereby covaries more with aggregate productivity in good (bad) states in which x_t is (low) high. Firms that make large (dis)investments are required to pay large capital adjustments costs that restrict the ability of shocks to be absorbed in investment. As shocks are not fully absorbed in quantities, they are absorbed in installed capital's price. Thus, both high and low investment rate firms are risky depending on the phase of the business cycle (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, the interaction between utilization and investment can mitigate (dis)investment adjustment costs and reduce the risk associated with altering capital. Consider a firm facing lower productivity. As discussed in Section 2.2, while the firm still has the incentive to reduce its capital stock, thereby exposing itself to potentially large quadratic capital adjustment costs, the firm also has two incentives to lower its utilization. First, equation (18) suggests that by lowering its utilization rate, the firm benefits from a reduction in its 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, because lower utilization implies lower depreciation (i.e., a lower natural rate of investment), the firm can pay a lower quadratic adjustment cost to disinvest. To see this, consider equation (9). If $\delta(u_{i,t})$ drops whenever $I_{i,t}/K_{i,t}$ drops, then the gap between the two rates shrinks, reduing the quadratic cost. Equation (17) implies that this creates a partial hedge for (dis)investment risk in (bad) good times by attenuating the fluctuations in q.

The incentives above create positive comovement and complementarity between a firm's need to disinvest and low utilization. Therefore, low utilization firms (that have low idiosyncratic productivity and high capital according to equation (19)), are riskier during aggregate economic downturns because they face large capital downscaling costs that they partially hedge through lower utilization (i.e., $\beta_{i,t} \uparrow$ if $u_{i,t} \downarrow$ and $x_t \downarrow$ since $\rho(u_{i,t}, i_{i,t}) > 0$). The converse holds 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 aggregate productivity. We break the symmetry in risk exposures between high and low utilization firms by introducing ingredients (2) and (3).

The role of the fixed disinvestment cost for risk. When we enrich the model with ingredient (2), the fixed cost of disinvestment ($f > 0$), we introduce a higher friction to disinvest. Reducing capital becomes a costly real option. As discussed in Section 2.2, firms facing a moderate drop in productivity do not disinvest immediately. They "wait and see" if productivity improves before exercising the costly disinvestment options. While not exercising the real option, the risk of the firms that "wait" rises (their capital is further from optimum with this friction). Simultaneously, these waiting firm substitute exercising the option to sell machines by temporarily lowering utilization. This helps to partially hedge capital risk. To see this, note that all waiting firms set their investment rate to the depreciation rate to maintain the capital stock. By lowering utilization, firms reduce $\delta(u_{i,t})$ and lowers the required investment rate needed to maintain capital. Lowered investment in bad states *raises* the current dividend. This creates a partial hedge compared to the case in which utilization is fixed.[24] Since the underlying frictions in the market for selling capital are even greater for low utilization firms when $f > 0$, their betas in bad states of the world exceed 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 role of the countercylical price of risk. 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 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. Combined, ingredients (2) and (3) yield a monotonic relation between utilization and risk premia.

Sensitivity analyses and discussions. Section OA.4 in the Online Appendix reports a host of additional model-related results. Section OA.4.1 discusses the model's assumptions in detail. Section OA.4.2 then shows that the utilization premium remains positive in the model after

[24] In other words, when utilization is fixed, firms that wait to sell capital set their investment rate to δ_k. When firms contemporaneously lower utilization, they set their investment rate to $\delta(u_{i,t}) < \delta_k$. As the current dividend and investment are negatively related, the firm's payout rises, all else equal.

conditioning on book-to-market (in line with the data). Section OA.4.4 illustrates the model's intuition for the spread numerically by perturbing key parameters and showing the spread falls when ϕ, ρ_x, σ_x, ρ_z, or σ_z drops, and with a constant market price of risk ($\gamma_1 = 0$). Finally, Section OA.4.5 designs a simulation-based experiment to put an upper bound on the degree to which ex-ante sectoral-level heterogeneity affects the utilization spread. The experiment suggest that heterogeneity in δ_k, λ, ϕ, or f induces only a marginal effect on the utilization premium.

3.3 Model extensions

Section OA.5 considers two model extensions. The first introduces a priced investment efficiency shock to the model, as in Papanikolaou (2011) and Justiniano, Primiceri, and Tambalotti (2011). Specifically, we change the quadratic adjustment cost function to

$$\frac{\phi - s_t}{2}\left(\frac{I_{i,t}}{K_{i,t}} - \delta\left(u_{i,t}\right)\right)^2 K_{i,t}, \tag{20}$$

where s_t represents the degree of aggregate investment efficiency, which evolves as an independent AR(1) process (for which $\phi >> \sigma_s$). Increases in s_t raise the marginal efficiency of investment, suggesting that investment can be transformed into capital in a more efficient manner.

