

The Utilization Premium^{*}

Fotis Grigoris and Gill Segal

February 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 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-

¹Notably, 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 quantitatively 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

²Firms 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.

³In 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.

⁴In 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 Bazardresch (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:

$$CU_{i,t} = \frac{IP_{i,t}}{Capacity_{i,t}}. \quad (1)$$

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

⁵While 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.

⁶The 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.

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

⁸A 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 10th (90th) 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 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

$$Ret_{i,t}^e = \beta_{0,i} + \beta_{1,i} \text{Agg-Prod}_t + \varepsilon_{i,t}, \quad (2)$$

where $Ret_{i,t}^e$ 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.

⁹Table 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.

¹⁰The 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		Equal-weighted	
	Mean	SD	Mean	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 (L-H)	5.67 (2.31)	17.71	5.44 (2.47)	15.51

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

Motivated by this evidence, along with the discussion below, our model in Section 2 features a single source of risk: aggregate productivity. As our model shows, exposures to aggregate productivity endogenously vary with firms' choices of utilization rates.

Time-varying exposures. 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).

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**

Portfolio	Market returns		Util.-adjusted TFP		Labor productivity	
	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: $Ret_{i,t}^e = \beta_0 + \beta_1 \text{Agg-Proxy}_t + \varepsilon_{i,t}$, where $Ret_{i,t}^e$ 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.

We also establish a negative relation between time-varying exposures and utilization rates.

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 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.

¹²While oil extraction frequently appears in the high utilization portfolio, Section OA.3.7 in the Online Appendix demonstrates that the utilization premium remains positive and significant if we exclude the entire mining sector.

Table 3: Characteristics of capacity utilization sorted portfolios

	Low (L)	Medium	High (H)	Diff(L-H)	<i>t</i> (Diff)
Panel A: Portfolio constituents					
CU (%)	67	79	91		
<i>N</i> (<i>Stocks</i>)	617	4423	348		
<i>N</i> (<i>Ind.</i>)	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.

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, 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

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).

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. 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, \mathbf{X}_t , measured at the end of June in year t . The cross-sectional regression specification is:

$$R_{i,t \rightarrow t+1} = \beta_{0,t} + \boldsymbol{\beta}'_t \mathbf{X}_{i,t} + \varepsilon_{i,t \rightarrow t+1} \quad (3)$$

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

Table 5: Fama-MacBeth regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CU	-1.66 (-2.31)	-1.51 (-2.13)	-1.64 (-2.27)	-1.53 (-2.11)	-1.93 (-2.44)	-1.41 (-2.06)	-1.51 (-2.12)	-1.42 (-2.01)	-1.64 (-2.07)	-1.33 (-2.01)
TFP		-1.91 (-2.23)					-1.66 (-1.94)	-1.29 (-1.46)	-1.25 (-1.59)	1.33 (2.81)
HIRE			-4.46 (-3.91)				-3.84 (-3.45)	-2.99 (-2.71)	-7.22 (-4.24)	-5.54 (-3.64)
I/K				-3.85 (-3.78)				-2.12 (-1.96)	-2.66 (-2.17)	-2.32 (-2.47)
OVER					-3.18 (-3.14)				-3.52 (-3.60)	-2.75 (-2.81)
ln(B/M)						2.55 (2.74)				3.87 (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.

assess the relation between utilization and future stock returns, while holding all other characteristics constant.¹³ 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

¹³We 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.

assigning industries to quintile portfolios, or three portfolios based on the 30th/70th 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 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.¹⁴ 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 idiosyncratic ($Z_{i,t} \equiv \exp(z_{i,t})$) productivity shocks. The production function for firm i at time t is

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

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.¹⁵

¹⁴A notable exception is Garlappi and Song (2017) who examine a different set of questions related to IST exposure.

¹⁵This 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

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

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

$I_{i,t}$ 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:

$$x_{t+1} = \rho_x x_t + \varepsilon_{t+1}^x, \quad (6)$$

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

$$z_{i,t+1} = \bar{z}(1 - \rho_z) + \rho_z z_{i,t} + \varepsilon_{i,t+1}^z, \quad (7)$$

where $\varepsilon_{i,t+1}^z \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_z^2)$. We assume that $\varepsilon_{i,t+1}^z$ and $\varepsilon_{j,t+1}^z$ are uncorrelated for $i \neq j$ and that idiosyncratic shocks are uncorrelated with ε_{t+1}^x . \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

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

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 is for a firm to alter its utilization rate in response to exogenous shocks. Holding all else constant, larger values of λ make increasing the capacity utilization rate more costly and ensures that firms choose a finite level of utilization.

Capital adjustment costs are given by the following function:

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

where $\phi > 0$, $f > 0$, and $\mathbf{1}_{\{\cdot\}}$ 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

(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)).

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:

$$W_t = \exp(\omega x_t), \quad (10)$$

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:

$$\ln(M_{t+1}) = \ln(\beta) - \gamma_t \varepsilon_{t+1}^x - \frac{1}{2} \gamma_t^2 \sigma_x^2, \text{ where } \ln(\gamma_t) = \gamma_0 + \gamma_1 x_t. \quad (11)$$

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:

$$V_{i,t} = \max_{\{I_{i,t}, L_{i,t}, u_{i,t}\}} D_{i,t} + E_t \left[\sum_{j=1}^{\infty} M_{t,t+j} D_{i,t+j} \right], \quad (12)$$

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-dividend value of firm i at time t . Finally, the gross stock return of firm i

$$R_{i,t+1}^S = \frac{V_{i,t+1}}{V_{i,t} - D_{i,t}} \quad (13)$$

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: $MPL_{i,t} = W_t$.¹⁶ Together with equation (10) this suggests

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

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

$$1 = E_t \left[M_{t,t+1} R_{i,t+1}^I \right], \quad (15)$$

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

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

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

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

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

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

The left hand side of equation (18) represents the benefit of utilization, as captured by its marginal product, or $MPU_{i,t}$.¹⁷ 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 $\delta_u u_{i,t}^\lambda K_{i,t}$ 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

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

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 $A_k = \frac{\theta-1}{1-\theta(1-\alpha_k)} < 0$.

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.

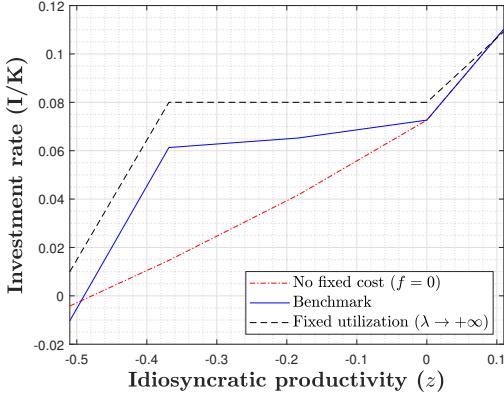
Consequently, firm-level investment and utilization comove positively (i.e., $\rho(\frac{I_{i,t}}{K_{i,t}}, u_{i,t}) > 0$). In particular, when firm-level productivity ($Z_{i,t}$) drops, firms want to reduce investment as $MPK_{i,t+1}$ is also smaller in expectation. Simultaneously, a drop in $Z_{i,t}$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

Fixed disinvestment cost ($f > 0$). When disinvestment is a real option and a firm's

¹⁶In our setup $MPL_{i,t} \equiv \frac{\partial Y_{i,t}}{\partial L_{i,t}} = \theta \alpha_L X_t Z_{i,t} (u_{i,t} K_{i,t})^{\theta \alpha_K} (L_{i,t})^{\theta \alpha_L - 1}$.

¹⁷This marginal product is represented by $\frac{\partial Y_{i,t}}{\partial u_{i,t}} = \theta \alpha_K K_{i,t} X_t Z_{i,t} (u_{i,t} K_{i,t})^{\theta \alpha_K - 1} (L_{i,t})^{\theta \alpha_L} > 0$.

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 applies. 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., $\lambda \rightarrow +\infty$).

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.¹⁸ 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 steady-state values.¹⁹ The figure shows that with fixed utilization (the case of $\lambda \rightarrow \infty$, 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 invest-

¹⁸Note 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 $I_{i,t}/K_{i,t}$.

¹⁹We 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.

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.

ment 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.

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 7.6% per annum (0.41) 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).²⁰ The long-run average level of idiosyncratic productivity (\bar{z}) is a scaling variable set so

²⁰The 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

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 20.89% 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 4.15% (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).

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 (0.66 in the model versus 0.67 in the data) and the second-order autocorrelation of investment (0.38 in the model versus 0.26 in the data).

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	0.66	
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	4.15	
Autocorrelation of aggregate capacity utilization level	0.65	0.92	
Volatility of aggregate sales growth	6.58	7.55	
Autocorrelation of aggregate sales growth	0.46	0.41	
Panel B: Asset prices			
Real risk-free rate	1.19	1.21	
Excess market return	6.28	5.39	
Volatility of excess market return	17.20	20.89	
Autocorrelation of excess market return	-0.05	-0.00	
Book-to-market spread	3.71	3.76	
Investment spread	3.70	4.27	

The table shows model-implied moments, obtained by simulating 1,000 firms for 40,000 periods (years), alongside their empirical counterparts, computed using data from 1967 to 2015. Panel A displays moments associated with firm-level investment rates, 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.

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 quan-

titatively reliable regarding cross-sectional risk premia. The value premium in the model is 3.76% per annum, whereas this spread is 3.71% per annum in the data. The model-implied investment premium of 4.27% per annum is also close to its empirical magnitude of 3.7% per annum 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 10th (90th) 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.²¹ To mimic the number of industries and firms per industry in the 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

²¹The industry-level results are materially unchanged when we perturb the correlation between the $z_{i,t}$ shocks of 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).

Table 8: Capacity utilization and stock returns: model

Portfolio	Population		Short sample $E[R^{CU}]$
	$E[R^{CU}]$	β	
Panel A: Industry-level analysis			
Low (L)	9.05	1.34	9.01 [4.24, 16.87]
Medium	6.93	1.25	6.93 [3.14, 14.07]
High (H)	5.11	1.17	4.97 [-0.10, 12.66]
Spread (L-H)	3.94	0.17	4.04 [0.38, 8.75]
Panel B: Firm-level analysis			
Low (L)	9.53	1.34	9.86 [5.47, 18.35]
Medium	6.83	1.23	7.11 [3.28, 14.56]
High (H)	4.41	1.09	4.64 [1.21, 11.18]
Spread (L-H)	5.12	0.25	5.21 [3.79, 7.09]

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 90th (10th) 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 40,000 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 95% confidence interval related to each moment across the 500 Monte Carlo simulations of the economy.

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 4.04% p.a. The empirical industry-level utilization spread of 5.6% per annum falls within the model-implied confidence interval. Panel B of Table 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 9.86% (4.64%) per annum. The model-implied firm-level utilization premium is about 5.2% 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 0.17 (0.26) in the model (data).²² Comparing Panels A 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 3.2% 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. However, flexible utilization, $u_{i,t}$, provides firms with a mechanism to reduce these capital frictions. Utilization also directly impacts the cyclicity 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 ($\lambda \rightarrow \infty$). 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 $\lambda \rightarrow \infty$). 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 $\beta_{i,t} = \frac{\partial V_{i,t}}{\partial \varepsilon_{x,t}} \approx \frac{\partial}{\partial \varepsilon_{x,t}} q_{i,t}(i_{i,t}) \dot{K}_i$.

²²Notably, 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). Moreover, the equity premium is time-varying, in both the model and the data, and comoves with the beta spread.

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, reducing 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.²³ 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 countercyclical 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 $E[R_{i,t+1}^e] \approx \beta_{i,t}\gamma(x_t) \uparrow$ 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 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 design 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).

²³In 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.

Specifically, we change the quadratic adjustment cost function to

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

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 calibrated 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

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

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 $\lambda \rightarrow \infty$) 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 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

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

Row	Model Data	Time-series				Cross-sectional		Risk premia		
		$\sigma_{TS}(ik)$	$S_{TS}(ik)$	$\rho_1(ik)$	$\sigma(u)$	$\sigma_{CS}(ik)$	$S_{CS}(ik)$	$E[R^{bm}]$	$E[R^{ik}]$	
		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)	($\lambda \rightarrow \infty$)	0.11	-0.09	0.64	-	0.07	0.11	2.93	2.30	
Baseline										
(2)	($\lambda = 3$)	0.14	0.66	0.58	4.15	0.11	1.19	3.76	4.27	
Different λ										
(3)	Very low ($\lambda = 2.00$)	0.16	1.12	0.56	6.24	0.13	1.60	4.11	4.69	
(4)	Low ($\lambda = 2.90$)	0.14	0.69	0.58	4.30	0.11	1.22	3.78	4.30	
(5)	High ($\lambda = 3.10$)	0.14	0.63	0.58	4.02	0.11	1.16	3.74	4.24	
(6)	Very high ($\lambda = 13.00$)	0.12	-0.05	0.62	0.96	0.08	0.39	3.18	3.53	
Panel B: Sensitivity to adjustment costs (ϕ)										
Different ϕ and fixed utilization										
(7)	Very low ($\phi = 0.75$)	0.19	0.50	0.59	-	0.12	0.06	2.04	2.03	
(8)	Low ($\phi = 1.40$)	0.12	-0.04	0.63	-	0.08	0.10	2.84	2.28	
(9)	High ($\phi = 1.60$)	0.11	-0.14	0.64	-	0.07	0.11	3.00	2.34	
(10)	Very high ($\phi = 3.00$)	0.06	-0.67	0.65	-	0.04	0.13	3.72	2.40	
Different ϕ and flexible utilization										
(11)	Very low ($\phi = 0.75$)	0.21	0.73	0.56	4.78	0.15	0.81	2.76	3.62	
(12)	Low ($\phi = 1.40$)	0.15	0.65	0.58	4.23	0.11	1.15	3.66	4.20	
(13)	High ($\phi = 1.60$)	0.14	0.67	0.58	4.09	0.10	1.23	3.85	4.34	
(14)	Very high ($\phi = 3.00$)	0.10	0.92	0.57	3.40	0.08	1.66	4.77	5.06	

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 ($\rho_1(ik)$) 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 ($E[R^{bm}]$) and investment premium ($E[R^{ik}]$) 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 40,000 periods (years). Finally, the top row of the table also reports the empirical counterpart of each moment.

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., Bundick, B., 2017. Uncertainty shocks in a model of effective demand. *Econometrica* 85, 937–958.
- 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.
- Bloom, N., 2009. The impact of uncertainty shocks. *Econometrica* 77, 623–685.
- 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 .
- Gabaix, X., 2011. The granular origins of aggregate fluctuations. *Econometrica* 79, 733–772.
- 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.
- Jermann, U. J., 1998. Asset pricing in production economies. *Journal of monetary Economics* 41, 257–275.
- 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.
- Patton, A. J., Timmermann, A., 2010. Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics* 98, 605–625.
- 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.

A Internet appendix

OA.1 Variable description and construction

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

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

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

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

Debt growth. We measure the growth rate of a firm's debt by calculating the annual percentage change in outstanding total debt, expressed in real terms. We define total real debt as the sum of long-term debt (Compustat Annual item DLTT) and debt in current liabilities (Compustat Annual item DLC), scaled by the value of the consumer price index. When computing this quantity we require firms to have at least \$10m of debt outstanding in year $t - 1$.

Depreciation rate (BEA implied). To compute the BEA-implied depreciation rate we take

the GDP-weighted average of the industry-level depreciation rates associated with equipment and structures.

Depreciation rate (Compustat implied). We compute Compustat-implied industry-level depreciation rates by first constructing firm-level depreciation rates. A firm's depreciation rate is defined as the firm's depreciation expense (Compustat annual item DP) minus the firm's amortization (Compustat annual item AM) scaled by net property, plant, and equipment (Compustat annual item PPENT). We then aggregate these firm-level depreciation rates to the industry-level by computing the value-weighted depreciation rate across all firms assigned to a particular industry.

Equity issuance. Gross equity issuance is defined as the sale of common and preferred stock (Compustat Annual item SSTK) divided by the lagged value of book equity, as per Belo et al. (2018).

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

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

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

Intangible investment rate (R&D / ME). We follow Lin (2012) and define a firm's intangible investment rate as the firm's R&D expense (Compustat Annual item XRD) divided by the firm's market capitalization.

Inventory growth. The inventory growth rate is defined following Belo and Lin (2012). That is, we compute the annual percentage change in each firm's inventory holdings (Compustat Annual item INVT) after converting the value of inventories to real terms.

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

Leverage. We define a firm's leverage ratio as long-term debt (Compustat Annual item DLTT) plus debt in current liabilities (Compustat Annual item DLC) divided by total assets (Compustat Annual item AT).

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

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

Organizational capital (OC). We construct the stock of a firm’s organizational capital by following the perpetual inventory inventory method described by Eisfeldt and Papanikolaou (2013). That is, we recursively accumulate a firm’s real selling, general and administrative expenses (Compustat Annual item XSGA) over time, and scale the stock of organizational capital by the firm’s total assets (Compustat Annual item AT).

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

TechMark. Recalling equation (OA.3.6), total factor productivity (TFP) is comprised of three distinct components: technology, time-varying markups, and time-varying capacity utilization rates. We isolate the components of TFP related to technology and markups, referred to as “TechMark,” as follows. First, we obtain firm-level estimates of the natural logarithm of TFP from Imrohoroglu and Tuzel (2014). We refer to this variable as $\ln(TFP_{i,t})$. Next, we assign industry-level capacity utilization rates to individual firms by following the matching algorithm described in Section OA.3.3. We take the natural logarithm of these firm firm-level capacity utilization rates, and denote this quantity $\ln(CU_{i,t})$. Finally, we define the TechMark variable for firm i at time t as $\text{TechMark}_{i,t} = \ln(TFP_{i,t}) - \ln(CU_{i,t})$.

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

OA.2 Capacity utilization data and summary statistics

The public report on industrial capacity utilization covers 57 industries. These industries are defined at different levels of aggregation ranging from two- to six-digit North American Industry Classification System (NAICS) codes. Specifically, 12 of the industries are crude aggregates that span *multiple* two-digit NAICS codes. For example, one of these 12 aggregates includes the average capacity utilization rate of all manufacturers in the United States. We remove these 12 crude aggregates from our benchmark sample for two reasons. First, these aggregates do not provide new information as they are spanned by more granularly defined sub-industries that are also included in the sample. Second, these aggregates represents a considerable proportion of total market value and would consequently dominate the returns of the value-weighted portfolios we form in Section 1. Removing these 12 crude aggregates leaves us with a benchmark cross-section of 45 industries that features a mix of durable manufacturers, nondurable manufacturers, and miners and utilities.²⁴

As the 45 industries included in the benchmark sample are defined from the relatively coarse two-digit NAICS code level to the most granular six-digit NAICS code level, the benchmark cross-section includes a number of overlapping industries.²⁵ For instance, the capacity utilization rate of food manufacturers is included in the utilization rate of two industries reported by the FRB: “Food,” as well as “Food, beverage, and tobacco.” Since removing overlapping industries from our benchmark sample would significantly reduce the number of cross-sectional assets, and thus, make

²⁴A list of these 45 industries, along with each industry’s sector, is provided in Table OA.2.1 of the Online Appendix.

²⁵In particular, our final sample consists of one sector defined at the two-digit NAICS level, 27 subsectors defined at the three-digit NAICS level, 13 industry groups defined at the four-digit NAICS level, two industries defined at the five-digit NAICS level, and two U.S. industries defined at the six-digit NAICS level. In Section OA.3 we ensure that our results are robust to this heterogeneity in classification levels.

certain asset pricing tests, such as portfolio double sorts, infeasible, we deal with this overlap in two ways. First, in robustness section OA.3.1 we remove the industries that overlap with others and conduct our baseline empirical tests in a sub-sample of 24 distinct non-overlapping industries, each of which corresponds to a unique three-digit NAICS code. Following the example above, this set of distinct industries includes both “Food” and “Beverage and Tobacco” manufacturers, but excludes the composite index that covers both groups of manufacturers. Table OA.3.3 shows that the utilization premium is significant even within this narrower sample. Second, we verify that our results are not driven by any particular industry that dominates the sample (see Tables OA.3.13).

Monthly utilization data for 32 industries are available in January 1967, and data for an additional 25 industries becomes available in January 1972.²⁶ The utilization data we collect ends in December 2015, when we commenced the empirical analysis of the paper. We verify in Section OA.3.1 that our results hold when we only consider the most recent half of the sample period.

OA.2.1 Summary statistics

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

Figure OA.2.1 shows the annual growth rate of aggregate capacity utilization over the sample period. The figure shows that utilization fluctuates significantly over time and that the growth rate of aggregate utilization is procyclical. The aggregate utilization rate drops during recessions, particularly during the Great Recession. The growth rate of aggregate utilization tends to slightly lead the business cycle, and has often served as an early warning for recessions. In five out of the seven recessions during our sample period the growth rate of utilization begins to drop prior to the start of the recession. The growth rate of capacity utilization increases during the technological revolution of the late 1990’s, the housing bubble, and the recovery from the Great Recession.

As illustrated by equation (1), the capacity utilization rate is a combination of both industrial production and capacity. The former variable is studied extensively in the macroeconomic and finance literature, and features prominently in the context of asset pricing. For instance, Cooper et al. (2008) document a premium for firms with lower total asset growth. The growth rate of assets is directly linked to firms’ output, and is consequently captured by the FRB’s measure of industrial production. To establish the empirical novelty in examining capacity utilization, we examine the extent to which utilization fluctuates independently of industrial production using:

$$\Delta CU_t = \beta_0 + \beta_1 \Delta IP_t + \varepsilon_t. \quad (\text{OA.2.1})$$

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

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

was likely facilitated by the technological advancements of the era (Bansak, Morin and Starr 2007).

Table OA.2.2 reports summary statistics for the capacity utilization rates of each sector in our benchmark sample. The average rate of aggregate capacity utilization rate is 79.91%. This figure for the U.S. is close to the average capacity utilization rates of 81.17%, 84.69%, and 82.49% for the Euro Zone, China, and Israel, respectively.²⁷

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

Among mining industries and utilities the average utilization rate is 84.13%. This average rate is slightly higher than, and statistically different from, the average rate across all industries. Due to this difference in average capacity utilization rates we verify that our empirical results are robust to excluding mining industries and utilities from our sample. We also verify that our results still hold when we conduct tests using the growth rate of utilization that eliminates differences in levels by construction. The results of both of these tests are reported in Section OA.3.

Table OA.2.2 also reports the volatility and autocorrelation of utilization for the different sectors in our sample. The volatility of the capacity utilization rate is comparable across sectors and ranges from 6.67% per annum for mining to 8.29% per annum for durables. The autocorrelation ranges from 0.52 to 0.61, with an all-industry mean of 0.58. These statistics affirm the notion that the level of utilization follows similar dynamics regardless of sector.

Overall, Table OA.2.2 shows that the *unconditional* average rate of utilization is only slightly different between sectors. In particular, most differences between the average utilization rate of a sector and the average aggregate utilization rate are statistically indistinguishable from zero. In contrast to these unconditional differences, the asset pricing tests we conduct rely on *conditional* variation in utilization rates. Thus, our tests exploit the fact that the relative ranking of industries in terms of utilization changes over time. Untabulated results show that if we assume that utilization rates are constant over time, and try to utilize the small unconditional differences in the average rate of utilization between industries to perform the empirical tests, our results cease to hold.

Finally, Table OA.2.3 shows the correlation between capacity utilization and other industry-level production-based characteristics for the average industry in our sample. The characteristics considered include book-to-market, TFP, the hiring rate, sales-to-assets, and the investment rate. While utilization has a positive correlation with productivity, hiring, sales, and investment, these average correlation are fairly low. For example, the average correlation between the investment rate and utilization is only 0.16. This suggests that varying utilization constitutes a separate degree of freedom for managers to smooth dividends, and that any interaction between utilization and expected returns is likely to be independent of the well-established book-to-market and investment rate effects on risk premia.

²⁷See <https://www.dallasfed.org/institute/oecd> for data on these utilization rates recorded by the Organization for Economic Cooperation and Development (OECD) and reported by the Federal Reserve Bank of Dallas.

Table OA.2.1: Sample composition and industry specification

Industry name	Sector	Overlap	Start year
Nonmetallic mineral product	D	No	1967
Primary metal	D	No	1967
Fabricated metal product	D	No	1967
Machinery	D	No	1967
Transportation equipment	D	No	1967
Motor vehicles and parts	D	Yes	1967
Aerospace and miscellaneous transportation eq.	D	Yes	1967
Furniture and related product	D	No	1967
Computers, communications eq., and semiconductors	D	Yes	1967
Wood product	D	No	1972
Iron and steel products	D	Yes	1972
Computer and electronic product	D	No	1972
Computer and peripheral equipment	D	Yes	1972
Communications equipment	D	Yes	1972
Semiconductor and other electronic component	D	Yes	1972
Electrical equipment, appliance, and component	D	No	1972
Automobile and light duty motor vehicle	D	Yes	1972
Miscellaneous	D	No	1972
Food, beverage, and tobacco	ND	Yes	1967
Leather and allied product	ND	No	1967
Paper	ND	No	1967
Petroleum and coal products	ND	No	1967
Chemical	ND	No	1967
Plastics and rubber products	ND	No	1967
Food	ND	No	1972
Beverage and tobacco product	ND	No	1972
Textile mills	ND	No	1972
Textiles and products	ND	Yes	1972
Textile product mills	ND	No	1972
Apparel	ND	No	1972
Apparel and leather goods	ND	Yes	1972
Printing and related support activities	ND	No	1972
Synthetic rubber	ND	Yes	1972
Plastics material and resin	ND	Yes	1972
Artificial and synthetic fibers and filaments	ND	Yes	1972
Mining	MU	No	1967
Metal ore mining	MU	Yes	1967
Nonmetallic mineral mining and quarrying	MU	Yes	1967

Continued on the next page...

Table OA.2.1 – Continued from the previous page

Industry name	Sector	Overlap	Start year
Electric power generation, transmission, and distribution	MU	Yes	1967
Electric and gas utilities	MU	Yes	1967
Natural gas distribution	MU	Yes	1967
Coal mining	MU	Yes	1967
Oil and gas extraction	MU	No	1972
Mining (except oil and gas)	MU	No	1972
Support activities for mining	MU	No	1972

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

Table OA.2.2: Summary statistics of the capacity utilization rate by sector

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

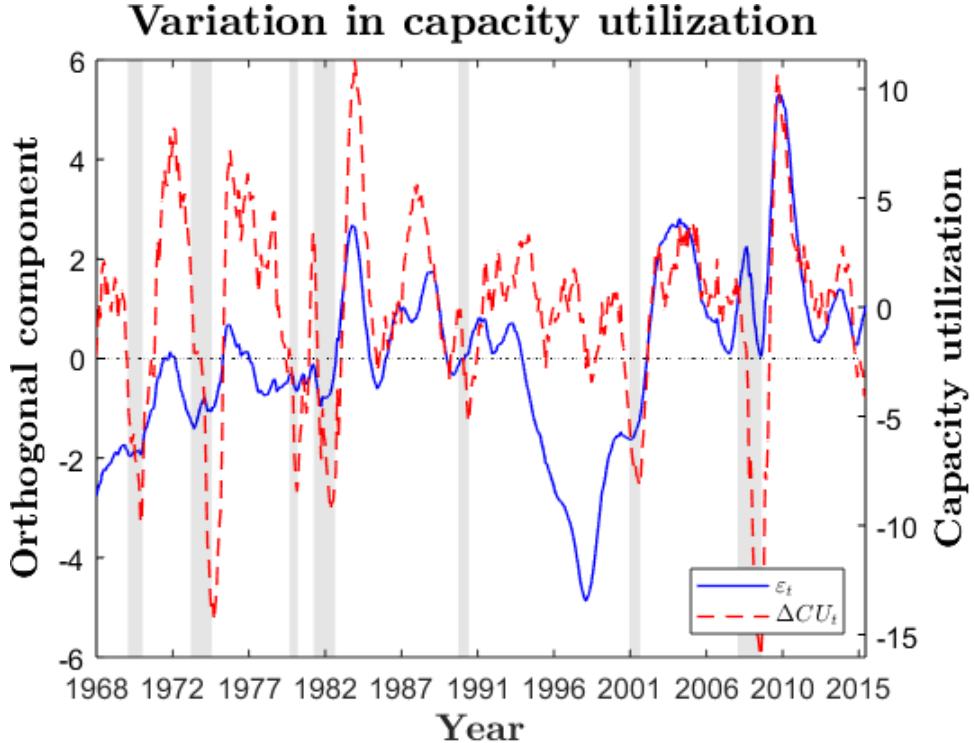
The table reports the mean, standard deviation (SD), and autocorrelation (AC(1)) of the annual capacity utilization rates for each sector in the sample. N represents the number of industries within each sector for which a time-series of capacity utilization data available on FRED. The column $t(\text{Sector-All})$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

Table OA.2.3: **Average industry-level correlation between production-based characteristics**

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

The table shows the correlation between pairs of industry-level characteristics, averaged over all industries in the sample. The characteristics are the capacity utilization rate (CU), the book-to-market ratio (BE/ME), total factor productivity (TFP), the hiring rate (Hire Rate), the ratio of sales-to assets (Sales / Assets), and the investment rate (I/K). At the end of each June from 1967 to 2015, each industry-level characteristic is constructed as the simple average of the characteristic of interest over all firms that belong to the industry at the point in time. For each industry, we then compute the correlation between industry-level characteristic X and industry-level characteristic Y over the sample period, and report the average value of this correlation across all industries in the sample. The data is annual and runs from 1967 to 2015. Additional details on the construction of each variable are provided in Section OA.1 of the Online Appendix.

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



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

OA.3 Additional empirical results

OA.3.1 Methodological variations in portfolio formation

In this section we show that the capacity utilization spread is also robust to several implementation choices related to the portfolio formation procedure described in Section 1.2.

Variation in breakpoints. In the benchmark analysis we use the 10th and 90th percentiles of the cross-sectional distribution of capacity utilization rates as breakpoints for the low and high utilization portfolios. Here, we modify these breakpoints and sort industries into portfolios using either (i) quintiles or (ii) the 30th and 70th percentiles of the cross-sectional distribution of capacity utilization rates. These choices of breakpoints either double or triple the number of industries in each of the low and high capacity utilization portfolios. The value- and equal-weighted returns of the quintile portfolios are reported in Panel A of Table OA.3.1, while the returns obtained by using the 30th and 70th percentiles of the distribution of utilization rates are reported in Panel B. Despite using coarser breakpoints, the value-weighted utilization spread in each panel is close to 5% per annum and statistically significant. This spread is less than 1% smaller in magnitude than the benchmark. Portfolio returns also tend to decrease as the average utilization rate of each portfolio

increases, with the equal-weighted returns showing strict monotonicity in utilization.

Variation in the sample period. While our sample period spans July 1967 to December 2015, we consider the impact of breaking the sample in half and examining the utilization spread in the most recent subsample that starts in July 1991. This subsample analysis is reported in Table OA.3.2 and shows that the magnitude of the spread is larger in the recent subsample than it is over the entire sample period. While the value-weighted spread has a mean return of 5.67% per annum between July 1967 and December 2015, its mean return between July 1991 and December 2015 is 9.09% per annum. As the second half of the sample is populated by two major recessions, the recession of the early 2000s and the Great Recession, this result also shows that the utilization spread is largely countercyclical.

Variation in the sample of industries. Our benchmark results are based on a cross-section of 45 industries. However, as explained in Section OA.2, some of these industries are comprised of firms that belong to multiple industries in the sample. To ensure that our results are not driven by this feature of the data, we repeat our baseline analysis using a subsample of industries whose constituent firms are distinct from one another. These 24 no-overlap industries are listed in Table OA.2.1 and the results of repeating our benchmark portfolio sorts in this subsample of industries are shown in Table OA.3.3. The value-weighted utilization spread is close to 6.65% per annum in this subsample, and even larger than our benchmark spread of 5.7% per annum. The value-weighted (equal-weighted) utilization spread is still statistically significant at the 5% (10%) level.

Relatedly, Table OA.3.13 exclude various sectors of the economy (e.g., durables or mining and utilities) and shows that the utilization premium remains economically and statistically significant. Likewise, Table OA.3.14 shows the premium is robust to sorting on the growth rate of utilization thereby eliminating industry fixed effect in utilization's level.

Collectively, the evidence clearly shows, to the best of our ability and available data, that the utilization premium is not sensitive to the specific choice of industries comprising the sample.

Quarterly portfolio sorting. Table OA.3.4 shows that forming portfolios on a quarterly, rather than an annual basis also leads to an economically large and statistically significant capacity utilization spread. Here, we form portfolios at the end of each March, June, September, and December using the lagged capacity utilization rate, and hold each portfolio for three months before rebalancing. The value-weighted (equal-weighted) capacity utilization spread is 6.45% (5.23%) per annum. Both are statistically significant at the 5% level.

Importance of conditional sorting. In untabulated results we demonstrate the importance of the *conditional* portfolio sorting procedure described in Section 1.2. Specifically, we consider an alternative procedure in which each industry is permanently assigned to the first portfolio it is sorted into. This *unconditional* portfolio sort leads to a capacity utilization spread that is both economically and statistically insignificant. This result highlights that there is a significant degree of conditional variation in industry-level capacity utilization rates, and that this variation is important for generating the capacity utilization spread.

Table OA.3.1: Capacity utilization spread: alternative portfolio breakpoints

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

Panel B: 30 th and 70 th percentile breakpoints					
Portfolio	Value-weighted		Equal-weighted		
	Mean	SD	Mean	SD	
Low (L)	12.62	17.86	9.66	19.72	
Medium	10.09	18.43	8.51	17.85	
High (H)	7.75	17.18	6.22	17.86	
Spread (L-H)	4.86 (2.47)	13.44	3.44 (2.26)	10.43	

The table reports annual returns of portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios. The construction of the portfolios is identical to the benchmark analysis, except for the portfolio breakpoints used. In Panel A, we sort the cross-section of industries into five portfolios based on quintile breakpoints. In Panel B, we sort the cross-section of industries into three portfolios based on the 30th and 70th percentiles of the cross-sectional distribution of capacity utilization rates. Mean refers to the average annual return and SD denotes the standard deviation of annual returns. Parentheses report *t*-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1991 to 2015 and are rebalanced annually.

Table OA.3.2: Capacity utilization spread: results based on the recent subsample

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

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

Table OA.3.3: Capacity utilization spread: results based on non-overlapping industries

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

The table reports annual returns of five portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios. The construction of the portfolios is identical to the benchmark analysis, except that the sample of industries is restricted to those industries whose constituents do not belong to multiple industries in the sample (see Table OA.2.1, Column 3, for the list of these non-overlapping industries). Mean refers to the average annual return, SD denotes the standard deviation of annual returns. Parentheses report *t*-statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each June from 1967 to 2015 and are rebalanced annually, with portfolio returns spanning July 1967 to December 2015.

Table OA.3.4: Capacity utilization spread: quarterly portfolio formation

Portfolio	Value-weighted		Equal-weighted	
	Mean	SD	Mean	SD
Low (L)	13.66	21.91	10.33	21.52
Medium	10.69	16.65	8.23	17.55
High (H)	7.20	20.48	5.10	21.00
Spread (L-H)	6.45 (2.41)	18.57	5.23 (2.21)	16.55

The table reports annual returns of three portfolios sorted on the basis of capacity utilization, as well as the spread between the low (L) and high (H) capacity utilization portfolios. Here, portfolios are formed at the end of each March, June, September, and December on the basis of each industry's lagged capacity utilization rate. Each of these portfolios is then held for three months before rebalancing. Mean refers to the average annual return, SD denotes the standard deviation of annual returns. Parentheses report t -statistics computed using Newey and West (1987) standard errors. The portfolios are formed at the end of each quarter from 1967 to 2015, with portfolio returns spanning April 1967 to December 2015.

OA.3.2 Conditional CAPM and time-varying betas to aggregate productivity

In this section we show that time-varying exposures to the market (aggregate productivity) fully absorb the utilization premium.

Conditional CAPM. We estimate the conditional CAPM to show that time variation in risk exposures are important to reconcile the utilization premium. Here, we estimate the conditional CAPM following the procedure outlined by Lewellen and Nagel (2006). That is, *within* each year of the sample period, we estimate an unconditional CAPM regression that projects each portfolio's excess returns on excess market returns. We then compute the time-series average of the corresponding CAPM alphas and betas. The results, reported in Table OA.3.5, confirm that a single-factor model is able to explain the utilization premium. Accounting for time-variation in CAPM betas *reduces* the economic magnitude of the average CAPM alphas compared to the unconditional specification and renders the CAPM alpha statistically insignificant.

Table OA.3.5: Capacity utilization spread: conditional CAPM alphas and betas

Portfolio	CAPM alpha		CAPM beta	
	α	$t(\alpha)$	β	$t(\beta)$
Low (L)	0.81	(0.33)	1.15	(9.66)
Medium	-1.05	(-0.48)	1.06	(11.70)
High (H)	-3.99	(-0.57)	0.92	(10.85)
Spread (L-H)	3.43	(1.34)	0.18	(2.35)

The table reports the alphas and betas obtained from estimating the conditional CAPM following the methodology proposed by Lewellen and Nagel (2006). We implement this analysis to obtain the conditional CAPM alphas and betas of each value-weighted utilization-sorted portfolio in two steps as follows. First, within each year of the sample period, we estimate an unconditional CAPM regression and record the alpha and beta associated with each portfolio. Second, we compute the time-series average of the estimated alphas and betas, and report these time-series averages in the table. We annualize each alpha by multiplying the time-series average of the monthly alphas obtained in the second step of the procedure by 12. Parentheses report Newey and West (1987) t -statistics, and the sample is from July 1967 to December 2015.

Non-linear model exposures. We complement the former evidence using other proxies of aggregate productivity and a non-linear model specification. By construction, the exposure of each portfolio to aggregate productivity varies over time and with the business cycle (if the loading on the non-linear term is non-zero). We augment equation (2) with a quadratic aggregate productivity term and estimate the following regression:

$$Ret_{i,t}^e = \beta_{0,i} + \beta_{1,i} \text{Agg-Prod}_t + \beta_{2,i} \text{Agg-Prod}_t^2 + \varepsilon_{i,t}. \quad (\text{OA.3.1})$$

Here, $Ret_{i,t}^e$ is the value-weighted excess return of the portfolio of interest, Agg-Prod_t is a proxy for aggregate productivity, and $\beta_{1,i}$ ($\beta_{2,i}$) captures the exposure of portfolio i to the linear (quadratic) effect of aggregate productivity.

We implement this analysis by considering three proxies for aggregate productivity: (i) the market return, (ii) utilization-adjusted TFP growth from Fernald (2012), and (iii) labor productivity from the BLS. We also combine the slope coefficients on the linear and quadratic terms to form the (total) productivity beta of portfolio i (denoted by $\beta_{i,prod}$) as:

$$\beta_{i,prod} = E \left[\frac{\partial Ret_i^e}{\partial \text{Agg-Prod}} \right] = \beta_{1,i} + 2\beta_{2,i} E[\text{Agg-Prod}], \quad (\text{OA.3.2})$$

and compute the standard errors associated with $\beta_{i,prod}$ using the Delta method. We report the results of this analysis in Table OA.3.6.

Table OA.3.6: **Exposure of CU-sorted portfolios to aggregate productivity proxies**

Portfolio	Market returns		Util.-adjusted TFP		Labor productivity	
	β_{prod}	$t(\beta_{prod})$	β_{prod}	$t(\beta_{prod})$	β_{prod}	$t(\beta_{prod})$
Low (L)	1.37	(9.12)	1.18	(2.78)	0.88	(1.81)
Medium	1.25	(11.12)	0.76	(2.27)	0.54	(1.33)
High (H)	1.07	(6.53)	0.78	(2.02)	0.54	(1.18)
Spread (L-H)	0.30	(3.53)	0.40	(1.91)	0.34	(1.94)
Intercept	1.27	(0.41)	4.46	(1.33)	3.08	(0.99)

The table reports the exposures of portfolios sorted on capacity utilization to three different aggregate productivity proxies. The regression we estimate is: $Ret_{i,t}^e = \beta_0 + \beta_{1,i} \text{Agg-Prod}_t + \beta_{2,i} \text{Agg-Prod}_t^2 + \varepsilon_{i,t}$, where $Ret_{i,t}^e$ is the value-weighted excess return of portfolio i , Agg-Proxy is a proxy of aggregate productivity, and β_1 (β_2) is the sensitivity of portfolio i 's excess return to the linear (quadratic) aggregate productivity term. We combine the linear and quadratic sensitivities to form the productivity beta (β_{prod}) following equation (OA.3.2), and report β_{prod} in the table. Here, Agg-Proxy 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 associated with each exposure are reported in parentheses, with the standard errors associated with β_{prod} computed using the delta method. “Intercept” refers to the annualized value of β_0 (obtained by multiplying β_0 by four) from projecting the utilization spread on each productivity proxy. Finally, the sample spans July 1967 to December 2015.

Comparing Table OA.3.6, which features a non-linear relation to aggregate productivity, to Table 2, which features only a linear relation, delivers three key takeaways. First, accounting for the non-linear term, the low utilization portfolio's exposure to aggregate productivity remains significantly larger than the high utilization portfolio's exposure to productivity. Second, with non-linear exposures to aggregate productivity, the differences in productivity betas (L-H) in Table OA.3.6 are at least as large, if not larger, than those reported in Table 2. Third, when we measure aggregate productivity using excess market returns, the economic magnitude of the intercept falls to 1.27% per annum and becomes statistically insignificant with a t -statistic of 0.41. Likewise, the intercept from projecting the utilization premium on utilization-adjusted TFP growth from Fernald (2012) becomes statistically insignificant.

Relation between conditional exposure and utilization. We show that capacity utilization correlates with industries' *conditional* exposure to aggregate productivity (CAPM betas). We establish this link in two steps. First, for each industry j and each month t , we estimate the following time-series regression using the past 60 months of stock return data

$$R_{j,t-60 \rightarrow t}^e = \alpha_{j,t} + \beta_{j,t} MKTRF_{t-60 \rightarrow t} + \varepsilon_{j,t-60 \rightarrow t}, \quad (\text{OA.3.3})$$

where, $R_{j,t-60 \rightarrow t}^e$ denotes the excess stock return of industry j , $MKTRF$ represents the market excess return, and $\beta_{j,t}$ is the conditional exposure of industry j to aggregate productivity at time t . Estimating this time-series regression for each industry and each month of our sample delivers a panel containing each industry's conditional risk exposure between July 1972 and December 2015.

Next, we estimate the following panel regression that projects the risk exposures onto capacity utilization, and other monthly industry-level variables, observable at time t

$$\beta_{j,t} = \alpha_j + \delta_t + \beta_{CU} CU_{j,t} + \boldsymbol{\beta}' \mathbf{X}_{j,t} + \varepsilon_{j,t}. \quad (\text{OA.3.4})$$

Here, $CU_{j,t}$ denotes the utilization rate of industry j in month t . β_{CU} captures the degree to which

capacity utilization is, on average, informative about conditional risk exposures. $\mathbf{X}_{j,t}$ is a vector of other industry-level controls that vary at the monthly frequency, including the size, idiosyncratic volatility, and momentum of industry j . α_j is an industry fixed effect, and δ_t is a time fixed effect that absorbs changes in conditional risk exposures common to all industries. We standardize each independent variable (to aid the interpretation of the marginal effects across specifications) and compute t -statistics using standard errors clustered by both industry and time.

Panel A of Table OA.3.7 shows that without controlling for other industry-level characteristics, the capacity utilization rate is highly informative about an industry's conditional CAPM beta. For instance, Column 1 shows that a one standard deviation increase in capacity utilization is associated with a -0.09 decrease in an industry's conditional market beta. Controlling for both time and industry fixed effects in Column 3 yields similar results. Moreover, Panel B indicates that the negative association between capacity utilization rates and conditional market betas remains significant when additional control variables are added to each specification.

Table OA.3.7: Sensitivity of conditional CAPM betas to capacity utilization

	Panel A: No controls			Panel B: Controls		
	(1)	(2)	(3)	(4)	(5)	(6)
CU	-0.09 (-3.87)	-0.09 (-3.09)	-0.05 (-2.33)	-0.09 (-3.86)	-0.07 (-2.76)	-0.05 (-2.38)
Size				-0.35 (-1.22)	-1.00 (-1.70)	0.20 (1.66)
IVOL				0.49 (1.93)	0.99 (1.62)	0.02 (0.22)
Momentum				-0.07 (-0.65)	0.07 (0.29)	-0.15 (-4.81)
Time FE	No	Yes	Yes	No	Yes	Yes
Industry FE	No	No	Yes	No	No	Yes
R2	0.05	0.12	0.55	0.07	0.17	0.55
Obs.	23363	23363	23363	23363	23363	23363

The reports the relation between conditional CAPM betas, capacity utilization rates, and other industry-level characteristics. We implement this analysis in two steps. In the first step, we obtain the conditional CAPM beta of each industry j in each month t by estimating the time-series regression represented by equation (OA.3.3) using the past 60 months of stock return data. In the second step, we project the conditional CAPM beta of each industry j at each time t on a set of industry-level characteristics including the industry's (i) capacity utilization rate, (ii) size, (iii) idiosyncratic return volatility, and (iv) return momentum. The specification of the panel regression we use in this second step is represented by equation (OA.3.4). The regressions in Panel A feature no additional control variables beyond the capacity utilization rate, whereas the regressions in Panel B include all control variables. We scale each independent variable by its unconditional standard deviation prior to estimating and reporting each slope coefficient. Parentheses represent t -statistics computed using standard errors clustered by industry and month. Finally, the sample period is from July 1967 to December 2015.