In equilibrium, investment efficiency shocks affect firms' investment policies but do not affect their choices of utilization rates. Because $u_{i,t}$ does not depend on the shock to s_t, the expected returns of the utilization-sorted portfolios remain tightly linked to their exposures to aggregate productivity. Contrarily, because $I_{i,t}/K_{i,t}$ depends on the shock to s_t, the expected returns of spreads related to firms' investment policies (e.g., the value premium) will depend more heavily on their exposures to the s_t shocks. This creates a stark separation between the two risk premia.

The calibrate model yields an unconditional utilization (value) premium of around 6% (4%) per annum. We then confirm that the two risk premia are materially distinct. Notably, projecting the utilization premium on the value premium results in a model-implied alpha of around 5% per annum, which is consistent with the data.

The second model extension introduces (reduced-form) firm-specific and time-varying markups to the model. Notably, we replace equation (4) with

$$Y_{i,t} = \exp(x_t + z_{i,t})\left[u_{i,t}^{(\theta\alpha_k\cdot\xi)} + \mu_{i,t}^{\xi}\right]^{\frac{1}{\xi}} K_{i,t}^{\theta\alpha_K} L_{i,t}^{\theta\alpha_L}, \tag{21}$$

where $\mu_{i,t}$ denotes the time-varying markups of firm i and ξ represents the elasticity of substitution between utilization and markups. We estimate the idiosyncratic productivity of each firm in the model through Solow residuals. Sorting firms into portfolios on the basis of the TFP we elicit results in a productivity premium of about 5.7% per annum. This model-implied productivity premium is largely separated from the model-implied utilization premium, as a projection of the utilization spread on the TFP spread results in alpha of about 6% (in line with the data).

4 Model implications for macro-finance modeling

4.1 Implications for investment moments

Flexible utilization plays a pivotal role for simultaneously matching investment and asset-pricing moments in the presence of real investment options. Panel A of Table 9 shows model-implied moments in an economy both without flexible utilization (i.e., with $\lambda \to \infty$) and with flexible utilization (i.e., with finite λ). For brevity, much of the discussion is relegated to Section OA.6 of the Online Appendix,. However, the key points are provided below.

Whenever utilization in the model is fixed (Row 1), the cross-sectional dispersion and skewness of investment are less than half of their empirical magnitudes. The time-series skewness of firm-level investment is negative, whereas it is positive in the data. This happens because disinvestment is a costly real option. During moderate economic slowdowns, firms "wait and see" if productivity improves before opting to sell capital. Under fixed utilization, these firms do not alter their capital and set their investment rates equal to the (constant) depreciation rate instead. Because a mass of waiting firms are then lumped around the center of investment's distribution, the cross-section of investment rates is compressed, and features low dispersion. If productivity is persistently negative, these waiting firms pass a tipping point in which they are overly burdened with unproductive capital and disinvest sharply. These disinvestment jumps create the counterfactual negative sign for the time-series skewness of investment. Moreover, as the distribution of investment rates is too compressed, firms' risk exposures to aggregate productivity do not feature enough heterogeneity. This shrinks investment-related spreads, such as the value premium.

Introducing flexible utilization to the model addresses the former misses (Row 2). When utilization is flexible, firms can 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 increase the cross-sectional dispersion of investment. Thus, 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. 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. Lastly, the greater dispersion in