Non-linear model: robustness using quintiles. Table OA.3.8 repeats the non-linear model analysis using five, rather than three, portfolios sorted on utilization rates. We use the market portfolio as a proxy for aggregate productivity. Even with quintile portfolios: (i) productivity exposures tend to decline with utilization, and (ii) the linear and non-linear productivity exposures of the low utilization portfolio are significantly higher than those of the high utilization portfolio.

Panel B also shows that the non-linear alpha is only 0.67% per annum and statistically insignificant.

Table OA.3.8: **Exposures of quintile portfolios to aggregate productivity**

Portfolio	Panel A: Linear		Panel B: Non-linear	
	β_{prod}	$t(\beta_{prod})$	β_{prod}	$t(\beta_{prod})$
Low (L)	1.29	(11.76)	1.33	(10.70)
2	1.27	(11.68)	1.30	(10.68)
Medium	1.28	(10.47)	1.28	(9.53)
4	1.16	(10.96)	1.19	(10.38)
High (H)	1.08	(8.72)	1.08	(7.38)
Spread (L-H)	0.21	(2.74)	0.25	(3.00)
Intercept (L-H)	3.82	(1.79)	0.67	(0.26)

The table reports the exposures of portfolios sorted on capacity utilization to aggregate productivity. The regression we estimate is: $Ret_{i,t}^e = \beta_{0,i} + \beta_{1,i} \text{Agg-Prod}_t + \beta_{2,i} \text{Agg-Prod}_t^2 + \varepsilon_{i,t}$, where $Ret_{i,t}^e$ is the value-weighted excess return of portfolio i , Agg-Proxy is the aggregate productivity proxy, and β_1 (β_2) is the sensitivity of portfolio i 's excess return to the linear (quadratic) aggregate productivity term. We combine the linear and quadratic sensitivities to form the productivity beta (β_{prod}) following equation (OA.3.2), and report β_{prod} in the table. Here, Agg-Proxy is measured using excess market returns. Monthly returns are aggregated to the quarterly frequency so that each regression is estimated using quarterly data. Newey and West (1987) t -statistics associated with each exposure are reported in parentheses, with the standard errors associated with β_{prod} computed using the delta method. “Intercept” refers to the annualized value of β_0 (obtained by multiplying β_0 by four) from projecting the utilization spread on the productivity proxy. Finally, the sample spans July 1967 to December 2015.

OA.3.3 Assigning industry-level utilization rates to CRSP/Compustat firms

When applicable, we assign each firm a capacity utilization rate that corresponds to the utilization rate of the industry to which the firm belongs. However, recall from Section OA.2 that our sample is comprised of industries that are defined with different degrees of granularity. This means that some firms may be matched to more than one industry. We execute the following matching algorithm to ensure that each firm is matched to the most granularity defined industry to which it belongs. We start by assigning utilization rates to all firms that belong to a six-digit NAICS code industry for which utilization data is available. We then consider the five-digit NAICS code industries and identify the constituents of these industries that were not previously assigned a utilization rate. These firms are then assigned a utilization rate corresponding to a five-digit NAICS code industry. This procedure then continues to the four-, three-, and two-digit NAICS code industries, in that order. If a previously unmatched firm belongs to two or more N -digit NAICS code industries, then we assign the firm the utilization rate of its “parent” ($N-1$)-digit NAICS code industry. Any firms unmatched at the end of this procedure are removed from the sample.

OA.3.4 Independence from value, investment, and organizational capital

Independence using Fama-Macbeth regressions. Table 5 reports the results of the Fama and MacBeth (1973) regression analysis outlined by equation (3). Columns (1) to (6) show that the coefficient on utilization is negative and significant at the 5% level when we include another investment-related characteristic in the regressions, including TFP, hiring rate, investment rate,

capital overhang, and book-to-market ratios. The signs of the average slope coefficients of the other characteristics are consistent with the spreads associated with each characteristic of interest.²⁸

Columns (7) to (9) control for several investment-related characteristics jointly. In all cases, utilization's predictive power for returns remains significant at the 5% level. Lastly, Column (10) considers all of the aforementioned characteristics in addition to size, organizational capital, and past returns.²⁹ Compared to Columns (1) to (9), the slope coefficient on utilization in Column (10) is largely similar. The slope coefficients on the other characteristics also remain similar, with the exception of the loading on TFP that flips sign from negative to positive. However, this change in sign does not compromise the validity of the TFP spread as a number of the investment-related characteristics included in this specification are relatively highly correlated. In all, the regression evidence shows that the relation between utilization and stock returns is materially distinct from to the known relations between returns and each of TFP, hiring, investment, and B/M.

Independence using portfolio double sorts. We corroborate the results of the Fama and MacBeth (1973) analysis by conducting conditional portfolio double sorts. That is, we conditionally sort the sample of industries into portfolios along two dimensions. The first dimension corresponds to either (i) book-to-market ratios, (ii) physical investment rates, or (iii) ratios of organizational capital to assets, while the second dimension reflects the capacity utilization rate.

The motivation behind this exercise is to control for one variable of interest (e.g., B/M) and show that controlling for the first dimension, the utilization premium remains significant. We focus on book-to-market ratios and investment rates as these are the only two characteristics in Panel B of Table 3 that are significantly different between the extreme utilization portfolios and also command a risk premium that is aligned with the utilization spread. While the difference in organizational capital is insignificant across the utilization portfolios, we conduct this double sort to rule out the possibility that low utilization is mechanically driven by a higher reliance on intangible capital.

The double sorts are implemented as follows. At the end of each June from 1967 to 2015 we first sort the cross-section of firms into three portfolios based on either their (i) book-to-market ratios, (ii) investment rates, or (iii) organizational capital-to-asset ratios. We use the 30th and 70th percentiles of the firm-level cross-sectional distribution of each characteristic to assign each firm to one of three portfolios. Next, within each of these three characteristic-sorted portfolios, we further sort firms into three additional portfolios on the basis of capacity utilization. We also use the 30th and 70th percentiles of the cross-sectional distribution of capacity utilization rates in March of the same year to determine portfolio membership in this second step. This process produces nine portfolios that are each held from the beginning of July in year t to the end of June in year $t + 1$, at which point in time all portfolios are rebalanced.³⁰

Table OA.3.9 reports the results of the bivariate portfolio sorts on the basis of both value- and

²⁸Imrohoroglu and Tuzel (2014) show that low TFP predicts high returns, Belo et al. (2014) find low hiring is associated with high returns, Titman, Wei, and Xie (2004) documents the relation between low investment rates and high returns, Aretz and Pope (2018) find higher capacity overhang predicts lower returns, and Fama and French (1993) discuss how both low market capitalization and high market-to-book ratios predict high returns.

²⁹Untabulated results show that adding sector fixed effects to column (10) of Table 5 produces quantitatively similar results.

³⁰The portfolio breakpoints used in the bivariate sorting procedure described above (the 30th and 70th percentiles) are necessary to ensure that there is a sufficient number of firms in each of the *nine* doubles-sorted portfolios (note that Table OA.3.1 shows that the utilization premium is sizable and significant using a univariate sort that is based on these breakpoints).

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

The results show that after controlling for either book-to-market ratios, investment rate, or organizational capital, the capacity utilization spread remains positive in 15 out of 18 cases. The utilization spread is also quantitatively large and statistically significant in most cases.

Panel A shows that, keeping book-to-market ratios relatively constant, the equal-weighted capacity utilization spread is significantly different from zero at the 10% level within the low book-to-market portfolio and at the 5% level for both the medium and the high book-to-market portfolios. The joint p -value across the three spread portfolios is under 8%. The value-weighted returns reported in Panel B show that the capacity utilization spread is most pronounced among growth firms. Within this low book-to-market portfolio, the capacity utilization spread exceeds 6% per annum. The p -value of 0.016 associated with the joint test in Panel B shows that the three value-weighted utilization spreads are statistically significant after conditioning on book-to-market.

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

Lastly, Panels E and F indicate that the utilization spread exists *within* organizational capital portfolios. In particular, the value-weighted utilization spread among firms with medium amounts of organization capital exceeds 4% per annum, and is statistically significant at better than the 1% level. Moreover, the joint test reported in each panel is rejected at the 5% level or better.

Table OA.3.9: Controlling for book-to-market ratios, investment rates, and organizational capital: double-sort analysis

		Panel A: Capacity Utilization (EW)					Panel B: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
BE/ME	Low (L)	9.40	8.10	6.26	3.14	(p=0.079)	11.63	10.53	5.55	6.08	(p=0.001)
	Medium	16.92	13.36	12.82	4.10	(p=0.012)	13.29	10.45	11.01	2.27	(p=0.084)
	High (H)	19.89	17.49	16.66	3.23	(p=0.039)	14.47	12.70	14.49	-0.02	(p=0.503)
		Joint test			(p=0.076)		Joint test			(p=0.016)	
		Panel C: Capacity Utilization (EW)					Panel D: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
I/K	Low (L)	20.06	17.74	12.62	7.44	(p < 0.001)	15.52	11.30	10.55	4.97	(p=0.007)
	Medium	17.07	13.65	13.08	3.99	(p=0.005)	13.35	10.64	9.07	4.28	(p=0.008)
	High (H)	10.60	7.73	9.02	1.59	(p=0.249)	9.96	9.26	7.74	2.23	(p=0.180)
		Joint test			(p < 0.001)		Joint test			(p=0.045)	
		Panel E: Capacity Utilization (EW)					Panel F: Capacity Utilization (VW)				
		Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)	Low (L)	Medium	High (H)	Spread(L-H)	p(Spread)
OC / AT	Low (L)	10.39	7.70	13.27	-2.87	(p=0.861)	10.18	10.11	11.16	-0.98	(p=0.661)
	Medium	17.30	13.55	14.52	2.78	(p=0.021)	15.57	12.46	11.31	4.26	(p=0.007)
	High (H)	20.19	17.50	16.60	3.59	(p=0.027)	16.57	12.20	14.63	1.94	(p=0.240)
		Joint test			(p=0.018)		Joint test			(p=0.038)	

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

Independence using projections. As an additional method of showing the independence between the utilization premium and other spread, we consider the following set of projections that regress the utilization premium at time t on another spread, including a constant

$$Utilization\text{premium}_t = \alpha + \beta X_t^{Other} + \varepsilon_t. \quad (\text{OA.3.5})$$

Here, X_t^{Other} represents the returns of the book-to-market, TFP, investment, size, profitability, momentum, or idiosyncratic volatility spread at the t . If the utilization premium is simply a linear transformation of the value premium (or any one of the other spreads represented by X_t^{Other}), then the constant in this regression (α) will be zero. In contrast, the results in Panel A of Table OA.3.20 indicate that the utilization premium is economically and statistically distinct from these spreads. For instance, projecting the utilization premium on the value premium results in an alpha of 5.11%

per annum (t -statistic of 2.04). Thus, the utilization premium is materially distinct from a host of (potentially related) risk premia.

OA.3.5 Independence from capital overhang

Aretz and Pope (2018) document that firms with higher capital overhang, or firms' whose installed productive capacities exceed their optimal amounts of capacity, have lower expected returns. The authors refer to these firms as possessing "capacity overhang." While the Fama and MacBeth (1973) regressions in Section 1.6 show that the utilization premium and the overhang spread are empirically distinct, the conceptual similarity between these margins motivates us to discuss how the notion of capacity utilization materially differs from that of capacity overhang. We also complement the regression analysis by showing that utilization and overhang each have a distinct impact on stock returns using portfolio double sorts.

Recalling equation (1), capacity utilization is defined as the ratio of a firm's actual output to its maximum potential output (its capacity). On the other hand, capacity overhang is the difference between a firm's *installed* capital stock and its *optimal* (value maximizing) level of capital. Intuitively, capacity utilization and capacity overhang are negatively related since a firm that desires to downscale can reduce its output by lowering the utilization of its existing capital. At the same time, the level of the firm's optimal capital stock also drops. If capital adjustments are not frictionless, then these frictions create a wedge between installed and optimal capacity, resulting in capacity overhang. Consequently, capacity utilization tends to decrease at the same time that overhang tends to increase.

The negative correlation between utilization and overhang is neither theoretically perfect nor empirically large in magnitude. Theoretically, the reason for this less than perfect correlation is that low capacity utilization is a result of a *costless and optimal* policy to keep some machines idle.³¹ This optimal decision to reduce the utilization of capital does not hinge on any installation frictions or adjustment costs. In contrast, capacity overhang depends crucially on the degree to which investment is irreversible, as influenced by frictions such as convex adjustment costs. While low capacity utilization is optimal in states of low productivity, a non-zero amount of capacity overhang can never represent the first-best outcome for a firm. Consequently, capacity overhang should always be zero in a frictionless economy, whereas capacity utilization may still fluctuate depending on a firm's productivity.

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

³¹Keeping machines idle in bad states is not only costless, but may also benefit the firm by preserving capital for future use in more productive states.

To implement this analysis we construct a measure of firm-level capital overhang based on the statistical procedure described by Aretz and Pope (2018), summarized in Section OA.1 of the Online Appendix. Following the discussion on the conceptual relation between capacity utilization and capacity overhang, Table OA.3.10 shows the correlation between overhang and utilization for each industry in our sample.³² The magnitude of the correlation between the two variables decreases with the degree of aggregation. When we aggregate all firms in our sample, the correlation between capacity utilization and capacity overhang is negative, as expected, and amounts to -0.52. When we compute the correlation between these two variables on an industry-by-industry basis and average these pairwise correlations, the result is a modest average correlation of -0.32. The 95% confidence interval for this cross-sectional correlation shows a high degree of dispersion and ranges from -0.71 to 0.11. Panel C of this table reports that the average firm-level correlation drops to -0.11, and shows that this correlation becomes even more dispersed in the cross-section of firms. These results collectively highlight the fact that while capacity utilization and overhang are conceptually negatively related, the empirical correlation between these two variables is low.

Table OA.3.11 reports the results of performing portfolio double sorts along the dimensions of capacity utilization and capacity overhang using the conditional double-sort analysis as described in Section OA.3.4. Panel A shows the average annual capacity utilization spread within three capacity overhang sorted portfolios when all returns are equal-weighted. The capacity utilization spread is positive and statistically significant within each overhang portfolio. The utilization spread is also jointly significant across all three overhang portfolios. Panel B shows that the results are similar when returns are value-weighted. Panels C and D report that the results are largely similar after changing the order of the sorts. Controlling for capacity utilization, the joint tests in Panels C and D show that the capacity overhang spread is positive and statistically significant on an equal-weighted basis, but is insignificant on a value-weighted basis.

³²The industry-level capacity overhang measure is obtained by computing the average overhang for all firms that belong to each industry at each point in time. We also note that our sample is only comprised of manufacturing, mining, and utilities firms, whereas the sample of Aretz and Pope (2018) includes the entire Compustat universe, excluding financial firms and utilities.

Table OA.3.10: Correlation between capacity utilization and capacity overhang

Panel A: Correlation by Industry		
Industry name	Sector	$\rho_{CU,OVER}$
Food, beverage, and tobacco	ND	-0.741
Printing and related support activities	ND	-0.728
Textile mills	ND	-0.690
Wood product	D	-0.687
Textiles and products	ND	-0.683
Beverage and tobacco product	ND	-0.642
Textile product mills	ND	-0.543
Computer and electronic product	D	-0.537
Food	ND	-0.534
Machinery	D	-0.528
Nonmetallic mineral product	D	-0.502
Support activities for mining	MU	-0.500
Coal mining	MU	-0.472
Computers, communications eq., and semiconductors	D	-0.469
Metal ore mining	MU	-0.450
Communications equipment	D	-0.392
Paper	ND	-0.366
Mining	MU	-0.348
Leather and allied product	ND	-0.345
Transportation equipment	D	-0.342
Mining (except oil and gas)	MU	-0.342
Semiconductor and other electronic component	D	-0.329
Automobile and light duty motor vehicle	D	-0.310
Motor vehicles and parts	D	-0.300
Primary metal	D	-0.254
Artificial and synthetic fibers and filaments	ND	-0.250
Chemical	ND	-0.248
Fabricated metal product	D	-0.214
Electrical equipment, appliance, and component	D	-0.195
Aerospace and miscellaneous transportation eq.	D	-0.183
Computer and peripheral equipment	D	-0.182
Apparel	ND	-0.170

Continued on the next page...

Table OA.3.10 – Continued from the previous page

Panel A: Correlation by Industry		
Industry name	Sector	$\rho_{CU,OVER}$
Nonmetallic mineral mining and quarrying	MU	-0.144
Apparel and leather goods	ND	-0.140
Furniture and related product	D	-0.134
Plastics and rubber products	ND	-0.047
Plastics material and resin	ND	0.002
Iron and steel products	D	0.004
Petroleum and coal products	ND	0.071
Miscellaneous	D	0.083
Oil and gas extraction	MU	0.157
Synthetic rubber	ND	0.242

Panel B: Industry-level Summary Statistics				
Statistic	Mean	Median	p5	p95
$\rho_{CU,OVER}$	-0.32	-0.34	-0.71	0.11

Panel C: Firm-level Summary Statistics				
$\rho_{CU,OVER}$	-0.11	-0.13	-0.66	0.51

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

Table OA.3.11: Double-sorted portfolios: capacity utilization versus capacity overhang

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

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

OA.3.6 Dissecting the TFP spread

Imrohoroglu and Tuzel (2014) show that low productivity (TFP) firms earn a high risk premium. While the results of Fama and MacBeth (1973) regressions reported in Table 5 show that utilization predicts return beyond TFP, this section explores the relation and the distinction between the capacity utilization premium and the productivity premium. The relation between the two can be motivated by the general form of a firm’s production function:

$$Y = \underbrace{\text{Technology} \times \text{Markups} \times \text{Utilization}}_{\text{Total factor productivity (TFP)}} \cdot F(K, L), \quad (\text{OA.3.6})$$

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

We begin by examining whether the TFP spread exists in our sample of manufacturing firms, mining firms, and utilities. This is necessary because our sample is more constrained compared to the sample in Imrohoroglu and Tuzel (2014). The results of replicating the TFP spread in our subsample of firms are reported in Panel A of Table OA.3.12. The equal-weighted TFP spread amount to 4.22% per annum and is statistically significant.³³

³³The value-weighted TFP spread is positive yet statistically insignificant using our subsample and time frame.

Table OA.3.12: Dissecting the productivity spread

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

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

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

Panel D: Unconditional correlations					
	$\rho(CU, TFP)$	$\rho(CU, TechMark)$	$\rho(TFP, TechMark)$		
	0.39	0.28	0.96		

Panel A reports both the annual returns of value- and equal-weighted portfolios formed on total factor productivity (TFP), and the spread between low and high TFP (or productivity) portfolios. Mean (SD) refers to the average (standard deviation) of annual returns, and parentheses report Newey and West (1987) robust *t*-statistics. Panel B reports equal-weighted portfolio returns obtained from a double sort procedure in which firms are first sorted into three portfolios on the basis of capacity utilization (CU). Within each portfolio, firms are further sorted into three portfolios on the basis of TFP. The rightmost column of the panel show the *p*-value from a test on the null hypothesis that each TFP spread is zero, as well as a test on null hypothesis that the three spreads are jointly equal to zero. Panel C reports the annual returns of three portfolios sorted on the technology and markups (TechMark) component of TFP. In each of Panels A, B, and C, portfolio breakpoints are based on the 30th and 70th percentiles of the cross-sectional distribution of the characteristic of interest. Panel D shows the pairwise correlations between equal-weighted univariate spreads formed on CU, TechMark, and TFP. The sample period is from July 1967 to June 2015, when the TFP data becomes unavailable. Additional details on the construction of each variable are provided in Section OA.1 of the Online Appendix.

Since capacity utilization is an underlying fundamental component of TFP, we begin by examining whether the TFP spread remains positive when controlling for capacity utilization. We conduct this analysis using a firm-level dependent double sort as described in Section OA.3.4. In other words, we construct the productivity spread within capacity utilization sorted portfolios. The results are reported in Panel B of Table OA.3.12 and show that the TFP spread is 4.56%, 4.21%,

For this reason we only focus on equal-weighted returns in this subsection.

and 1.78% per annum within the portfolio of firms with low, medium, and high rates of capacity utilization, respectively. A joint test on the magnitude of the productivity premium across the three capacity utilization portfolios is statistically significant at the 5% level. This suggests that the productivity premium is distinct from the capacity utilization spread.

Next, we construct a measure for the technology and markup (TechMark) components of TFP by taking the difference between TFP and the capacity utilization rate, as motivated by equation (OA.3.6). This allows us to isolate the component of TFP that is separate from capacity utilization, and examine the relation between this orthogonal component of TFP and stock returns. We sort firms into portfolios based on the TechMark measure at the end of each June and report the results of these univariate sorts in Panel C of Table OA.3.12. The annualized spread between low and high TechMark firms is 3.29% and statistically significant.

Taken together, the results above indicate that the TFP premium is driven by two *distinct* underlying spreads: the TechMark and utilization spreads. Each of these spreads is statistically significant and economically large. We shed light on the contribution of each of these components to the overall productivity spread in Panel D of Table OA.3.12. This panel shows that the correlation between the TFP spread and the capacity utilization (TechMark) spread is 0.39 (0.96).

OA.3.7 Controlling for sectoral effects: Within-sector spread

Table 4 shows that some durable industries are often sorted into the low utilization portfolio, whereas mining industries and utilities often exhibit high capacity utilization rates. The former fact raises the concern that the utilization premium may be a manifestation of the durability spread of Gomes, Kogan, and Yogo (2009). That is, the utilization spread may reflect the know fact that durable manufacturers are riskier than nondurable manufacturers. The latter fact raises the concern that the utilization spread is dominated by one particular sector and may reflect ex-ante heterogeneity between different sectors, as opposed to reflecting a risk premium that exists within sectors. We alleviate both concerns below.

First, only three (two) of the five industries that are most commonly sorted into the low (high) capacity utilization portfolio are durable (nondurable) manufacturers (recall Table 4). Furthermore, the most common industry constituents of the high capacity utilization portfolio are not nondurable manufacturers, as may be expected if the utilization spread were strongly associated with the durability premium.

Second, in the left panel of Table OA.3.13 we examine the utilization premium within a subsample of industries that only includes durable manufacturers. Specifically, we sort the cross-section of 18 durable manufacturers into three portfolios based on the level of capacity utilization following our benchmark sorting procedure. The capacity utilization spread *within* this subsample of durable manufacturers amounts to 5.85% per annum, and is statistically significant. This demonstrates that the utilization spread is also a within-sector phenomenon that is materially unrelated to the ex-ante heterogeneous exposures of durable and nondurable manufacturers to aggregate risk.

Table OA.3.13: Capacity utilization spread: inclusion and exclusion of major sectors

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

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

Third, we examine the magnitude of the capacity utilization spread when we exclude the only sector that heavily populates the high utilization portfolio: mining and utilities. The mining and utilities sector is also unique in that its average level of capacity utilization over the sample period is statistically different from that of all other industries (see Table OA.2.2). The right Panel of Table OA.3.13 shows the results of sorting all non-mining industries into three portfolios on the basis of capacity utilization. Excluding mining industries and utilities from the sample does not change our baseline results. The utilization spread remains positive, yielding an average return of about 5.3% annually, and statistically significant at the 5% level.

Fourth, in our benchmark analysis we sort industries into portfolios based on the level of each industry's utilization rate. Here we modify this approach by sorting industries into portfolios based on the year-on-year *growth* rate, instead of the *level*, of utilization. Using the growth rate removes any (potential) differences in the average level of utilization across industries. The portfolio formation procedure follows that in Section 1.2, apart from the use of growth rates. The results are reported in Table OA.3.14 and show that the value-weighted (equal-weighted) utilization spread is 4.80% (5.74%) per annum and is significant at the 5% level. Portfolio returns are also monotonically decreasing in the utilization growth rate.

Table OA.3.14: Capacity utilization spread: sorting on growth rates

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

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

Lastly, we complement the empirical evidence above with a theoretical exercise in Section OA.4.5. We consider the implications of ex-ante parameter heterogeneity on the model-implied utilization premium. Parameter heterogeneity captures any cross-sectoral differences in depreciation, adjustment costs, or elasticity of depreciation to utilization. We show that such heterogeneities contribute only marginally to the utilization premium.

OA.3.8 Robustness: firm-level capacity utilization

For robustness, we construct a proxy for the unobservable capacity utilization rate at the *firm*-level. First, for each industry, we project the utilization rate of industry j at time t ($CU_{j,t}$) on salient industry-level production-related characteristics, contained in the vector $\mathbf{X}_{j,t}$. For the dependent variable, $CU_{j,t}$, we use either the raw industry utilization or industry-demeaned utilization rate. The latter approach ensures that the fitted value of this projection is not affected by fixed differences in average utilization across industries. The choice of $\mathbf{X}_{j,t}$ is motivated by the model in Section 2. We use the logarithms of size and book-to-market, the investment-to-capital ratio, IVOL, and TFP.³⁴ The regression is

$$CU_{j,t} = \beta_{j,0} + \boldsymbol{\beta}_j \mathbf{X}_{j,t} + \varepsilon_{j,t} \quad (\text{OA.3.7})$$

By estimating this projection separately for each industry, the relation between utilization and the characteristics, as measured by $\hat{\boldsymbol{\beta}}_j$, is specific to industry j . The average R^2 of this projection across industries is sizable at 33%, suggesting that the regressors well-span utilization at the industry level.

Second, the proxy for the utilization rate of a firm i that belongs to industry j at time t

³⁴Our empirical analysis shows that utilization affects the exposure of firms to aggregate productivity. Consequently, for $\mathbf{X}_{j,t}$ we choose firm-level variables that are known to also correlate with firms' exposure to this factor. In particular, Zhang (2005) shows that aggregate productivity exposure interacts with book-to-market and investment. Imrohoroglu and Tuzel (2014) establish the relation between firm-level TFP and aggregate productivity. Ai and Kiku (2016) document that idiosyncratic volatility serves as proxy for underlying growth options, whose riskiness depends on aggregate consumption. Importantly, in untabulated results we verify that our findings are robust to either removing particular characteristics (e.g., IVOL) or adding additional characteristics (e.g., hiring rates).

(denoted $\hat{CU}_{i,j,t}$) is obtained by combining the estimated slope coefficients for industry j , obtained via equation (OA.3.7), with the observable characteristics of firm i , denoted $\mathbf{X}_{i,j,t}$,

$$\hat{CU}_{i,j,t} = \hat{\beta}_{j,0} + \hat{\beta}_j \mathbf{X}_{i,j,t}. \quad (\text{OA.3.8})$$

This procedure allows the utilization proxy to vary between firms *within* the same industry. We use the firm-level utilization proxy to sort firms into portfolios as per Section 1.2, and report the results in Table OA.3.15.

Table OA.3.15: Capacity utilization spread: proxy for firm-level utilization rates

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

The table reports the annual value-weighted returns of portfolios sorted on the basis of estimated firm-level capacity utilization rates, as well as the spread between the low (L) and high (H) utilization portfolios. The table reports the average value-weighted return (Mean) and standard deviation (SD) of each portfolio's returns, and all portfolios are formed by following the procedure described in Section 1.2. t -statistics, reported in parentheses, are computed using Newey and West (1987) standard errors. The sample period is between July 1967 to December 2015.

The table shows that the firm-level utilization premium is about 5% per annum and statistically significant. Similar results are obtained using both raw or industry-demeaned utilization rates in projection (OA.3.7) (i.e., excluding industry fixed effects). The relation between firm-level utilization and average returns also remains monotonically decreasing in either case.

The findings above help to further illustrate that the benchmark utilization premium is not driven by ex-ante heterogeneity across industries.

OA.3.9 Supplemental tables

Table OA.3.16: Transition matrix of constituents between capacity utilization portfolios

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

The table shows the probability of an industry sorted into portfolio $i \in \{\text{Low, Medium, High}\}$ in year t , where i is the row index, being sorted into portfolio $j \in \{\text{Low, Medium, High}\}$ in year $t + 1$, where j is the column index. The transition probabilities are computed using annual capacity utilization data from June 1967 to December 2015. Industries are sorted into portfolios at the end of each June following the portfolio formation procedure described in Section 1.2.

Table OA.3.17: Capacity utilization spread: decile portfolios

Portfolio	Panel A: Portfolio returns		Panel B: Market betas	
	Mean	SD	β	$t(\beta)$
Low (L)	13.64	21.23	1.24	(11.74)
2	12.74	18.84	1.16	(14.51)
3	10.89	17.81	1.14	(14.00)
4	11.39	19.48	1.17	(17.34)
Medium	10.56	19.97	1.20	(11.78)
6	9.26	18.35	1.16	(14.51)
7	10.65	18.07	1.08	(12.71)
8	9.02	18.69	1.07	(11.39)
9	9.94	17.54	1.03	(13.57)
High (H)	7.96	20.22	1.05	(11.59)
Spread (L-H)	5.67 (2.31)	17.71	0.19	(2.96)

The table reports annual returns (Panel A) and CAPM betas (Panel B) of portfolios sorted on the basis of capacity utilization, as well as the spread between the returns and risk exposures of the low (L) and high (H) capacity utilization portfolios. The construction of the portfolios is identical to the benchmark analysis, except for the portfolio breakpoints used. Specifically, we sort the cross-section of industries into portfolios based on decile breakpoints. The mean in Panel A refers to the average annual return, obtained by multiplying the average monthly return by 12, and SD denotes the annualized standard deviation of monthly returns. Parentheses report Newey and West (1987) t -statistics. All portfolios are formed at the end of each June and are rebalanced annually. The sample is from July 1967 to December 2015. The risk exposures in Panel B correspond to β_1 from the regression $Ret_{i,t}^e = \beta_0 + \beta_1 MKTRF_t + \varepsilon_{i,t}$, where $Ret_{i,t}^e$ is the value-weighted excess return of a portfolio and $MKTRF_t$ is the excess market return, a proxy of aggregate productivity. Parentheses report Newey and West (1987) t -statistics. All portfolios are formed at the end of each June and are rebalanced annually. The sample is from July 1967 to December 2015.

Table OA.3.18: Fama-MacBeth regressions: industry-level

	Panel A: Annual		Panel B: Monthly	
	(1)	(2)	(3)	(4)
β_{CU}	-1.77 (-2.16)	-1.52 (-2.01)	-0.15 (-2.04)	-0.14 (-1.95)
Sector FE	-	Yes	-	Yes
Controls	Yes	Yes	Yes	Yes
R^2	0.475	0.540	0.459	0.533

The table reports the results of Fama-MacBeth regressions in which future industry-level excess returns are regressed on current utilization rates and industry-level controls. We run the following cross-sectional regression in which the dependent variable is either an industry's annual excess return from the start of July in year t to the end of June in year $t+1$ (Panel A), or an industry's monthly excess return from the end of month t to the end of month $t+1$ (Panel B). The independent variables are the utilization rate and a vector of the industry's characteristics (controls), \mathbf{X}_t measured at the end of June in year t (Panel A) or every month if available (Panel B): $R_{i,t \rightarrow t+1} = \beta_0 + \beta_u u_{i,t} + \beta'_t \mathbf{X}_{i,t} + \varepsilon_{i,t \rightarrow t+1}$. The controls \mathbf{X}_t include 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 $_{t-1}$). Here, all industry-level characteristics (other than CU) are calculated as the average characteristic across all firms assigned to a given industry at each point in time. After running these cross-sectional regressions we compute the time-series average of each element of the vectors the estimated slope coefficients. Parenthesis report Newey and West (1987) t -statistics. Columns 1 to 4 show the results when all characteristics are used in multivariate regressions, while Columns 2 and 4 also including sector fixed effects. Each control variable is standardized by dividing it by its unconditional standard deviation. The table also report the time-series average of the R^2 obtained from each set of cross-sectional regressions.

Table OA.3.19: Fama-MacBeth regressions: excluding small and microcap firms

	(1)	(2)	(3)
CU	-1.56 (-1.92)	-1.52 (-2.26)	-2.02 (-3.81)
TFP		1.27 (2.84)	1.20 (2.65)
HIRE		-4.20 (-2.72)	-4.15 (-2.84)
I/K		-2.30 (-2.17)	-2.22 (-2.36)
OVER		-3.62 (-2.98)	-3.83 (-3.32)
OC / AT		0.80 (0.65)	0.62 (0.65)
ln(ME)		-0.18 (-0.17)	-0.32 (-0.29)
ln(B/M)		3.71 (4.18)	3.69 (4.27)
RET_{t-1}		-0.34 (-0.22)	-0.29 (-0.21)
Sector FE	-	-	Yes
R^2	0.012	0.119	0.147

The table reports the results of Fama-MacBeth regressions in which future excess returns are regressed on current characteristics. In each year t we first identify firms in our sample that have a market capitalization rate less than the median market capitalization at that point in time, and remove these firms from the cross-section of firms (i.e., we restrict our focus to firms with large market capitalizations). We then run the following cross-sectional regression in which the dependent variable is a firm's annual excess return from the start of July in year t to the end of June in year $t + 1$, and the independent variables are a vector of the firm's characteristics, \mathbf{X}_t measured at the end of June in year t : $R_{i,t \rightarrow t+1} = \beta_0 + \beta_t' \mathbf{X}_{i,t} + \varepsilon_{i,t \rightarrow t+1} \quad \forall t \in \{1967, \dots, 2014\}$. The characteristics considered are capacity utilization (CU), total factor productivity (TFP), the hiring rate (HIRE), the natural investment rate (I/K), capacity overhang (OVER), the 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}). After running these cross-sectional regressions we compute the time-series average of each element of the vectors the estimated slope coefficients, $\{\hat{\beta}_t\}_{t=1967}^{2014}$. Each column reports the average slope coefficients for the characteristics of interest. Parenthesis report Newey and West (1987) t -statistics. Column 1 shows the results when the capacity utilization rate is the only predictor, while Columns 2 and 3 show the results when all characteristics are used in multivariate regressions. Column 3 also including sector fixed effects. Each control variable is standardized by dividing it by its unconditional standard deviation. The table also report the time-series average of the R^2 obtained from each set of cross-sectional regressions. The first regression is run in 1967 and the last regression is run in 2014, when the TFP data becomes unavailable.

Table OA.3.20: Distinction between the utilization premium and other spreads: projection evidence

Other spread	BE/ME	TFP	I/K	ME	Profit.	MOM	IVOL
Panel A: Empirical results							
α^{Data}	5.11 (2.04)	5.40 (2.20)	4.74 (1.99)	5.39 (2.21)	6.33 (2.63)	7.06 (2.84)	6.76 (2.72)
Panel B: Model-implied results							
α^{Model}	4.90	6.64	3.90	6.58	-	-	-

The table considers the distinction between the utilization premium and other spreads in both the data (Panel A) and the model (Panel B) through the lens of time-series regressions. Specifically, each panel reports the results of a univariate time-series regression in which returns associated with the utilization premium are projected on returns associated with another spread. The regression we estimate is given by $R_t^{CU} = \alpha + \beta R_t^{\text{Other}} + \varepsilon_t$, where R_t^{CU} represents the utilization premium in month t and R_t^{Other} represents the return associated with another spread in the same month. The other spreads we consider are: (i) the value premium (BE/ME), (ii) the productivity premium (TFP), (iii) the investment premium (I/K), (iv) the size effect (ME), (v) the profitability premium, (vi) the momentum effect, and (vii) the idiosyncratic volatility effect. The time-series regressions in Panel A are estimated using monthly data between July 1967 and December 2015, and then annualized. Parentheses report Newey and West (1987) t -statistics. The time-series regressions in Panel B are estimated using model-implied data from population-sample simulations of our extended models in Section OA.5, that feature two aggregate shocks (aggregate productivity and investment efficiency). Specifically, BE/ME, I/K and ME spreads are implied by the model in Section OA.5.1, while the TFP premium is implied by the model in Section OA.5.2. Panel B then reports the values of the model-implied alphas obtained by estimating the aforementioned regressions.

Table OA.3.21: Return predictability using utilization and productivity proxies

	Panel A: Industry-level			Panel B: Firm-level		
CU	-1.85 (-1.79)	-2.28 (-2.06)	-2.34 (-2.24)	-1.69 (-2.09)	-2.30 (-2.69)	-2.55 (-3.10)
ln (Sales / Capital)	-0.14 (-0.43)		-0.09 (-0.30)	0.50 (0.25)		0.21 (0.10)
ln (Sales / Employees)		0.02 (1.34)	0.02 (1.31)		2.82 (2.90)	3.19 (3.19)
R^2	0.126	0.104	0.170	0.027	0.030	0.043

The table reports the results of Fama-MacBeth regressions in which future industry-level (Panel A) or firm-level (Panel B) excess returns are regressed on current utilization rates and productivity proxies. We run the following cross-sectional regression in which the dependent variable is either an industry's annual excess return from the start of July in year t to the end of June in year $t+1$ (Panel A), or a firm's annual excess return from the start of July in year t to the end of June in year $t+1$ (Panel B): $R_{i,t \rightarrow t+1} = \beta_0 + \beta'_t \mathbf{X}_{i,t} + \varepsilon_{i,t \rightarrow t+1} \quad \forall t \in \{1967, \dots, 2014\}$. Here, the independent variables X_t are the utilization rate, the logarithmic ratio of sales to physical capital, and the logarithmic ratio of sales-to-employees, proxies for productivity. The industry-level characteristics (other than CU) in Panel A are calculated as the average characteristic across all firms assigned to a given industry at each point in time. After running these cross-sectional regressions we compute the time-series average of each element of the vectors the estimated slope coefficients. Parenthesis report Newey and West (1987) t -statistics. The table also report the time-series average of the R^2 obtained from each set of cross-sectional regressions. Each control variable is standardized by dividing it by its unconditional standard deviation.

OA.4 Additional theoretical results: Asset pricing

OA.4.1 Discussion of the model's assumptions

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

Fixed cost. The model features a real option to disinvest. While we could, in principle, also introduce a fixed cost for expanding capacity to the model (and make positive investment a real option also), we refrain from doing so to keep the dimensionality of the model's parameters low. Since the prior literature emphasizes that the adjustment costs of disinvestment are larger than those of investment (e.g., Zhang (2005)), our model captures this notion in a parsimonious manner. However, we relax this assumption and use a more comprehensive adjustment cost function in Section OA.6.4. We also provide an extensive discussion on the role of the disinvestment option under Section OA.4.3. There, we show that the non-convex adjustment cost breaks the link between utilization and investment margins.

Sources of risk. Motivated by the empirical evidence in Sections 1.4, our framework relies on exposures to a single priced state variable: productivity. Despite having only a single aggregate shock, the model can generate a distinct value premium and utilization premium, as shown in Section OA.4.2. The distinction arises from the real option nature of investment in the model. This real option to disinvest creates a substitution between a firm's utilization and investment policies, and also introduces additional degrees of non-linearity to model. However, we relax the assumption of a single aggregate shock in Section OA.5.1. Introducing an additional priced source of risk enhances the quantitative distinction between the utilization and value premia.

Wages and labor. Lastly, we demonstrate the importance of wages and flexible labor in the model. Our wage specification is shared with Jones and Tuzel (2013), Belo et al. (2014), and Imrohoroglu and Tuzel (2014), among many other studies. As in the aforementioned studies, our model does not feature a fixed cost that is paid unconditionally every period as in Zhang (2005) (thereby creating operating leverage). In contrast, our model relies on the labor margin which leads to a similar amplification of risk. To see this, consider equation (19). Utilization depends on aggregate productivity with the elasticity A_x that rises when $\omega < 1$. That is, the optimal choice of utilization (of all firms) is sensitive to changes in aggregate productivity though the dispersion in firms' choices of labor. This influences the risk exposures of firms' to aggregate productivity without explicitly introducing operating leverage to the model.

Table OA.4.1 show the model's sensitivity to both flexible labor and ω . Lower (higher) ω has a negligible effect on the utilization premium, but increases (decreases) the equity premium. For example, the equity premium changes by 0.80% per annum when we switch between $\omega = 0$ and $\omega = 1$. In contrast, the utilization premium only changes by about 0.2% per annum. This is in line with equation (19): lower ω increases the risk for all firms. In line with Imrohoroglu and Tuzel (2014), we solve a model with fixed labor (that is, with $L_{i,t} = 1$ for all firms). The resulting utilization premium is considerably smaller (about 1.35% per annum). Thus, flexible labor (or fixed costs of production) are required to create sufficient dispersion in risk exposures across firms.

Table OA.4.1: Model-implied moments across alternative calibrations for labor and wages

Row	Model	Time-series		Cross-sectional		Risk premia	
		$S_{TS}(ik)$	$\rho_1(ik)$	$\sigma_{CS}(ik)$	$S_{CS}(ik)$	$E[R^M]$	$E[R^{CU}]$
Data							
(1)		0.67	0.52	0.16	1.89	6.28	5.67
Baseline							
(2)		0.66	0.58	0.11	1.19	5.39	5.12
Different ω							
(3)	Low ($\omega = 0.00$)	0.71	0.59	0.11	1.19	5.49	5.07
(4)	High ($\omega = 1.00$)	0.57	0.56	0.07	0.10	4.70	5.28
Fixed labor							
(5)		-1.87	0.75	0.01	0.23	7.90	1.35

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 that perturb wages and labor. The table reports the time-series skewness ($S_{TS}(ik)$) and the first-order autocorrelation ($\rho_1(ik)$) of firm-level investment rates and the cross-sectional dispersion ($\sigma_{CS}(ik)$) and skewness ($S_{CS}(ik)$) of investment rates. In addition, the table also reports the equity premium ($(E[R^{CU}])$) and the capacity utilization premium ($E[R^{CU}]$) obtained by sorting the cross-section of model-implied returns association with each calibration on capacity utilization rates. 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 sensitivity of wages to aggregate productivity (ω) or fixing the quantity of labor available to each firm. All moments are based on a simulations of 1,000 firms over 40,000 periods (years). Finally, the top row of the table reports the empirical counterpart of each moment, while the second row of the table reports the value of each moment in our benchmark calibration and model.

OA.4.2 Model-implied distinction between utilization and book-to-market (or investment)

This section discusses the distinction between the utilization premium and the value premium (i.e., investment-related premia) within our benchmark model. To illustrate this point, we report the results of a conditional portfolio double sort of model-implied stock returns on book-to-market ratios and capacity utilization rates. The portfolio formation procedure follows the discussion in Section 3.1. The results of the analysis are reported in Table OA.4.2 and show that the utilization premium also exists *within* book-to-market portfolios.

Table OA.4.2 shows the model produces a sizable utilization premium *within* book-to-market portfolios. There are two reasons why our single-shock model is capable of simultaneously generating a spread along these two separate dimensions. First, despite the comovement between investment, utilization, and book-to-market in the model (all relate to Tobin's q), the correlation between the latter two margins is not perfect. Notably, our model features a real option (i.e., the fixed cost f) that induces “wait and see” periods of investment inaction. In these “wait and see” periods of investment inaction, utilization and investment do not comove as utilization *substitutes* the decision to exercise the costly real option to shed capital. Second, while both utilization and book-to-market are linked to the same aggregate shock, these relations to the aggregate shock are *non linear* because of time-varying betas. Both margins vary over time, and as they do, both

Table OA.4.2: Conditional double sort in the model

	Low CU	Medium CU	High CU	Spread (L-H)
Low B/M	6.42	4.96	3.48	2.95
Medium B/M	9.28	7.34	5.69	3.59
High B/M	10.02	8.99	7.60	2.42

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

margins change firms' conditional exposures to aggregate productivity shocks.