Table 9: **Model-implied moments across alternative calibrations of the model**

Row	Model	Time-series $\sigma_{TS}(ik)$	Time-series $S_{TS}(ik)$	Time-series $\rho_1(ik)$	Time-series $\sigma(u)$	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	4.09	0.16	1.89	3.71	3.70
	Panel A: Sensitivity to utilization (λ)								
	Baseline without utilization								
(1)	($\lambda \to \infty$)	0.11	-0.27	0.63	-	0.07	0.07	3.00	2.29
	Baseline								
(2)	($\lambda = 3$)	0.14	0.65	0.58	4.13	0.11	1.19	3.77	4.29
	Different λ								
(3)	Very low ($\lambda = 2.00$)	0.16	1.12	0.55	6.21	0.13	1.60	4.13	4.72
(4)	Low ($\lambda = 2.90$)	0.14	0.68	0.58	4.27	0.11	1.22	3.80	4.32
(5)	High ($\lambda = 3.10$)	0.14	0.62	0.58	4.00	0.11	1.16	3.75	4.26
(6)	Very high ($\lambda = 13.00$)	0.12	-0.07	0.62	0.95	0.08	0.37	3.18	3.56
	Panel B: Sensitivity to adjustment costs (ϕ)								
	Different ϕ and fixed utilization								
(7)	Very low ($\phi = 0.75$)	0.18	0.14	0.58	-	0.12	0.03	2.25	2.03
(8)	Low ($\phi = 1.40$)	0.12	-0.23	0.62	-	0.07	0.07	2.91	2.28
(9)	High ($\phi = 1.60$)	0.11	-0.31	0.63	-	0.07	0.09	3.06	2.32
(10)	Very high ($\phi = 3.00$)	0.06	-0.73	0.64	-	0.04	0.14	3.74	2.37
	Different ϕ and flexible utilization								
(11)	Very low ($\phi = 0.75$)	0.21	0.76	0.56	4.77	0.15	0.81	2.76	3.63
(12)	Low ($\phi = 1.40$)	0.15	0.64	0.58	4.20	0.11	1.14	3.67	4.21
(13)	High ($\phi = 1.60$)	0.14	0.66	0.58	4.06	0.10	1.23	3.87	4.36
(14)	Very high ($\phi = 3.00$)	0.10	0.91	0.56	3.37	0.08	1.64	4.77	5.08

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 ($S_{TS}(ik)$), the first-order autocorrelation ($\rho(ik)$) of firm-level investment rates, the time-series volatility of utilization ($\sigma(u)$), 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 elasticity of marginal depreciation (λ) or the quadratic capital adjustment cost (ϕ). 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.

investment suggests a larger dispersion in firms' risk exposures, boosting cross-sectional spreads.

More generally, and as we show in Rows (3) to (6), the value of λ has a substantial quantitative impact on matching the data. As utilization becomes more flexible (i.e., as λ decreases), the time-series/cross-sectional skewness and volatility of investment rise, and risk premia increase as well. This generally moves each moment towards its empirical counterpart when compared to Row (1). However, utilization cannot be overly flexible. When λ is very low, as in row (3), the volatility of utilization exceeds the 95% confidence interval of this quantity in the data.

Panel B of Table 9 shows model-implied moments when we perturb the adjustment costs with the fixed and flexible utilization models. The panel shows that the problem of matching moments under fixed utilization is not alleviated by recalibrating the model. Rows (7) and (8) consider the case of lower frictions compared to the benchmark. Sufficiently lower friction can help turn the time-series skewness of investment to a positive value, but the cross-sectional skewness of investment is still too small. Lower capital adjustment frictions also cause risk premia to fall.

The diminished value premium under fixed utilization can be raised by increasing the capital adjustment costs (Row 10). However, the adjustment costs required to match the value premium are 100% higher than those with flexible utilization. This alternative calibration has counterfactual implications for investment's dispersion. When utilization is flexible, the dispersion of investment rates rises. This increases the dispersion of firms' exposures to aggregate productivity and allows the model to rely on lower capital adjustment frictions while generating cross-sectional risk premia.

More general adjustment costs. We also augment the fixed-utilization model with a more complex adjustment cost function that is inspired by Cooper and Haltiwanger (2006) and requires extra model parameters. Section OA.6.4 in the Online Appendix shows that while this model's fit to the data is improved, the model still fails to fully reconcile the data in spite of the additional exogenous calibration parameters. As such, utilization "saves" on the degree of the exogenous parameters needed to explain the data.

4.2 Implications for depreciation fluctuations

Equation (8) of the model suggests that firms' depreciation rates are positively related to firms' utilization rates. In Section OA.7 of the Online Appendix we present empirical evidence that supports this prediction. In fact, recent studies show that BEA- and Compustat-implied depreciation rates are strikingly different (e.g., Clementi and Palazzo (2019); Bai et al. (2019)). We show in Section OA.7.1 that the correlation between these two measures of depreciation increases when controlling for utilization. We also propose a method to measure the *aggregate* depreciation rate based on utilization data in Section OA.7.2. Unlike the BEA rate, which is only available at the annual frequency, our method provides a high-frequency (i.e., monthly) measure.

Finally, as the correlation between utilization and depreciation rates is not perfect, Section OA.5.3 of the Online Appendix augments the model from Section 2 to include depreciation shocks. This means that a firm's depreciation rate becomes a combination of its utilization rate and an exogenous and systematic capital depreciation shock. While this shock breaks the high model-implied correlation between utilization and depreciation rates, this additional shock has a minimal impact on the magnitude of the model-implied utilization premium.