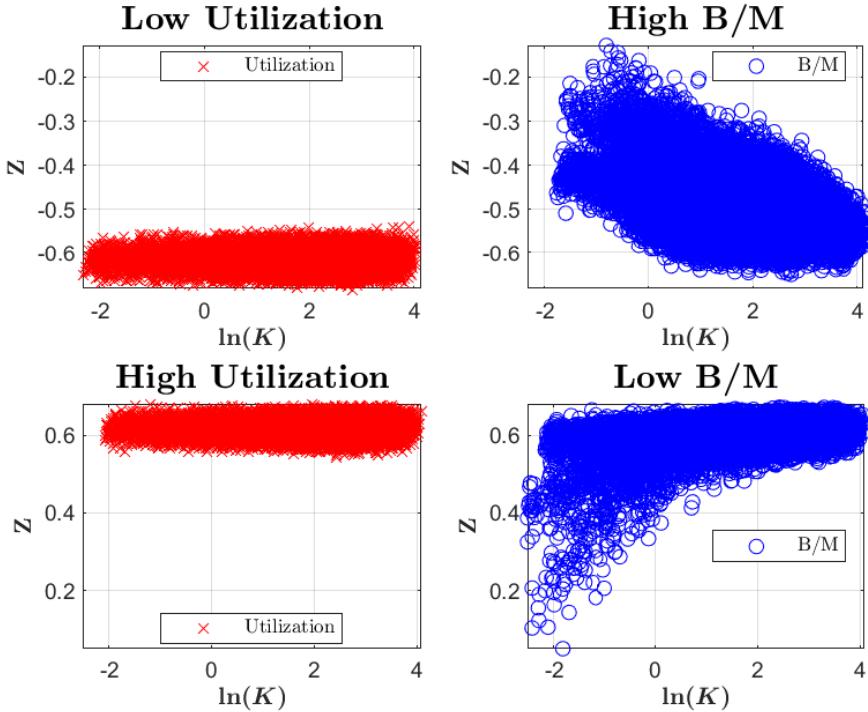
To further illustrate the distinction between low/high utilization firms and high/low B/M firms in our benchmark single-shock model, Figures OA.4.1 and OA.4.2 plot how firms sorted into portfolios according to their utilization rates and B/M ratios differ along two key dimensions: the natural logarithm of their capital stocks (represented by the x-axis and denoted by K) and the level of their idiosyncratic productivities (represented by the y-axis and denoted by Z). We focus on these two variables because they represent the two firm-level state variables in our model.

We produce these figures in three steps. First, we sort the cross-section of model-implied firms in each year t into portfolios based on either their utilization rates or their book-to-market ratios in year $t - 1$. When forming these portfolios we use the 10th and 90th percentiles of the cross-sectional distribution of each firm-level characteristic as portfolio breakpoints. This mimics the portfolio sort procedure we implement empirically. Next, we compute the mean value of both the capital stock and the idiosyncratic productivity of all firms assigned to each portfolio. Finally, we produce scatterplots and heat maps that visually describes the level of capital and idiosyncratic productivity for each key portfolio on the portfolio formation dates. Each plot is obtained from a model-implied population simulation that spans 40,000 periods and 1,000 firms.

The scatterplots in the top panel of Figure OA.4.1 display the properties of the low utilization portfolio and the high B/M portfolio. The figure shows that while both low utilization and high B/M firms tend to have large capital stocks and low idiosyncratic productivities, there are differences between the two portfolios in the *magnitude* of these two margins. In particular, low utilization firms are associated with low idiosyncratic productivity (Z) regardless of their capital stocks, whereas there is a substitution between Z and K for high book-to-market (i.e., value) firms. The idiosyncratic productivity of a value firm can be higher than that of a low utilization firm provided the value firm's capital stock is sufficiently small (i.e., low K). Similarly, the bottom panel of the figure shows the types of firms assigned to the high utilization and low B/M portfolios. Once again, high utilization firms are often not synonymous with low B/M firms.

Notably, a firm's exposure to aggregate productivity shocks is ultimately a function of the firm's underlying *state variables* Z and K . By projecting the cross-section of firms onto the (Z, K) plane, these plots help to illustrate why our single-shock model is able to generate both a utilization

Figure OA.4.1: Utilization and B/M sorted firms: Scatter plots of Z and $\ln(K)$

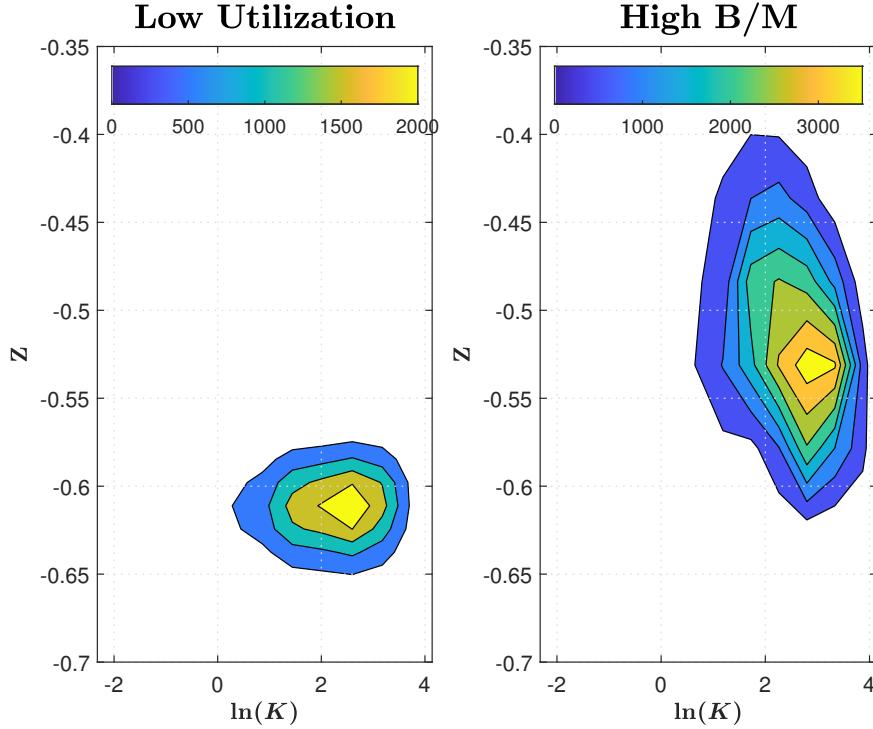


The figure reports scatterplots that display the average capital stock (represented by the x-axis) and the average level of idiosyncratic volatility (represented by the y-axis) of model-implied firms sorted into portfolios on the basis of their capacity utilization rates and book-to-market ratios. We produce these figures in two steps. First, we sort the cross-section of model-implied firms in each year t into portfolios based on either their utilization rates or their book-to-market ratios in year $t - 1$. When forming these portfolios we use the 10th and 90th percentiles of the cross-sectional distribution of each firm-level characteristic as portfolio breakpoints. This mimics the portfolio sort procedure we implement empirically. Second, we compute the mean value of both the capital stock and the idiosyncratic productivity of all firms assigned to each portfolio. Each plot is obtained from a model-implied population simulation that spans 40,000 periods and 1,000 firms.

premium and value premium. This is because the types of firms that are sorted into the low utilization portfolio are often quite different from the types of firms that are sorted into the high B/M portfolio. As such, these portfolios have visually different exposures to aggregate productivity risk and consequently earn different expected returns. This ultimately gives rise to relatively distinct utilization and investment premia within our framework.

While the former scatterplots show a number of distinct differences between the firms assigned to the various utilization- and B/M-sorted portfolios, the large number of overlapping points on these diagrams can mask the underlying *distributions* of firms in the economy. To address this point, Figure OA.4.2 plots heat maps that display the types of firms assigned to the key portfolios, along with associated the frequencies of portfolio membership as a function of Z and K . The contours of the low (high) utilization (B/M) portfolios are almost non-overlapping when we account for the frequency of portfolio membership. This once again highlights how a value firm in our single-shock model is not necessarily synonymous with a low utilization firm.

Figure OA.4.2: Low utilization versus value firms: heat maps of Z and $\ln(K)$



The figure reports heat maps that display the average capital stock (represented by the x-axis) and the average level of idiosyncratic volatility (represented by the y-axis) of model-implied firms sorted into portfolios on the basis of their capacity utilization rates and book-to-market ratios. We produce these figures in two steps. First, we sort the cross-section of model-implied firms in each year t into portfolios based on either their utilization rates or their book-to-market ratios in year $t - 1$. When forming these portfolios we use the 10th and 90th percentiles of the cross-sectional distribution of each firm-level characteristic as portfolio breakpoints. This mimics the portfolio sort procedure we implement empirically. Second, we compute the mean value of both the capital stock and the idiosyncratic productivity of all firms assigned to each portfolio. Each plot is obtained from a model-implied population simulation that spans 40,000 periods and 1,000 firms.

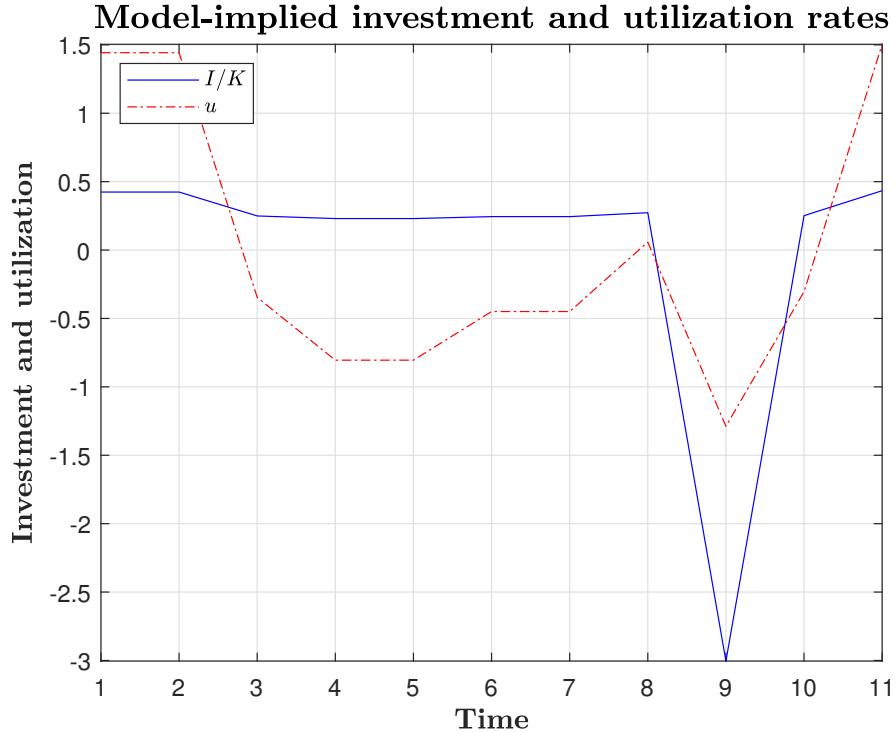
OA.4.3 The role of non-convex disinvestment costs

When the fixed cost of disinvestment is positive ($f > 0$), the decision to shed capital becomes a real option. This means that firms may not necessarily want to sell their capital immediately when a negative shock hits the economy. This is because there are multiple periods of investment inaction in which firms “want and see” if productivity will recover before committing to the partially irreversible decision to shed physical capital. It is these periods of investment inactions that show the role of variable capacity utilization as serving as a *substitute* for selling capital.

Consequently, the fixed cost of disinvestment creates economically important distinctions between capital *utilization* and capital *investment*. Namely, a positive disinvestment cost implies that (i) utilization leads investment during economic downturns, an empirical observation that is commonly used by forecasters, and (ii) the utilization premium features distinction from investment-based spreads, even in a single-shock model. We discuss each role of $f > 0$ in turn.

Utilization as a leading economic indicator. Consider Figure OA.4.3 that displays the (normalized) paths of physical investment and capacity utilization for a model-simulated firm

Figure OA.4.3: Substitution between investment and utilization



The figure displays the (normalized) paths of physical investment and capacity utilization for a model-simulated firm around an economic downturn.

around an economic downturn. While this representative firm *slightly* reduces its physical investment rate between periods two and three, its physical investment rate is unchanged between periods three and eight. This is because the firm is stuck in a “wait and see” period of investment inaction while the firm decides whether to exercise its costly real option to shed capital. When the firm ultimately commits to shed physical capital in period nine, its investment rate drops sharply before *slightly* increasing in period 11. Overall, the firm’s investment policy is characterized by a lengthy period of investment inaction followed by a “lumpy” decision to disinvest capital.

In contrast, the firm’s choice of utilization drops substantially from period two to period six and therefore *predicts* the sharp disinvestment of physical capital that is only observed in period nine. Moreover, the firm’s utilization rate continues to vary over time even while its physical investment rate is fixed between periods three and eight. From periods two to eight, the lower utilization rate *substitutes* the firm’s investment decision for the purpose of smoothing its dividends. That is, the lower utilization rate implies a lower depreciation that, all else equal, raises the firm’s dividend. This lower depreciation rate has the same impact on the firm’s dividends as selling machines.

The fact that capacity utilization *leads* investment is a prominent empirical regularity. For instance, the Federal Reserve Board often cites the capacity utilization rate as serving as a useful predictor of output (Koenig et al., 1996) and other business cycle fluctuations (Corrado and Mattey, 1997). This leading behavior would be difficult to observe in a model *without* non-convex adjustment costs and real options. Moreover, without the substitution between investment and utilization that is induced by the real option to disinvest, the firm’s investment-related decisions would be too

Table OA.4.3: Utilization premium within B/M-sorted portfolios: the role of f

Panel A: $f > 0$					
	Low CU	Medium CU	High CU	Spread (L-H)	
Low B/M	6.42	4.96	3.48	2.95	
Medium B/M	9.28	7.34	5.69	3.59	
High B/M	10.02	8.99	7.60	2.42	
Panel B: $f = 0$					
	Low CU	Medium CU	High CU	Spread (L-H)	% change
Low B/M	6.97	6.35	5.69	1.28	56.61%
Medium B/M	8.17	7.30	6.83	1.33	62.95%
High B/M	8.95	8.24	7.89	1.06	56.20%

The table reports the results of conditional portfolio double sorts that construct the utilization premium *within* book-to-market sorted portfolios using our benchmark model from Section 2 with the fixed cost of disinvestment (i.e., $f > 0$) in Panel A and without the fixed cost of disinvestment (i.e., $f = 0$) in Panel B. The portfolio sorts in each panel are conducted as follows. First, we simulate a cross-section of 1,000 model-implied firms for 40,000 time periods. Then, we then conduct a portfolio double sort in which we first sort the cross-section of firms on the basis of book-to-market ratios. That is, on each date t we sort the cross-section of firms into one of three portfolios on the basis of their book-to-market ratios on date $t - 1$. We assign firms to these portfolios using the 20th and 80th percentiles of the cross-sectional distribution of book-to-market ratios. Next, *within* each of these three book-to-market-sorted portfolios, we further sort the cross-section of firms into one of three portfolios based on the cross-sectional distribution of capacity utilization rates on date $t - 1$. Here, we use the same portfolio breakpoints as in the previous step. This produces a series of nine portfolios that we hold for one period, at which point in time all portfolios are rebalanced. The table then reports the average return of each portfolio, as well as the spread in the returns between the low and high utilization-sorted portfolios. Moreover, in Panel B, we report the percentage change between the conditional utilization premium obtained in Panel A, from the economy in which $f > 0$, to the conditional utilization premium obtained in Panel B, from the economy in which $f = 0$.

highly correlated with utilization rates (counterfactually so).

Distinction from investment-based spreads. The real option to disinvest also introduces a large degree of non-linearity to the production-based model in Section 2. In turn, this non-linear relation between the fundamental aggregate productivity shock and both investment and capacity utilization decisions allows the model to produce book-to-market and utilization spreads that are relatively distinct despite the model only featuring a single source of risk: aggregate productivity. Section OA.4.2 further elaborates on the economic mechanism that separates the utilization premium and the value premium within the baseline model.

To illustrate the importance of $f > 0$ for this between the utilization premium and investment premia, we start by simulating a cross-section of 1,000 model-implied firms for 40,000 time periods in an economy that features a positive fixed cost of disinvestment. We then conduct a portfolio double sort in which we first sort the cross-section of firms on the basis of book-to-market ratios. That is, on each date t we sort the cross-section of firms into one of three portfolios on the basis of their book-to-market ratios on date $t - 1$. We assign firms to these portfolios using the 20th and 80th percentiles of the cross-sectional distribution of book-to-market ratios. Next, *within* each of these three book-to-market-sorted portfolios, we further sort the cross-section of firms into one of three portfolios based on the cross-sectional distribution of capacity utilization rates on date $t - 1$. We once again use the same portfolio breakpoints as the previous step. This produces a series of nine portfolios that we hold for one period, at which point in time all portfolios are rebalanced.

Panel A of Table OA.4.3 reports the average annual portfolio returns associated with this

Table OA.4.4: The role of flexible utilization for key moments: the case of $f = 0$

Row	Model	Time-series		Cross-sectional		Risk premia	
		$S_{TS}(ik)$	$\rho_1(ik)$	$\sigma_{CS}(ik)$	$S_{CS}(ik)$	$E[R^{CU}]$	$E[R^{bm}]$
	Data						
(1)	Baseline	0.67	0.52	0.16	1.89	5.67	3.71
(2)	Baseline without fixed cost	0.66	0.58	0.11	1.19	5.12	3.76
(3)	Baseline without utilization	1.04	0.63	0.11	1.32	4.53	3.63
(4)	Baseline without utilization and fixed cost	-0.09	0.64	0.07	0.11	-	2.93
(5)		0.66	0.71	0.08	0.59	-	2.67

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 of the model featuring no fixed disinvestment cost ($f = 0$). The table reports the time-series skewness ($S_{TS}(ik)$), and the first-order autocorrelation ($\rho_1(ik)$) of firm-level investment rates, as well as the cross-sectional dispersion ($\sigma_{CS}(ik)$) and skewness ($S_{CS}(ik)$) of investment rates. In addition, the table also reports the value premium ($E[R^{bm}]$) and investment premium ($E[R^{ik}]$) 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 $f = 0$ in rows (3) to (5), and $\lambda \rightarrow \infty$ in rows (4) and (5). All moments are based on a simulations of 1,000 firms over 40,000 periods (years). Finally, the top row of the table also reports the empirical counterpart of each moment.

analysis. The panel indicates that even though our baseline model only features a *single* priced source of risk, the model is capable of generating a utilization premium that is largely distinct from the value premium. Notably, the utilization premium ranges from about 2.5% to 3.5% per annum even after we condition on book-to-market ratios. While smaller than the unconditional spread, this conditional spread falls within the empirical confidence interval. When we set $f = 0$ and repeat this analysis, Panel B shows that the relative magnitude of the utilization premium falls by between 55% and 62% compared to Panel A. Moreover, a utilization premium of about 1.2% is below the lower bound of the empirical confidence interval. The stark difference between Panels A and B highlights the key role of non-convex disinvestment costs for separating the utilization premium from the value premium in the context of our benchmark model.

Sensitivity to fixed disinvestment costs. Table OA.4.4 reports a sensitivity analysis to show how the key investment-related moments and risk premia change when we switch both the fixed cost of disinvestment and variable capacity utilization on and off. There are two key takeaways.

First, row (3) shows the unconditional level of the utilization premium remains sizable even when $f = 0$. That is, without the fixed cost of disinvestment, the utilization premium falls by 0.59% (from 5.12% in row (2) to 4.53% in row (3)). This represents as 12% decline in the relative magnitude of the utilization premium compared to the baseline model. While this decline in the magnitude of the spread is quite sizable, the model-implied utilization spread in row (3) still falls within the empirical confidence interval of the spread in the data.

Second, an economy with flexible utilization can improve the model's fit to the data compared to an economy with fixed utilization, even if there is no fixed cost of disinvestment and $f = 0$. Comparing row (3) to row (5) shows that setting $f = 0$ but letting utilization vary helps the model to generate a sizable value premium, holding all else constant. The value premium in row (5) is 0.96% smaller than the value premium in row (3). This indicates that fixing utilization induces the

value premium to fall, in relative terms, by 26% compared to the economy underlying row (3). Row (5) also shows that the cross-sectional skewness of investment is more than halved when utilization is fixed, and investment becomes too autocorrelated compared to the data.

OA.4.4 Sensitivity analysis of the model for risk premia

Below, we numerically illustrate the intuition for the utilization spread discussed in Section 3.2. We show the sensitivity of the spread to ingredients (1)–(3) of our model (the quadratic capital adjustment cost, fixed cost of disinvestment, and countercyclical market price of risk, respectively). We also show that the utilization spread is largely unaffected by perturbing the parameters governing the evolution of aggregate or idiosyncratic productivity. Table OA.4.5 presents these results, and reports the mean value-weighted of the utilization spread. The table also reports the mean and volatility of the equity risk premium in the model under each alternative calibration.

The results in rows (2) and (3) show that when the extent of the first friction, the quadratic capital adjustment costs, is perturbed, the magnitudes of the utilization spread changes but β is largely unaffected. As this friction is increased in row (3), the magnitude of the utilization spread increases. With higher adjustment costs, firms can less readily alter the level of their capital stocks, and low utilization implies more underlying capital risk.

Row (4) considers an economy in which the second ingredient, the fixed cost of capital disinvestment, is removed but the remaining two frictions are held constant. The utilization spread still exists, although its magnitude is decreased by almost 1% per annum. The decrease in the utilization spread reflects how removing the fixed cost of disinvestment better allows firms to shed their capital stock instead of substituting disinvestment with temporary declines in utilization. However, the fact that the utilization spread remains sizable indicates that firms still cannot fully absorb productivity shocks into their capital stock. Next, rows (5), (6), and (7) consider the role of the third ingredient, the countercyclical market price of risk. In particular, row (6) illustrates how a more countercyclical market price of risk translates into a higher equity risk premium and volatility of aggregate market returns, as well as an increased utilization spread. This occurs because the asymmetry between good and bad aggregate productivity is widened. Row (7) indicates that both the equity risk premium and capacity utilization spread are severely diminished with an acyclical market price of risk.

Rows (8) and (9) show how the utilization spread changes as the persistence of aggregate productivity changes. The results in row (8) show that when aggregate productivity is less persistent, the magnitude and the volatility of the equity risk premium decrease. Similarly, the mean utilization premium falls slightly. The opposite patterns emerge in row (9), when the persistence of aggregate productivity increases.

Rows (10) and (11) of the table display how the utilization premium and equity risk premium both fall (rise) when aggregate productivity becomes less (more) volatile. The same patterns hold true for the volatility of the equity risk premium. Importantly, in rows (8)–(11), the model implied utilization premium changes by at most 0.6% in absolute value compared to the benchmark case, and falls within the empirical 95% confidence interval.