5 Conclusion

We show that flexible utilization induces sizable implications for cross-sectional risk profiles and investment choices. Empirically, we document two facts: (1) A low capacity utilization portfolio earns a higher expected return of about 5% per annum, resulting in a *Utilization Premium*. Utilization predicts returns beyond production-based characteristics, such as investment and hiring rates; (2) There is a monotonically decreasing relation between utilization and aggregate productivity exposures. The low utilization portfolio is more sensitive to changes in productivity. Theoretically, we construct a production-based model that can quantitatively reconcile the new facts.

In the model, downscaling capital in the presence of capital adjustment costs increases firms' exposures to aggregate risk. Lowering utilization allows firms to partially hedge this risk. First, lower capacity utilization causes the depreciation rate to decrease and conserves more capital for future periods that are more productive. Second, the decrease in the depreciation rate 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 substitute selling capital by lowering utilization. Overall, low utilization firms are risky because a low utilization rate is indicative of a firm that wants to drop investment, faces high frictions in the market for selling capital, and tries to partially alleviate these frictions through utilization.

Flexible utilization also has broader implications for macro-finance models. In a real option setup with fixed utilization, the cross-section of investment rates features too little dispersion and skewness. By inducing a time-varying depreciation rate, flexible utilization increases the dispersion and asymmetry in investment's distribution, and induces the dispersion of firms' risk exposures to rise. Consequently, a flexible utilization model generates large return spreads, while relying on parsimonious adjustment costs. Overall, our results show the importance of time-varying utilization for expected returns and real quantities.

References

Ai, H., Li, J. E., Li, K., Schlag, C., 2019. The collateralizability premium. The Review of Financial Studies, forthcoming .

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

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.

Bai, H., Li, E. X. N., Xue, C., Zhang, L., 2019. Does costly reversibility matter for U.S. public firms? Working Paper .

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., Li, J., Lin, X., Zhao, X., 2017. Labor-force heterogeneity and asset prices: The importance of skilled labor. The Review of Financial Studies 30, 3669–3709.

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.

Belo, F., Lin, X., Yang, F., 2018. External Equity Financing Shocks, Financial Flows, and Asset Prices. The Review of Financial Studies 32, 3500–3543.

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

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

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.

Carlstrom, C. T., Fuerst, T. S., 1997. Agency costs, net worth, and business fluctuations: A computable general equilibrium analysis. The American Economic Review pp. 893–910.

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

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., 2006. On the nature of capital adjustment costs. 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.

Dou, W., Ji, Y., Reibstein, D., Wu, W., 2019. Customer capital, financial constraints, and stock returns. Journal of Finance, forthcoming .

Eisfeldt, A. L., Papanikolaou, D., 2013. Organization capital and the cross-section of expected returns. The Journal of Finance 68, 1365–1406.

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.

Fernald, J., 2012. A quarterly, utilization-adjusted series on total factor productivity. Working paper. Federal Reserve Bank of San Francisco .

Garlappi, L., Song, Z., 2017. 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.

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.

Justiniano, A., Primiceri, G. E., Tambalotti, A., 2011. Investment shocks and the relative price of investment. Review of Economic Dynamics 14, 102–121.

Kilic, M., 2017. Asset pricing implications of hiring demographics. Working paper .

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.

Koenig, E. F., et al., 1996. Capacity utilization as a real-time predictor of manufacturing output. Economic Review-Federal Reserve Bank of Dallas pp. 16–23.

Kogan, L., Li, J., Zhang, H. H., 2019. A production-based economic explanation for the gross profitability premium. Working paper .

Kuehn, L.-A., Schmid, L., 2014. Investment-based corporate bond pricing. The Journal of Finance 69, 2741–2776.

Lewellen, J., Nagel, S., 2006. The conditional capm does not explain asset-pricing anomalies. Journal of financial economics 82, 289–314.

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

Lin, X., 2012. Endogenous technological progress and the cross-section of stock returns. Journal of Financial Economics 103, 411 – 427.

Loualiche, E., et al., 2019. Asset pricing with entry and imperfect competition. Working paper .

Morin, N., Stevens, J. J., 2005. Diverging measures of capacity utilization: An explanation. Business Economics 40, 46–54.

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

Olley, G. S., Pakes, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica 64, 1263–1297.

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.

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.

Tuzel, S., Zhang, M. B., 2017. Local risk, local factors, and asset prices. The Journal of Finance 72, 325–370.

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