Finally, rows (12) to (15) display the sensitivity of key asset-pricing moments to perturbations in the parameters governing the dynamics of idiosyncratic productivity. The results indicate that when the persistence (ρ_z) or the volatility (σ_z) of idiosyncratic productivity increases, the capacity

Table OA.4.5: Model-implied capacity utilization spread across alternative calibrations of the model

Row	Model	$E[R^M]$	$\sigma(R^M)$	$E[R^{CU}]$
	Baseline			
(1)		5.39	20.89	5.12
	Different ϕ			
(2)	Low ($\phi = 1.40$)	5.39	20.85	5.06
(3)	High ($\phi = 1.60$)	5.39	20.94	5.18
	No fixed cost			
(4)		5.72	20.39	4.53
	Different γ_1			
(5)	Low ($\gamma_1 = -8.60$)	5.26	20.48	5.06
(6)	High ($\gamma_1 = -9.00$)	5.52	21.31	5.18
(7)	Acyclical ($\gamma_1 = 0$)	1.88	9.94	2.68
	Different ρ_x			
(8)	Low ($\rho_x = 0.899$)	4.51	17.48	4.69
(9)	High ($\rho_x = 0.945$)	6.59	25.54	5.55
	Different σ_x			
(10)	Low ($\sigma_x = 0.0137$)	4.96	19.77	4.91
(11)	High ($\sigma_x = 0.0143$)	5.84	22.03	5.34
	Different ρ_z			
(12)	Low ($\rho_z = 0.585$)	5.58	20.90	4.88
(13)	High ($\rho_z = 0.615$)	5.19	20.89	5.38
	Different σ_z			
(14)	Low ($\sigma_z = 0.2925$)	5.54	20.89	4.87
(15)	High ($\sigma_z = 0.3075$)	5.23	20.85	5.37

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

utilization premium rises.

OA.4.5 Sectoral heterogeneity and the utilization premium

We design a simulation-based experiment to put an upper bound on how much ex-ante sectoral-level differences (e.g., difference in adjustment costs and the flexibility of utilization across different industries) may affect the utilization spread. Let x denote a parameter of interest, let x_0 denote the value of x in our benchmark calibration, and let N represent the number of firms in the economy. First, we solve the model twice: once when x is doubled to $x_U = 2x_0$, and once when x is halved to $x_D = 0.5x_0$. Next, we simulate $N/2$ firms implied by the model solved for each x_U and x_D , and combine these two simulations into one economy of N firms. These two simulations aim to capture *extreme* differences between industries in terms of parameter x . Finally, we sort the N firms on

the basis of utilization in an identical fashion to our baseline model results. The difference between the model-implied utilization premium here versus our benchmark quantifies an upper bound on the effect that ex-ante heterogeneity in parameter x has on the premium.

Table OA.4.6: Capacity utilization spread: sensitivity to ex-ante heterogeneity

Portfolio	Hetero. in ϕ		Hetero. in λ		Hetero. in δ_k		Hetero. in f	
	$E[R^{CU}]$	β	$E[R^{CU}]$	β	$E[R^{CU}]$	β	$E[R^{CU}]$	β
Low (L)	9.82	1.21	9.60	1.19	9.71	1.26	9.81	1.21
Medium	6.99	1.10	6.99	1.09	6.15	1.10	7.01	1.12
High (H)	4.52	0.97	4.23	1.01	3.60	0.97	4.50	1.01
Spread (L-H)	5.30	0.23	5.37	0.18	6.11	0.29	5.31	0.21

The Table show the model-implied capacity utilization spread when firms in the economy show ex-ante heterogeneity in some parameter of interest x . The parameter of interest x is either ϕ , the quadratic capital adjustment costs, λ , the elasticity of depreciation to utilization, δ_k , the average depreciation rate, or f , the fixed cost of disinvestment. For each parameter, we follow the simulation procedure described in Section OA.4.5 and set N to 1,000. Each simulation encompasses 40,000 periods, and we sort all firms on the basis of utilization in an identical fashion to Table 8. A firm is sorted into the high (low) utilization portfolio if its level of capacity utilization is above (below) the 90th (10th) percentile of the cross-sectional distribution of capacity utilization rates in the previous period. 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.

We consider heterogeneity in four parameters: the depreciation rate, δ_k , elasticity of depreciation to utilization, λ , convex adjustment cost, ϕ , and fixed disinvestment cost, f . The results are reported in Table OA.4.6. In comparison to Panel B of Table 8, heterogeneity in ϕ , λ , δ_k and f add up to 0.06%, 0.13%, 0.87% and 0.20% per annum, respectively, to the premium. Thus, ex-ante sectoral heterogeneity only implies a marginal amplification of the model-implied utilization spread.

OA.4.6 Numerical model solution

We solve the model numerically using value function iteration. The value function and the optimal policies implied by the firm's maximization problem in equation (12) are solved on a grid in a discrete state space. The grid for capital stock, K , features 501 grid points, with the endpoints of the grid chosen to be nonbinding. The aggregate productivity process, x , and the idiosyncratic productivity process, z , are each driven by an independent and identically distributed normal distribution. While each of these state variables has continuous support in the model, each variable needs to be transformed into a finite number of states to implement the numerical solution algorithm. We use the method of Tauchen and Hussey (1991) to discretized the z process into 11 states. Because the method of Tauchen and Hussey (1991) does not work well for persistent processes, namely those with a persistence parameter greater than 0.90, we use the method of Rouwenhorst (1995) to discretize x into 5 states. Once the discrete state space has been constructed, conditional expectations are computed using matrix multiplication and the firm's maximization problem is solved using a global search routine. All results are robust to choosing finer grids.

OA.5 Model extensions

OA.5.1 A model with investment efficiency shocks

We extend our baseline single aggregate shock model from Section 2 to a model with two aggregate shocks in order to generate distinct economic mechanisms that drive the utilization premium and other investment-based spreads. Specifically, we introduce an aggregate “investment efficiency” shock to our model that captures the spirit of the capital-embodied shocks in Papanikolaou (2011), Garlappi and Song (2017), and many other studies.

Introducing these investment efficiency shocks to the model leads to periods in which the marginal efficiency of investment is higher, meaning that investment can be transformed into capital in a more efficient manner as in Papanikolaou (2011) and Justiniano et al. (2011). This increased investment efficiency can arise as a result of an improvement in the quality of investment goods or through lower frictions to the process through which investment goods are transformed into productive capital. Justiniano et al. (2011) argue that these investment frictions can also be related to the efficiency of the financial sector (i.e., a reduced-form financial accelerator as in Carlstrom and Fuerst (1997)).

Consequently, investment efficiency shocks reduce the installation cost of new vintages of capital (i.e., lower adjustment costs). This means that investment efficiency shocks will affect firms’ investment policies but will not affect their choices of utilization rates. Thus, introducing these investment efficiency shocks creates a separation between the investment-related spreads (e.g., the value premium), which are driven by these investment efficiency shocks, and the utilization premium, which remains primarily driven by aggregate productivity shocks.

Model setup. We make two main changes to the setup of our baseline model described in Section 2.1. First, we alter the convex capital adjustment cost function to

$$\frac{\phi - s_t}{2} \left(\frac{I_{i,t}}{K_{i,t}} - \delta(u_{i,t}) \right)^2 K_{i,t}, \quad (\text{OA.5.1})$$

where s_t is the degree of aggregate investment efficiency. Note that a positive s_t shock decreases the amount of adjustment costs. We assume that $\phi \gg \sigma_s$, such that $\phi - s_t$ is always positive. Alternatively, we could posit that adjustment costs are proportional to $\phi/\exp(s_t)$ (see, e.g., Belo et al. (2014)). However, we opt for the former linear specification to avoid any asymmetric effects between positive versus negative s_t shocks due to convexity. The process for s_t follows

$$s_{t+1} = \rho_s s_t + \varepsilon_{t+1}^s, \text{ where } \varepsilon_{t+1}^s \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_s^2). \quad (\text{OA.5.2})$$

Second, we also alter the SDF to feature this priced “investment efficiency” shock

$$\ln(M_{t+1}) = \ln(\beta) - \gamma_t \varepsilon_{t+1}^x - \text{SIGN}_S \cdot \gamma_s \varepsilon_{t+1}^s - \frac{1}{2} \gamma_t^2 \sigma_x^2 - \frac{1}{2} \text{SIGN}_S^2 \gamma_s^2 \sigma_s^2. \quad (\text{OA.5.3})$$

Here, γ_t follows identical dynamics to those described in Section 2.1 (recall equation (11)), γ_s is the absolute value of the market price of investment efficiency risk, and $\text{SIGN}_S \in \{-1, +1\}$ determines whether investment efficiency shocks carry a negative or a positive market price of risk.

Optimality conditions and model intuition. To see why the utilization and investment-related spreads are distinct in this two-shock model, it is useful to examine the first-order conditions associated with a firm’s choice of investment and utilization. The first-order condition for invest-

ment is given by the Euler equation

$$q_t = \mathbb{E}_t \left[M_{t,t+1} \left(\frac{\partial Y_{i,t+1}}{\partial K_{i,t+1}} - \frac{I_{i,t+1}}{K_{i,t+1}} - \frac{\phi - s_{t+1}}{2} \left(\frac{I_{i,t+1}}{K_{i,t+1}} - \delta_{i,t} \right)^2 + q_{t+1} \left(1 - \delta_{i,t+1} + \frac{I_{i,t+1}}{K_{i,t+1}} \right) \right) \right], \quad (\text{OA.5.4})$$

where

$$q_t = 1 + (\phi - s_t) \left[\frac{I_{i,t}}{K_{i,t}} - \delta_{i,t} \right]. \quad (\text{OA.5.5})$$

The first-order condition for utilization is

$$0 = \frac{\partial Y_{i,t}}{\partial u_{i,t}} + (\phi - s_t) \left[\frac{I_{i,t}}{K_{i,t}} - \delta_{i,t} \right] \frac{\partial \delta(u_{i,t})}{\partial u_{i,t}} - q_t \frac{\partial \delta(u_{i,t})}{\partial u_{i,t}} K_{i,t}, \quad (\text{OA.5.6})$$

which can be simplified to

$$\frac{\partial Y_{i,t}}{\partial u_{i,t}} = \frac{\partial \delta(u_{i,t})}{\partial u_{i,t}} K_{i,t}. \quad (\text{OA.5.7})$$

Thus, equation (OA.5.7) shows that the firm's choice of utilization is independent of investment efficiency as both $\frac{\partial Y_{i,t}}{\partial u_{i,t}}$ and $\frac{\partial \delta(u_{i,t})}{\partial u_{i,t}}$ are independent of s_t and K_t is predetermined.

In contrast, equation (OA.5.4) highlights that the firm's choice of investment critically and directly depends on the degree of investment efficiency s_{t+1} , and indirectly depends on s_t since $\rho_s > 0$ (i.e., shocks to investment efficiency are persistent). When investment efficiency *rises* (that is, when s_t increases), the cost of installing new capital *falls*. Thus, a positive investment efficiency shock can be interpreted as a technological improvement that induces the firm to invest (disinvest) more than the firm otherwise would when subject to a positive (negative) productivity shock.

Because the choice of utilization does not depend on shocks to s_t , the expected returns of the utilization-sorted portfolios remain tightly related to their exposures to aggregate productivity shocks. Thus, the utilization premium in this two-shock model will arise for the same reasons that we obtain a utilization premium in our single-shock model (see Section 3.1). In contrast, because a firm's investment policy depends on the degree of investment efficiency s_t , the expected returns of portfolios sorted on the basis of firms' investment-related policies (e.g., the value premium) will depend more heavily on the degree to which these portfolios are exposed to the investment efficiency shocks. To see this, we can fix a firm's investment policy and write the exposure of the ex-dividend firm value to s_t shocks as

$$\beta_{i,s}^{\text{ex-div}} = \frac{\partial q_{i,t}}{\partial s_t} = -\frac{I_{i,t}}{K_{i,t}}. \quad (\text{OA.5.8})$$

Equation (OA.5.8) show that a firm with a positive investment rate that is expanding its capital stock has a *negative* exposure to the investment efficiency shock s_t . Intuitively, a firm that is expanding its capital stock has a relatively larger growth option value in the presence of capital adjustment costs because shocks are absorbed in firm value rather than investment. Consequently, a positive shock to s_t lowers capital adjustment costs (recall equation OA.5.1) and, in turn, induces a negative valuation effect by decreasing the growth option value of the firm (in fact, there is no growth option value in the limit where $\phi = s_t$ and $q_{i,t} = 1$). The opposite logic applies to a firm that is disinvesting, whose exposure to s_t shocks is negative. Moreover, a positive s_t shock makes an investing firm have a higher I/K and a disinvesting firm have a lower I/K , all else equal. This interaction effect enhances the exposure to investment efficiency shocks (in absolute value).

As growth (value) firms invest (disinvest) more heavily, they have a negative (positive) exposure to the investment efficiency shocks. This creates a strong dependence of the value premium (and

other investment-related spreads) on shocks to s_t .

Calibrating the two-shock model. Adding the “investment efficiency” shock introduces three new parameters to the model. In the spirit of Li (2018), who proposes a unified explanation for the value premium and the moment effect, we set the volatility of these investment efficiency shocks to 7.25% per annum and the autocorrelation of these shocks to 0.90. We also set $SIGN_s$ to +1, indicating that investment efficiency shocks carry a positive market price of risk. A positive market price of risk for investment efficiency shocks is not only consistent with Li (2018), but also mirrors an important result from Garlappi and Song (2017). Namely, investment-specific technology (IST) shocks carry a *negative* price of risk if utilization is fixed, but a *positive* price of risk if utilization is sufficiently flexible.³⁵ The latter case fits our framework with time-varying utilization. We calibrate the magnitude of the market price of this investment efficiency shock (γ_s) to match the book-to-market spread in the data. Since our primary objective are to (i) establish an economically sizable utilization premium, and (ii) establish that the utilization premium is independent from the value premium, we emphasize that neither this choice nor the choice of any other calibration parameter targets the utilization spread directly. Thus, the utilization premium remains a testable prediction of our two-shock model.

In the two-shock model we opt to set the fixed cost of disinvestment (f) to zero. Recall that in our single-shock model, part of the difference between a firm’s choice of utilization and investment arises because of this fixed cost of disinvestment. This cost induces periods in which firms are “waiting to see” if aggregate productivity improves before exercising their costly options to sell capital. In the two-shock framework, the investment efficiency shock introduces a sufficient degree of heterogeneity between firms’ investment and utilization policies. By setting the fixed cost of disinvestment to zero, we narrow the economic distinction between the utilization and the value premia to (i) hinge on the investment efficiency shocks (rather than on real option nature of investment), and (ii) limit the degrees of freedom in the model. Because the model no longer features a real option, the model is smooth and differentiable with respect to the state and control variables. We therefore solve the model using a fourth order perturbation method in Dynare++.

We report the remaining parameters used to calibrate the two-shock model in Table OA.5.1. These remaining parameters are either identical to, or in the close vicinity of, the parameters used in our baseline one-shock model that are reported in Table 6.

Table OA.5.2 reports the key moments implied by the two-shock model when the investment efficiency shocks carry either a positive or a negative market price of risk. We compare these moments to the same moments implied by our baseline one-shock model and the data. The table shows that the two-shock model provides a close match to the data along many dimensions related to firm-level investment rates, aggregate utilization rates, aggregate sales growth rates, and aggregate asset-pricing data. In particular, and compared to the single-shock model, the two-shock model yields an improvement in the model’s fit for the second-order autocorrelation and the inter-decile

³⁵Garlappi and Song (2017) show that a positive IST shock increases the relative productivity of the investment sector compared to the consumption sector, and hence diverts labor away from consumption firms and into investment firms. This will, on the one hand, drop the supply of labor in the consumption sector and reduce the amount of the consumption good that is available, increasing the marginal utility of households. On the other hand, if utilization is flexible, then the consumption sector can offset the decline in labor by increasing the utilization of capital (which becomes relatively cheaper), thereby producing more of the consumption good. This decreases the marginal utility of the household. If capital utilization is sufficiently flexible, the second effect more than offsets the negative effect of labor diverting to the investment sector and results in a positive price of risk.

Table OA.5.1: A model with aggregate investment efficiency shocks: calibration

Symbol	Description	Value
Stochastic processes		
ρ_x	Persistence of aggregate productivity	0.922
σ_x	Conditional volatility of aggregate productivity	0.014
ρ_z	Persistence of idiosyncratic productivity	0.700
σ_z	Conditional volatility of idiosyncratic productivity	0.150
ρ_s	Persistence of capital efficiency	0.900
σ_s	Conditional volatility of capital efficiency	0.073
β	Time discount factor	0.988
γ_0	Constant price of aggregate productivity risk	2.175
γ_1	Time-varying price of aggregate productivity risk	-89.250
γ_s	Time-varying price of capital efficiency risk	2.000
Technology		
α_k	Capital share	0.333
α_l	Labor share	0.667
θ	Returns to scale of production	0.900
δ_k	Fixed capital depreciation rate	0.080
λ	Elasticity of marginal depreciation	2.350
ω	Sensitivity of wages to aggregate productivity	0.200
ϕ	Adjustment cost parameter	0.653

The table reports the model parameters for the model extension that features two aggregate priced sources of risk: aggregate productivity shocks and investment efficiency shocks.

range of investment, and the volatility of market returns. Most moments are largely identical when $SIGN_s = 1$ compared to when $SIGN_s = -1$, although the equity premium is about 0.2% per annum larger in the former case (which is our benchmark calibration of this two-shock model).

Model-implied spreads. As a first step, we check whether the model produces both a sizable utilization premium and a sizable value premium. Using model-simulated data, we sort the cross-section of firms into three univariate portfolios on the basis of either capacity utilization rates or book-to-market ratios. Here, our portfolio formation procedure follows the benchmark approach used in Section 1.2 of the main text. That is, in each period t , we sort firms into portfolios on the basis of the cross-sectional distribution of the characteristic of interest in period $t - 1$. We then hold each of these portfolios for one period, at which point in time all portfolios are rebalanced.

Table OA.5.3 reports the results of these univariate portfolio sorts in the two-shock model, and shows that the model produces both a sizable utilization premium and a sizable book-to-market spread. The utilization (book-to-market) spread is 6.40% (3.82%) per annum when $SIGN_s = 1$, while the utilization (book-to-market) spread is 6.52% (3.05%) per annum when $SIGN_s = -1$.

Two differences between the case of $SIGN_s = 1$ (our benchmark) and the case of $SIGN_s = -1$ are worth noting. First, the value premium is sensitive to the price of risk of the investment efficiency shocks. If these shocks carry a positive price of risk ($SIGN_s = 1$), then the value premium matches the data. In contrast, if these shocks carry a negative price of risk ($SIGN_s = -1$), then the premium shrinks by 20% (in relative terms) from 3.82% per annum to 3.05% per annum. This finding is consistent with the model's optimality condition for investment in equation (OA.5.4) and the ensuing discussion that shows that growth (value) firms have a negative (positive) exposure to s_t shocks. When $SIGN_s = 1$, growth firms hedge the investment efficiency risk and command a lower risk premium.

Table OA.5.2: A model with aggregate investment efficiency shocks: model-implied moments versus data

Variable	Benchmark			
	Data	Model	$SIGN_s = 1$	$SIGN_s = -1$
Vol. of investment rate (time-series)	0.14	0.14	0.14	0.15
Vol. of investment rate (cross-sectional)	0.16	0.11	0.12	0.12
AC(1) of investment rate	0.52	0.58	0.56	0.57
AC(2) of investment rate	0.26	0.38	0.28	0.29
Inter-decile range of investment rate	0.32	0.22	0.30	0.30
Vol. of aggregate capacity utilization level	4.09	4.15	4.02	4.19
AC(1) of aggregate capacity utilization level	0.65	0.92	0.93	0.94
Vol. of aggregate sales growth	6.58	7.55	7.03	7.12
AC(1) of aggregate sales growth	0.46	0.41	0.40	0.41
Real risk-free rate	1.19	1.21	1.15	1.15
Excess market return	6.28	5.39	4.05	3.84
Volatility of excess market return	17.20	20.89	17.11	17.83

The table shows model-implied moments alongside their empirical counterparts, computed using data from 1967 to 2015. The column “Benchmark model” reports the moments implied by our single-shock benchmark model from 2 that does not include aggregate investment efficiency shocks. In contrast, the column $SIGN_s = 1$ ($SIGN_s = -1$) refers to moments implied by the two-shock model in which the investment efficiency shocks carry a positive (negative) market price of risk.

Second, the utilization premium is materially unchanged when the market price of risk of the s_t shocks turns negative. Specifically, the utilization premium with $SIGN_s = 1$ is only 1% smaller (in relative terms) than the utilization premium when $SIGN_s = -1$. This finding is once again consistent with the model’s optimality conditions that show that the choice of utilization is independent from the investment efficiency shocks (recall equation OA.5.7). The fact changing $SIGN_s$ to a negative value has a considerable effect on the value premium but only a negligible effect on the utilization premium provides an early indication that the two spreads are relatively distinct. Comparing the drivers of the two spreads also shows that the utilization premium is more closely tied to aggregate productivity risk, whereas the value premium (and other investment-related spreads) is more closely tied to investment efficiency risk. We verify this intuition below.

Distinction between the utilization premium and investment-related spreads. To demonstrate the distinction between the value and utilization premia we begin by implementing portfolio double sorts using model-simulated returns. Table OA.5.4 confirms that we continue to obtain a sizable utilization premium even if we control for book-to-market ratios. We establish this fact in three steps. First, in each period t , we control for book-to-market ratios by sorting the cross-section of model-implied firms into three portfolios on the basis of their book-to-market ratios in period $t - 1$. Here, we use the 20th and 80th percentiles of the cross-sectional distribution of book-to-market ratios as portfolio breakpoints. Next, *within* each book-to-market sorted portfolio, we further sort the cross-section of firms on the basis of their capacity utilization rates in period $t - 1$ using the same portfolio breakpoints. Finally, each portfolio is held for one period, at which point in time all portfolios are rebalanced.

The table shows that the capacity utilization premium remains almost unchanged at about 6% to 7% per annum after controlling for differences in book-to-market ratios across firms. This

Table OA.5.3: A model with aggregate investment efficiency shocks: univariate spreads

Portfolio	Panel A: Utilization premium		Panel B: B/M spread	
	$SIGN_s = 1$	$SIGN_s = -1$	$SIGN_s = 1$	$SIGN_s = -1$
Low (L)	7.80	7.60	3.35	3.51
Medium	3.99	3.84	3.53	3.56
High (H)	1.40	1.08	7.17	6.56
Spread	6.40	6.52	3.82	3.05

The table reports the average annual value-weighted returns of portfolios sorted on capacity utilization rates (Panel A) and book-to-market ratios (Panel B), as implied by the model with investment efficiency shocks. As in the empirical analysis, a firm is sorted into the high (low) utilization portfolio if its level of capacity utilization is above (below) the 90th (10th) percentile of the cross-sectional distribution of capacity utilization rates in the previous period. Similarly, a firm is sorted into the high (low) book-to-market portfolio if book-to-market ratio is above (below) the 90th (10th) percentile of the cross-sectional distribution of book-to-market ratios in the previous period. The columns $SIGN_s = 1$ ($SIGN_s = -1$) refer to the cases in which the market price of risk of the aggregate investment efficiency shocks is positive (negative).

distinction between the value and utilization spreads in this two-shock model is quantitatively more pronounced than the same distinction between these spreads in the single-shock model. This is because the sort on capacity utilization in the two-shock model is akin to a sort on each firm's exposure to aggregate productivity shocks, while the sort on book-to-market ratios is akin to a sort on each firm's exposure to investment efficiency shocks.

Table OA.5.4: A model with aggregate investment efficiency shocks: double sorts

	Utilization premium within B/M portfolios			
	Low CU	Medium CU	High CU	Spread (L-H)
Low B/M	6.61	3.17	0.33	6.28
Medium B/M	7.35	3.27	0.18	7.19
High B/M	10.48	5.72	2.78	7.70

The table shows the returns obtained from a conditional portfolio double sort procedure implied by the model with investment efficiency shocks. Specifically, the control variable, or the first dimension sorting variable, is a firm's book-to-market ratio, and the second dimension sorting variable is the capacity utilization rate. The portfolios are constructed as follows. First, in each period t , firms are first sorted into three portfolios on the basis of the cross-sectional distribution of the control variable in period $t - 1$. Here, we use the 20th and 80th percentiles of the cross-sectional distribution as portfolio breakpoints. Next, within each portfolio, firms are further sorted into three additional portfolios on the basis of the second stage sorting variable in period $t - 1$. Here, we once again use the 20th and 80th percentiles of the cross-sectional distribution of this variable as portfolio breakpoints. This procedure produces nine portfolios that are held for one period, at which point in time all portfolios are rebalanced.

Projection evidence. To further highlight the distinction between the two spreads we project the utilization premium onto the value premium via the following time-series regression

$$\text{Utilization premium}_t = \alpha + \beta \text{B/M spread}_t + \varepsilon_t. \quad (\text{OA.5.9})$$

If the utilization premium is merely a linear transformation of the value premium (i.e., if both spreads are linearly driven by the same fundamentals), then the constant in this regression will be zero. In contrast, the results from a population-sample simulation show that α in the regression above is 4.90% per annum, an economically sizable quantity. This alpha is consistent with the

empirical counterpart of this estimate that is 5.11% per annum (see Panel A of Table OA.3.20).

OA.5.2 A model with time-varying firm-level markups

Firm-level TFP is a combination of (i) technology, (ii) markups, and (iii) utilization. Our benchmark model is somewhat silent on the underpinnings of this decomposition. However, the empirical analyses in Section OA.3.6 suggest that the distinction between the TFP premium and the utilization premium originates from the fact that the former spread reflects a sort on the first two components of TFP (technology and markups) while the latter spread reflects a sort on the third component of TFP (utilization).

To mirror this type of result in the model, we introduce markups to the two-shock model described in Section OA.5.1. Because the model is a partial equilibrium, we cannot directly model monopolistic competition to generate endogenous markups. Therefore, we assume that markups fluctuate exogenously. These fluctuations may reflect time-varying degrees of price adjustment (menu) costs or substitutability between the firm's product and competitors products. Specifically, we introduce these time-varying markups by replacing equation (4) with

$$Y_{i,t} = \exp(x_t + z_{i,t}) \left[u_{i,t}^{(\theta\alpha_k \cdot \xi)} + \mu_{i,t}^\xi \right]^{\frac{1}{\xi}} K_{i,t}^{\theta\alpha_K} L_{i,t}^{\theta\alpha_L}, \quad (\text{OA.5.10})$$

where the markup process $\mu_{i,t}$ follows the dynamics

$$\mu_{i,t+1} = \rho_\mu \mu_{i,t} + \varepsilon_{i,t+1}^\mu, \text{ where } \varepsilon_{i,t+1}^\mu \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\mu^2). \quad (\text{OA.5.11})$$

To keep the setup generic, the augmented production function in equation (OA.5.10) introduces a constant elasticity of substitution (CES) aggregator in which ξ represents the elasticity of substitution between utilization and markups for output. We model the augmented output function in this way for two reasons. First, to emphasize that utilization is a factor of production that is related to the *intensity* of capital usage while markups simply induce a wedge between price and marginal cost, we raise $u_{i,t}$ but not $\mu_{i,t}$ to the power of α_k . This means that markups exist even if the firm relies entirely on labor to produce output (i.e., if the share of capital in output $\alpha_k \rightarrow 0$). Second, setting σ_μ^2 in equation (OA.5.11) to zero reduces the augmented output function in equation (OA.5.10) to our baseline production function in equation (4). Thus, this augmented production function nests our baseline production function.

Since our goal is to *numerically* demonstrate that markups further boost the independence between the utilization premium and the productivity premium, we set the dynamics of the firm-level markup process $\mu_{i,t}$ to follow an autoregressive progress with $\sigma_\mu = 0.012$, and $\rho_\mu = 0.48$. This suggests that markups are moderately persistent (in line with Basu et al. (2006)), but not very volatile (i.e., the volatility of firm-level technology shocks is much larger the volatility of markup shocks). To make the quantitative distinction the starker, we let $\xi \rightarrow 1$ so that the output function is the simple linear aggregation of the effects of utilization and markups. In particular, when $\xi \rightarrow 1$, the first-order condition for utilization in unchanged compared to equation (OA.5.7). Since $\partial Y_{i,t} / \partial u_{i,t}$ is independent of $\mu_{i,t}$, the firm's choice of utilization does not depend on the current markup. However, as $\partial Y_{i,t} / \partial K_{i,t}$ depends on $\mu_{i,t}$, firms' investment policies depend on markups and technology via the Euler equation for investment. We solve this augmented model using the same method as that described in Section OA.5.1.

Estimating firm-level TFP. Before evaluating the degree to which the utilization and productivity premia are related, we estimate firm-level TFP in the model. We do this in a similar

fashion to the empirical approach of Imrohoroglu and Tuzel (2014). While Imrohoroglu and Tuzel (2014) apply the Olley and Pakes (1996) estimator to elicit the firm-specific component of productivity, we employ a simplified OLS-based estimator since our production-based model is stripped of many of the underlying identification issues that the Olley and Pakes (1996) estimator is designed to overcome. For instance, there are no entry and exit decisions in our setup so we do not need to account for the probability that a firm survives until the next period. As such, we elicit firm-level TFP as the residual from the following panel regression estimated using model-simulated data:

$$\ln(Y_{i,t}) = \delta_t + \beta_K \ln(K_{i,t}) + \beta_L \ln(L_{i,t}) + \epsilon_{i,t}. \quad (\text{OA.5.12})$$

Here, $Y_{i,t}$ represents the output of firm i , $K_{i,t}$ represents the firm's (predetermined) capital stock, and $L_{i,t}$ represents the firm's labor at time t , respectively. δ_t is a time fixed effect that subsume the effects of aggregate productivity shocks that hit all firms at the same time. By mapping equation (OA.5.10) to projection (OA.5.12), the residual $\epsilon_{i,t}$ captures technology ($z_{i,t}$), markup ($\mu_{i,t}$), and endogenous utilization ($u_{i,t}$).

Distinction between the utilization and productivity premia. As described above, a firm's choice of utilization is (i) independent of the degree of investment efficiency s_t (whereas the firm's investment decision depends on these shocks) and (ii) independent of markups. This creates two differences between the utilization and TFP premia. First, even without markups (i.e., if $\sigma_\mu = 0$), low TFP firms disinvest more and consequently face larger exposure to s_t shocks. In contrast, the choice of utilization is independent of these s_t shocks. Second, with markups (i.e., if $\sigma_\mu > 0$), firm-level TFP is a combination of $z_{i,t}$ and $\mu_{i,t}$, while utilization depends only on $z_{i,t}$. This creates another separation: controlling for TFP, the utilization premium in the model is driven by technology and capital, while controlling for utilization, the TFP premium is mostly driven by markups. As each state variable affects the conditional exposure of firms to priced shocks, the model produces a sufficient amount of dispersion in the conditional betas of the TFP-sorted versus the utilization-sorted portfolios. In the extreme case, a low utilization firm can also be a high markup firm, rendering the firm to have high TFP in spite of having a low utilization rate. Thus, a low utilization firm is not necessarily synonymous with a low productivity firm.

Before showing the distinction between the TFP and utilization premia, we first verify that this model produces a sizable TFP spread. To do this, we construct the TFP premium by sorting the cross-section of model-implied firms on their values of $\epsilon_{i,t}$ from equation (OA.5.12), which captures the firm-level productivity in the model. In line with the results in Imrohoroglu and Tuzel (2014), we find that the low (high) productivity portfolio earns an average annual return of 5.87% (0.11%), resulting in a model-implied productivity premium of 5.76% per annum.

To show that the utilization premium is quantitatively distinct from the productivity premium we first conduct a portfolio double sort using model-simulated returns. The first stage of this procedure sorts the cross-section of model-implied firms at each time t into three portfolios on the basis of firm-level productivity at time $t - 1$. Here, we use the 20th and 80th percentiles of the cross-sectional distribution of firm-level productivity to form these portfolios. The second stage of this procedure then constructs the utilization premium *within* each of the productivity-sorted portfolios using the same percentiles of the cross-sectional distribution of utilization rates at time $t - 1$. We hold each portfolio for one period, at which point in time all portfolios are rebalanced.

The results of this double sort procedure are reported in Table OA.5.5. Panel A of the table implements these double sorts in the model that *excludes* firm-specific markups (i.e., using the

model from Section OA.5.1 in which $\sigma_\mu = 0$). We obtain an economically small but distinct utilization premium *within* each productivity-sorted portfolio. That is, controlling for firm-level productivity in the model without markups results in a utilization premium of about 1.5% per annum. As mentioned above, the two-shock model in Section OA.5.1 produces a utilization premium that is distinct from spreads that are related to investment policies, including TFP.

Table OA.5.5: Double sorts controlling for TFP: A model with and without markups

Panel A: Including markups				
	Low CU	Medium CU	High CU	Spread (L-H)
Low TFP	10.27	8.01	6.16	4.11
Medium TFP	5.60	3.44	1.44	4.16
High TFP	2.72	0.77	-1.04	3.76
Panel B: Excluding markups				
	Low CU	Medium CU	High CU	Spread (L-H)
Low TFP	1.31	0.21	-0.58	1.88
Medium TFP	4.65	3.49	2.96	1.69
High TFP	9.41	8.58	9.12	0.28

The table reports the results from a double sort procedure within a model that features investment efficiency shocks, and either includes time-varying markups (Panel A) or excludes these markups (Panel B). In each panel, we first sort the cross-section of model-implied firms at each time t into three portfolios on the basis of firm-level productivity at time $t - 1$, estimated using equation (OA.5.12). Here, we use the 20th and 80th percentiles of the cross-sectional distribution of firm-level productivity (TFP) to form these portfolios. Next, we construct the utilization premium *within* each of the three productivity-sorted portfolios by further sorting the cross-section of firms into three portfolios on the basis of their capacity utilization rates at time $t - 1$. Here, we once again use the same percentiles of the cross-sectional distribution of utilization rates as portfolio breakpoints. This procedure produces nine portfolios that are held for one period, at which point in time all portfolios are rebalanced.

Moreover, Panel B of Table OA.5.5 shows that adding firm-specific markups to the model significantly boosts the distinction between the utilization and TFP premia. Including markups in the model results in a utilization premium of around 4% per annum even when we condition on firm-level productivity. This highlights a significant degree of distinction between the two spreads.

Lastly, in this extended model, the unconditional utilization premium is above 6% (similar to the former subsection). Projecting the model-implied utilization premium on the model-implied productivity premium through a time-series regression similar to equation (OA.5.9) results in an alpha that is very close to the unconditional spread of about 6%. This model-implied alpha is in line with the empirical estimate of the same alpha reported in Panel A of Table OA.3.20. This time-series approach once again affirms the takeaways from the portfolio double sorts and Fama and MacBeth (1973) regressions: the utilization and productivity premia are distinct spreads.

OA.5.3 A model with depreciation shocks

We pursue an extension of our benchmark model motivated by the empirical evidence in Section OA.7.1. Specifically, while our baseline model assumes that depreciation rates correlate positively with utilization rates (recall equation (8)), Table OA.7.1 shows that these two quantities comove together, but not *perfectly*. As a result, we modify equation (8) to also feature an exogenous shock

to the depreciation rate. This implies that a firm's depreciation rate becomes a combination of (i) its choice of utilization rate, and (ii) a stochastic shock. This augmented depreciation function is

$$\delta(u_{i,t}) = \delta_k + \delta_u \left[\frac{u_{i,t}^{1+\lambda} - 1}{1 + \lambda} \right] + d_t, \quad \text{where } d_{t+1} = \rho_d d_t + \sigma_d \varepsilon_{t+1}^\delta, \quad (\text{OA.5.13})$$

and ε_t^δ is a standard normal i.i.d. shock to depreciation. Since depreciation shocks can be correlated with aggregate productivity, we consider two extreme cases: shocks that are both perfectly positively and perfectly negatively correlated with aggregate productivity.

We calibrate ρ_d and σ_d such that the largest value of $|d_t|$ in a discrete state space with a truncated support (i.e., two standard deviations from zero) causes the depreciation rate to change by up to 2.58%. As the unconditional depreciation rate in the model (δ_k) is 8%, the impact of the largest depreciation shock causes a firm's depreciation rate to change by roughly 33% compared to its unconditional mean. Given the large magnitude of these calibrated depreciation shocks, this exercises places an upper bound on the effect that these shocks can have on the utilization premium.

Table OA.5.6 shows that depreciation shocks that are perfectly positively (negatively) correlated with aggregate productivity decrease (increase) the model-implied utilization spread by 0.60% per annum. In either the case the model-implied spread, which is based on industry-level portfolio returns, falls within the confidence interval of the utilization premium in the data. Since the true correlation between depreciation and utilization rates is unknown, but must necessarily fall within the range of correlations we consider (i.e., $[-1, +1]$), exogenous shocks to depreciation rates do not materially impact the magnitude of the utilization premium.

Table OA.5.6: Capacity utilization spread: sensitivity to depreciation shocks

Portfolio	Positively correlated $\varepsilon_{i,t}^\delta$		Negatively correlated $\varepsilon_{i,t}^\delta$	
	$E[R^{CU}]$	β	$E[R^{CU}]$	β
Low (L)	7.70	1.06	10.21	1.64
Medium	5.94	1.01	7.79	1.51
High (H)	4.38	0.96	5.74	1.38
Spread (L-H)	3.32	0.09	4.46	0.26

The table reports the average model-implied annual value-weighted returns of portfolios sorted on capacity utilization, as well as the exposure of each utilization portfolio to market returns (β), at the industry level. As in the empirical analysis, an industry is sorted into the high (low) utilization portfolio if its level of capacity utilization is above (below) the 90th (10th) percentile of the cross-sectional distribution of capacity utilization rates in the previous period. Here, the model economy is identical to the benchmark case and calibration with one exception: the depreciation rate of each firm is subject to an exogenous shock, as represented by equation (OA.5.13). In the left (right) portion of the table these depreciation rate shocks are perfectly positively (negative) correlated with the aggregate productivity shocks. Industry-level returns are simulated using the procedure described in Section 3.1. Population moments are obtained from one simulation of 50 industries for 40,000 periods (years).

OA.6 Model implications for macro-finance modeling

The implications of flexible utilization for asset prices span beyond the utilization premium. We highlight the roles of flexible utilization for jointly targeting cross-sectional risk premia and

investment moments in the presence of real options while relying on a parsimonious adjustment cost specification. We first show the failures of the model without utilization to target asset-pricing and production moments. We then explain how utilization provides a solution to these misses. We demonstrate that flexible utilization permits us to target key moments with a lower degree of adjustment costs vis-à-vis a model with fixed utilization.

OA.6.1 A fixed utilization model: the failures

The benchmark model’s success in jointly fitting (i) the volatility and skewness of investment, both across time and across firms, and (ii) risk premia, crucially hinges on variable capacity utilization rates. To illustrate this point, row (1) of Table 9 shows model-implied moments in an economy without flexible utilization (i.e., $\lambda \rightarrow \infty$).

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

The fixed utilization model fails to capture the aforementioned moments since (1) the fixed adjustment cost makes disinvestment a real option, and (2) without flexible utilization, a firm’s *only* way to respond to a negative productivity shock is by exercising this option.

As discussed in Section 2.2, if a drop in productivity at time t is not extremely severe, then a “wait and see” effect tends to dominate. Thus, declines in productivity typically lead to periods of investment-policy inaction in which many firms do not alter their capital stock. Each waiting firm j sets its investment rate, $i_{j,\tau}$, to the constant depreciation rate of δ_k for all $\tau \in [t, t + \hat{t})$, where \hat{t} is the ending time of the endogenous inaction period. Since a mass of waiting firms are clustered around the center of investment’s distribution (i.e., around δ_k), investment’s dispersion and cross-sectional skewness both decrease.

Furthermore, if productivity remains persistently low, then at time $t + \hat{t}$ waiting firms pass a tipping point in which they are overly burdened with unproductive capital and choose to disinvest this capital sharply. This implies that $i_{j,t+\hat{t}} << \delta_k$.³⁶ Thus, these periods of inaction are often followed by negative investment spikes, producing the negative skewness of firm-level investment that is inconsistent with the data.³⁷

The distorted distribution of investment rates in the model with fixed utilization also has an

³⁶Put differently, when firms are close to the disinvestment threshold, then the investment policy is locally concave and the expected value of investment becomes negative.

³⁷Importantly, the counterfactuals in the model with fixed utilization cannot be remedied simply via the aggregation level. As many real options operate at the plant level, one can argue that a collection of production units in the model comprise one firm. Aggregation of many units into a single firm does indeed smooth model-implied investment rates by shrinking periods of investment inaction, and lowering the size of disinvestment jumps. However, these model-implied moments remain unaligned with the data. We verify this in untabulated simulation by aggregating 100 production units into a firm. The resulting model-implied volatility of investment is smaller than the data. While the skewness of investment turns positive, this quantity is close to zero (remaining significantly lower than the data).

adverse impact on risk premia. Because investment's distribution features too little dispersion and asymmetry, there is too little heterogeneity between firms' risk exposures. Sorting firms into portfolios based on investment (or valuation ratios) implies that both the top and bottom quintiles (tails) contain fewer extreme outcomes compared to the benchmark with flexible utilization. As differences in cross-sectional risk premia are fundamentally driven by heterogeneity in investment, cross-sectional spreads get smaller with fixed utilization.

OA.6.2 Flexible utilization: a solution

Our benchmark model with flexible utilization overcomes the counterfactuals outlined in Section OA.6.1 by making the depreciation rate endogenously stochastic. This improved model fit is highlighted in row (2) of Table 9 by showing model-implied moments under the benchmark value of λ . When firms can choose utilization, they have an extra mechanism by which to scale down production in response to adverse productivity shocks, even as they "wait and see" if productivity recovers. That is, firms can respond to moderate drops in productivity by utilizing their existing machines less intensively rather than selling machines. As underutilized capital depreciates slower, more capital is preserved for more productive future periods (i.e., $\delta(u_{j,\tau}) < \delta_k$ if $u_{j,\tau} < 1 \forall \tau \in [t, t + \hat{t}]$).

Lower utilization reduces the natural investment rate needed to maintain the current capital stock. Thus, even as firms wait to sell capital, the investment required to maintain existing machines, $i_{j,\tau} = \delta(u_{j,\tau})$, becomes endogenously stochastic. This time-varying depreciation eliminates the long periods of constant investment. The time-series volatility of firm-level investment rises, and the cross-sectional dispersion of investment increases. To see the latter, note that firms' utilization rates depend on idiosyncratic productivity shocks. Since these shocks differ between waiting firms, $u_{j,\tau} \neq u_{k,\tau} \Rightarrow i_{j,\tau} \neq i_{k,\tau}$ for firms j and k .

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

The increases in the skewness and dispersion of investment under flexible utilization also boost risk premia spreads, as seen by comparing the value premium between rows (2) and (1) of Table 9. A larger value premium under flexible utilization can be attributed to the fact that the cross-section of investment rates is more dispersed and almost 12 times as skewed in the model with flexible utilization. Greater dispersion in investment leads to more heterogeneity in risk exposures to aggregate productivity, which increases return spreads.³⁸

³⁸The increase in the cross-sectional skewness of investment also has an impact on the value premium. With fixed utilization, and *symmetric* cross-sectional distribution, the portfolio of growth firms (bottom 20% of book-to-market) includes both firms with very high and moderately high investment rates (or Tobin's Q). With flexible utilization, and *asymmetric* cross-sectional distribution, the right tail of investment's distribution becomes thicker, and the portfolio of growth firms includes firms with only very high investment rates. As these firms expand capacity significantly (suggesting a much higher Q) precisely when the price of risk is high (bad states), their skewed investment behavior provides an excellent hedge against bad states. This decreases the risk premium of the growth portfolio, and increases the magnitude of the value premium.

More generally, as shown in rows (3) to (6) of Panel A in Table 9, the value of λ has a substantial quantitative impact on matching the data. As utilization becomes more flexible (i.e., λ decreases), the time-series/cross-sectional skewness and volatility of investment rise, the autocorrelation of investment slightly declines, and risk premia increase as well. This generally moves each moment towards its empirical counterpart when compared to the case of $\lambda \rightarrow \infty$. In particular, row (6) shows that when utilization becomes less flexible (i.e., λ is finite but high), investment's skewness and dispersion are too low. Rows (4) and (5) show that our results are only mildly affected by small perturbations of the benchmark value of λ . 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.

OA.6.3 Required adjustment costs under flexible utilization

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

Perturbing adjustment costs. In rows (7) to (10) of Panel B in Table 9 we alter the quadratic adjustment cost while keeping utilization fixed. Rows (7) and (8) consider the case of lower frictions compared to the benchmark. Sufficiently lower friction (row (7)) 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 frictions also cause risk premia to fall. The value premium, which is already too low in the model with fixed utilization, falls in row (7) to almost half of its empirical magnitude.

The diminished value premium in the model with fixed utilization can be boosted by increasing the quadratic capital adjustment cost. With higher adjustment costs, shocks are absorbed in asset prices rather than investment quantities. We demonstrate this in rows (9) and (10). We search for an adjustment cost parameter ϕ to match the value premium in the model with fixed utilization. Our structural search suggests that ϕ needs to be around 3.00 to match this spread (see row (10)). While this parameter is broadly consistent with existing literature, this value is *double* the value of ϕ under our benchmark model with flexible utilization. Moreover, doubling ϕ simultaneously distorts investment's distribution. Investment's dispersion becomes a quarter of its empirical magnitude. The time-series skewness of investment becomes even more counterfactually negative.³⁹

Flexible utilization provides a channel that addresses the aforementioned concerns. Row (2) of Table 9 shows that flexible utilization allows our baseline model to feature adjustment costs that are smaller than those required with fixed utilization. These smaller costs are sufficient to simul-

³⁹In untabulated results, we further verify that under fixed utilization, the value premium cannot be targeted successfully in the model (jointly with investment's moments) by changing the fixed cost f . Specifically, we find that lowering the fixed cost f causes investment's skewness to rise and move closer to the data. However, even in the extreme case of $f = 0$, the cross-sectional skewness of investment is only a half of its empirical magnitude. Moreover, lowering f decreases the magnitude of the value premium (e.g., when $f = 0.01$, the implied value premium is merely 2.7%). Similarly, we find that while raising the fixed cost f to 0.06 allows us to obtain a value premium of 3.2% (roughly consistent with the data), this causes the model's investment mismatch to become even more severe. The time-series and cross-sectional skewness of investment become counterfactually negative (about -0.75), and investment's volatility falls to 10%.

taneously produce sizable risk premia spreads, including many periods of depressed investment.⁴⁰ This happens because of the following key mechanisms.

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

The second mechanism involves the changes in the cross-sectional distribution of investment rates that are caused by flexible utilization. As we outline in Section OA.6.2, utilization makes the cross-sectional distribution of investment more dispersed and skewed. By featuring more extreme observations in the tails of investment's distribution, one can obtain quantitatively large return spreads with moderated values of adjustment costs.

The third mechanism is utilization's ability to enhance payout cyclicity. Value firms have low idiosyncratic productivity and high capital, and desire to reduce excess capital in bad states. With fixed utilization, these firms are riskier because they need to pay large quadratic costs that reduce their payoff precisely when aggregate productivity is low. With flexible utilization, these firms also desire to lower utilization to conserve capital for future periods. The reduction in utilization implies today's output is even lower, contemporaneously with a bad aggregate state. Thus, the cyclicity of output is larger, amplifying risk.

Alternative adjustment costs. A model with fixed utilization can potentially match the aforementioned moments using a more elaborate adjustment cost function. Extra adjustment costs may include piecewise quadratic and linear terms (e.g., Belo and Lin (2012)).

While it is hard to rule out the possibility of an admissible adjustment cost function, the next subsection provides evidence that adjustment costs in the form of Cooper and Haltiwanger (2006) are unlikely to reconcile risk premia alongside the time-series and cross-sectional moments of investment. We augment the model to feature asymmetric quadratic and linear adjustment costs. The downside quadratic (linear) coefficient is ten times (two to three times) larger than the upside coefficient, in line with extant papers. We find that the asymmetric fixed cost has a negligible effect on all moments. By contrast, when the quadratic adjustment cost is asymmetric, and re-calibrated to successfully match the cross-sectional skewness of investment, the time-series skewness and volatility of investment are much larger than the data, and return spreads are still too low.

⁴⁰Moreover, when utilization is flexible, this result is not very sensitive to small perturbations of ϕ . Rows (12) and (13) of Table 9 show that when the quadratic adjustment cost (ϕ) slightly increases (decreases), the volatility of investment drops (rises), the value premium slightly increases (decreases), but these moments are almost identical to the benchmark in row (2). Row (14) shows that if ϕ is set to 3.00 (which is required in the case of fixed utilization), then the implied value premium is almost 5%, about 1.5% above its value under fixed utilization. In this case, the cross-sectional skewness of investment aligns very well with the data, but the cross-sectional dispersion is too low.

OA.6.4 Alternative adjustment cost specification

We document that even if the capital adjustment cost function given by equation (9) is augmented to include piecewise linear and quadratic terms, the model with fixed utilization is not likely to match key investment-related moments, while generating high risk-premia spreads.

Consistent with the works of Cooper and Haltiwanger (2006), Belo and Lin (2012), and Belo et al. (2014), among others, we generalize the adjustment cost function employed by our benchmark analysis to include asymmetric linear and quadratic adjustment cost. Specifically, the adjustment cost function we employ is given by

$$G_{i,t} = \left[\frac{\phi^+}{2} (ni_{i,t})^2 \mathbf{1}_{\{ni_{i,t}>0\}} + \frac{\phi^-}{2} (ni_{i,t})^2 \mathbf{1}_{\{ni_{i,t}<0\}} + f^+ \mathbf{1}_{\{ni_{i,t}>0\}} + f^- \mathbf{1}_{\{ni_{i,t}<0\}} \right] K_{i,t}. \quad (\text{OA.6.1})$$

Here, $ni_{i,t} \equiv \frac{I_{i,t}}{K_{i,t}} - \delta_k$ denotes the net investment rate of firm i at time t , while the functions denoted by $\mathbf{1}_{\{ni_{i,t}>0\}}$ ($\mathbf{1}_{\{ni_{i,t}<0\}}$) are indicator variables that take on a value of one when the firm increases (reduces) its stock of capital. The adjustment parameter ϕ^+ (ϕ^-) capture the quadratic adjustment cost of increasing (decreasing) capacity, while f^+ (f^-) captures the fixed adjustment cost of increasing (decreasing) capacity. While our benchmark adjustment cost specification imposes the constraints that $\phi^+ = \phi^-$ and $f^+ = 0$, we relax these constraints below and consider more general specification in which $\phi^+ \neq \phi^-$ and $f^+ > 0$.

The model in Zhang (2005) suggests the quadratic adjustment cost for investment is around one tenth of the magnitude of that for disinvestment. Accordingly, we focus on the case that $\phi^+ = \phi^- \times \frac{1}{10}$. Likewise, the fixed cost of investment in Belo and Lin (2012) and Belo et al. (2014) is approximately one third to one half of the magnitude of that for disinvestment. Consequently, we consider the cases in which $f^+ = f^- \times \frac{1}{3}$ or $f^+ = f^- \times \frac{1}{2}$. For completeness, we also consider calibrations in which the ratios f^+/f^- or ϕ^+/ϕ^- differ from these prior studies.

Table OA.6.1 reports the model-implied time-series and cross-sectional moments of investment rates, as well as the magnitude of cross-sectional risk premia, in a fixed-utilization model featuring asymmetric capital adjustment costs. Rows (2) to (6) of the table consider calibrations in which $\phi^+ = \phi^-$ in equation (OA.6.1), identical to our baseline calibration reported in row (1), but $f^+ > 0$. We keep in f^- at a positive value, consistent with the reason outlined in Section 2.1. By and large, perturbing introducing an upside linear adjustment cost does not help to reconcile the data. For instance, when $f^+ = f^- \times \frac{1}{2}$ in row (6) the dispersion and volatility of investment rates, as well as the magnitudes of the risk premia, remain far lower than their counterparts in the data.

Rows (7) to (9) of the table keep $f^+ = 0$ (identical to our baseline model) but allow for asymmetry in the quadratic adjustment costs (i.e., $\phi^+ \neq \phi^-$). Focusing on row (8), a calibration in which $\phi^+ = \phi^- \times \frac{1}{10}$, shows that this asymmetry can reconcile *cross-sectional* moments of investment rates with the data (e.g., the cross-sectional skewness is about 1.8 in both the model and the data). However, the same calibration fails to match the *time-series* moments of investment rates and risk premia. Specifically, the model-implied time-series volatility (skewness) of investment rates is almost twice (five times) as high as its empirical counterpart. Moreover, risk premia in this model are too low compared to the data.

In rows (10) and (11) the adjustment cost function is calibrated to feature both asymmetric linear *and* quadratic adjustment costs. Here, the degree of asymmetry follows the degrees of asymmetry considered in the literature (see, e.g., Zhang (2005), Belo and Lin (2012), and Belo et al.

(2014)). The results indicate that even with the most general form of the asymmetric adjustment costs, the fixed-utilization model is only able to reconcile the cross-sectional moments at the expense of not matching the time-series of investment and risk premia. For instance, row (10) is quite similar to row (8) in terms of matches and mismatches.

To complement the above evidence, we also search for a value of ϕ^- that is able to match the value premium, similar to the exercise presented in Section OA.6.3. The search result yields two conclusions that are almost identical to Section OA.6.3. First, even though we feature a more elaborate adjustment cost, as in Cooper and Haltiwanger (2006), we still need to increase the exogenous amount of capital friction compared to the flexible utilization model. We find that the downside quadratic adjustment cost ϕ^- has to be *doubled*. That is, a calibration in which $\phi^- = 3$, $\phi^+ = \phi^-/10$ and $f^+ = f^-/3$ yields a value premium of 3.5%, which is close to the data. Second, this higher value of ϕ^- renders two prominent counterfactual moments for investment: the time-series skewness of investment is much larger than the data (about 3.5), and the cross-sectional dispersion is about a half of the data (about 0.08).

To conclude, the evidence points out that asymmetry in the linear cost has only a marginal effect on the results, while asymmetry in ϕ helps only for the cross-section of investment. In particular, row (8) shows that without deviating from standard values of α , δ_k , or α_l , a model with fixed utilization and real options cannot match the data. Importantly, even if such calibration was were, flexible utilization would still offer valuable merit: flexible utilization provides a way to rely on a lower dimensional adjustment cost function and endogenize the implications of additional exogenous calibration parameters using a micro-founded margin. The only additional model parameter needed to accommodate flexible utilization is λ .

OA.7 Utilization and depreciation: Empirical evidence

OA.7.1 Utilization and depreciation dynamics

Equation (8) suggests that firms' depreciation rates should correlate positively with their utilization rates. In this section we check this prediction and explore its implications for the accuracy and the frequency of depreciation's measurement.

We examine the relation between utilization and depreciation via the projection

$$\Delta\delta_{j,t} = d_j + d_u\Delta u_{j,t} + d_x X_{j,t} + \varepsilon_{j,t},$$

where j denotes an industry index, $\Delta\delta_{j,t}$ is the log growth of industry j 's depreciation rate from BEA, $\Delta u_{j,t}$ is the log growth of industry j 's utilization rate, and $X_{j,t}$ is a control variable. We use the log growth of depreciation and utilization to reduce persistence in these variables, and to account for a non-linear relation between the level of the two. All variables are standardized for the ease of interpretation. The results of the projection are reported in Table OA.7.1. In columns (1) and (2) of the table we run the projection without any controls. We show that the simple correlation between log depreciation growth and log utilization growth is 30%, and that this correlation is not affected by the inclusion of industry fixed effects.

Recent studies in production-based asset pricing show that BEA- and Compustat-implied depreciation rates are strikingly different. The use of one over the other can lead to economically sizable differences in the distribution of gross investment rates (e.g., Clementi and Palazzo (2019); Bai et al. (2019)). In line with these papers, columns (3) and (4) of Table OA.7.1 demonstrate the

Table OA.6.1: Model-implied moments and asymmetric capital adjustment costs

Row	Model	Time-series			Cross-sectional		Risk premia	
		$\sigma_{TS}(ik)$	$S_{TS}(ik)$	$\rho_1(ik)$	$\sigma_{CS}(ik)$	$S_{CS}(ik)$	$E[R^{bm}]$	$E[R^{ik}]$
	Data	0.14	0.67	0.52	0.16	1.89	3.71	3.70
Baseline without utilization								
(1)	$(\lambda \rightarrow \infty)$	0.11	-0.09	0.64	0.07	0.11	2.93	2.30
Unchanged ϕ and $f^+ > 0$								
(2)	$f^+ = f^- \times \frac{1}{20}$	0.11	-0.03	0.63	0.07	0.16	2.92	2.11
(3)	$f^+ = f^- \times \frac{1}{10}$	0.11	0.03	0.62	0.07	0.23	2.91	2.01
(4)	$f^+ = f^- \times \frac{1}{5}$	0.11	0.13	0.61	0.07	0.33	2.91	1.85
(5)	$f^+ = f^- \times \frac{1}{3}$	0.11	0.26	0.60	0.07	0.47	2.91	1.74
(6)	$f^+ = f^- \times \frac{1}{2}$	0.11	0.40	0.59	0.07	0.63	2.90	1.60
Asymmetric ϕ and $f^+ = 0$								
(7)	$\phi^+ = \phi^- \times \frac{1}{20}$	0.30	4.83	0.34	0.21	2.56	2.28	2.18
(8)	$\phi^+ = \phi^- \times \frac{1}{10}$	0.24	3.60	0.42	0.17	1.83	2.46	2.22
(9)	$\phi^+ = \phi^- \times \frac{1}{5}$	0.19	2.45	0.51	0.13	1.21	2.62	2.26
Asymmetric ϕ and $f^+ > 0$								
(10)	$f^+ = f^- \times \frac{1}{3}$, and $\phi^+ = \phi^- \times \frac{1}{10}$	0.25	3.66	0.39	0.18	2.08	2.41	1.92
(11)	$f^+ = f^- \times \frac{1}{2}$, and $\phi^+ = \phi^- \times \frac{1}{10}$	0.25	3.71	0.37	0.18	2.19	2.41	1.85

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 of the model featuring an asymmetric capital adjustment cost function. Specifically, the augmented adjustment cost function is

$$G_{i,t} = \left[\frac{\phi^+}{2} (n_{i,t})^2 \mathbf{1}_{\{n_{i,t}>0\}} + \frac{\phi^-}{2} (n_{i,t})^2 \mathbf{1}_{\{n_{i,t}<0\}} + f^+ \mathbf{1}_{\{n_{i,t}>0\}} + f^- \mathbf{1}_{\{n_{i,t}<0\}} \right] K_{i,t},$$

where $n_{i,t} \equiv \frac{I_{i,t}}{K_{i,t}} - \delta_k$ represents the net investment rate of the firm, and the functions denoted by $\mathbf{1}_{\{n_{i,t}>0\}}$ ($\mathbf{1}_{\{n_{i,t}<0\}}$) are indicator variables that take on a value of one when the firm increases (reduces) its capacity. Here, . We consider various specifications of the function above, whereby we alter ϕ^+ (f^+) to either be unchanged relative to the baseline model, or a fixed multiple of ϕ^- (f^-). The table reports the time-series volatility ($\sigma_{TS}(ik)$), skewness ($S_{TS}(ik)$), and the first-order autocorrelation ($\rho_1(ik)$) of firm-level investment rates, as well as the cross-sectional dispersion ($\sigma_{CS}(ik)$) and skewness ($S_{CS}(ik)$) of investment rates. In addition, the table also reports the value premium ($E[R^{bm}]$) and investment premium ($E[R^{ik}]$) 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 the degree of asymmetry in the alternative adjustment cost function. All moments are based on a simulations of 1,000 firms over 40,000 periods (years). Finally, the top row of the table also reports the empirical counterpart of each moment.

discrepancy between these two depreciation measures. We set $X_{j,t}$ to be the log growth of industry j 's Compustat-based depreciation rate, and restrict d_u to zero. The correlation between the growth of these depreciation measures is only 3%.⁴¹

In columns (5) and (6) we do not restrict d_u to be zero. First, the positive correlation between BEA-implied depreciation growth and utilization growth remains positive and sizable when controlling for Compustat-implied depreciation. Second, the (partial) correlation between the growth rates of BEA- and Compustat-implied depreciation increases to 14%. Thus, utilization narrows the wedge between these two measures. While measurement error may exist in both measures, the fact that the correlation between the two increases when controlling for utilization suggests that

⁴¹We describe the measurement of the these depreciation rates in Section OA.1 of the Online Appendix.

Table OA.7.1: Empirical relation between capacity utilization and depreciation rates

	(1)	(2)	(3)	(4)	(5)	(6)
β_{UTIL}	0.30 (3.11)	0.30 (3.14)			0.28 (2.80)	0.28 (2.83)
β_{COMP}			0.03 (3.50)	0.03 (3.53)	0.13 (3.54)	0.14 (3.58)
Industry FE	No	Yes	No	Yes	No	Yes
R^2	0.09	0.09	0.02	0.02	0.10	0.10

The table reports the empirical relation between the industry-level capacity utilization rate, industry-level depreciation rate from the BEA, and industry-level depreciation rate from Compustat. In each specification considered in the table we run projections of the log-growth rate of BEA-implied depreciation on the log-growth rates of capacity utilization and Compustat-implied depreciation rates, and standardize all variables for ease of interpretation. Columns (1), (3), and (5) of the table estimate pooled-OLS regressions, whereas columns (2), (4), and (6) of the table estimate panel regressions including industry fixed effects. t -statistics, reported in parentheses, and computed using standard errors clustered at the industry level. Finally, the time span underlying the regressions is from January 1967 to December 2015.

utilization can be used to accurately filter the true depreciation rate. We briefly illustrate this point in the next subsection.

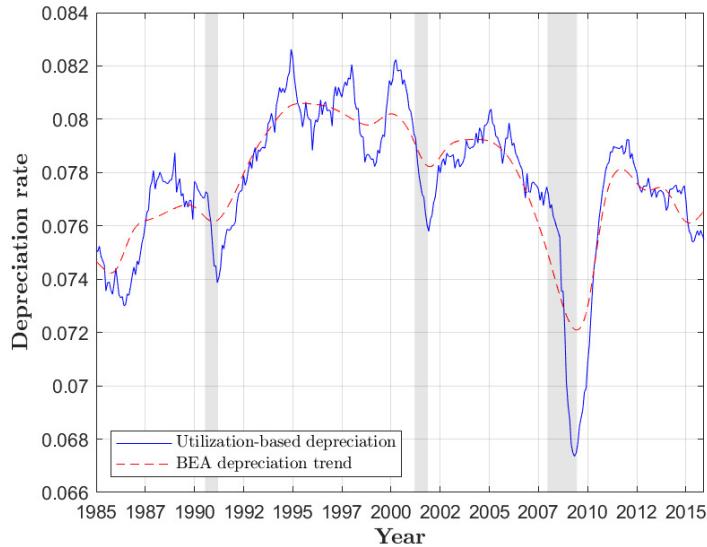
OA.7.2 High-frequency depreciation based on utilization data

As utilization data is available at the monthly frequency, the utilization-implied depreciation rate we propose is computed at a higher frequency than depreciation rates implied by either BEA or Compustat. First, to be consistent with the model, we adjust each industry's utilization rate to have a mean of one. Then, for each industry j , we obtain a utilization-implied depreciation rate, $\delta(u_j)$, by applying equation (8) to the industry's utilization data. Here, we use the model parameters in Table 6. Second, we average these depreciation rates across all industries to obtain an aggregate utilization-implied depreciation rate, $\delta(u_{agg})$. Third, we adjust $\delta(u_{agg})$ to share the same trend as the aggregate depreciation rate from the BEA. We do this by combining the business-cycle component of $\delta(u_{agg})$ with the stochastic trend component of the BEA's aggregate depreciation rate.⁴² We obtain the components of each time-series using the Hodrick and Prescott (1997) filter.

In Figure OA.7.1 we plot the monthly time-series of $\delta(u_{agg})$ alongside the trend of the BEA's aggregate depreciation rate. By construction, the two time-series share the same trend, but $\delta(u_{agg})$ shows high-frequency business-cycle fluctuations around this common trend. These fluctuations could be important as they amplify the volatility of gross investment rates and can help to reconcile the dynamics of BEA- and Compustat-implied depreciation rates.

⁴²The third step is optional and meant to ensure both series follow the same trend. An alternative measure of the utilization-implied depreciation rate involves only the first two steps, with similar results.

Figure OA.7.1: Utilization-based high-frequency depreciation rate



The figure shows a high-frequency measure of the aggregate depreciation rate. The figure reports the time series of the high-frequency depreciation rate from 1985, the beginning of the Great Moderation, to 2015.

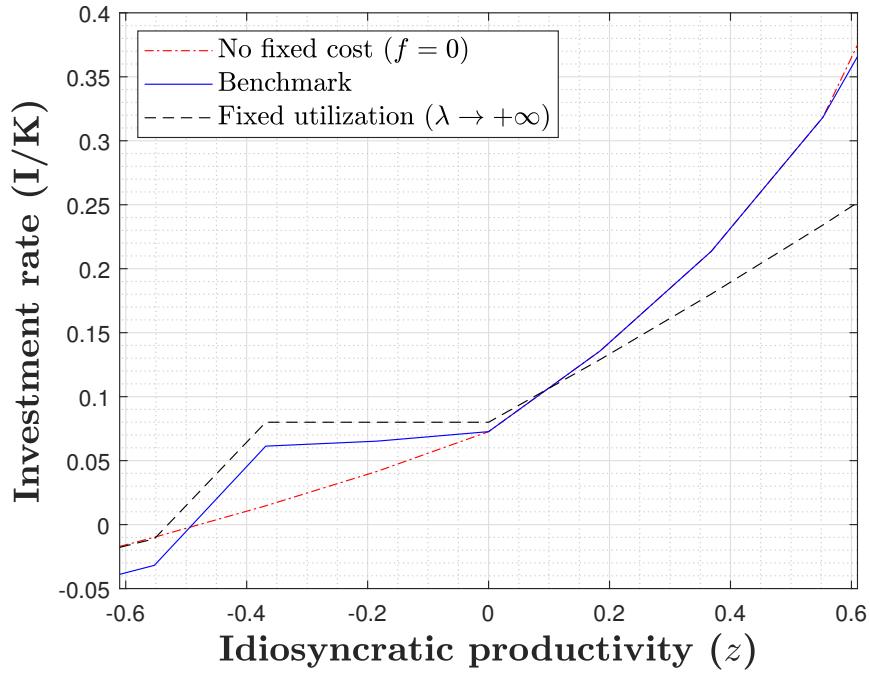
OA.8 Supplemental tables and figures

Table OA.8.1: Model-implied CAPM alpha

Portfolio	$[R^{CU}]$	α_{CAPM}
Low (L)	9.01 [4.24,16.87]	0.36 [-1.69,2.55]
Medium	6.93 [3.14,14.07]	-1.24 [-1.81,-0.72]
High (H)	4.97 [-0.10,12.66]	-2.80 [-6.26,0.59]
Spread (L-H)	4.04 [0.38,8.75]	3.16 [-0.94,7.42]

The table reports the average annual value-weighted returns and CAPM alphas (α_{CAPM}) of portfolios sorted on capacity utilization at the industry-level across short-sample simulations of our model economy. As in the empirical analysis, an industry is sorted into the high (low) utilization portfolio if its level of capacity utilization is above (below) the 90th (10th) percentile of the cross-sectional distribution of capacity utilization rates in the previous period. Industry-level returns are simulated using the procedure described in Section 3.1, and short-sample moments are obtained by averaging moments across 500 simulations of 50 industries for 50 periods (years). Finally, square brackets report the 95% confidence interval related to each moment across the 500 Monte Carlo simulations of the economy.

Figure OA.8.1: 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 let z varying between two standard deviations of its mean value. We consider the I/K policy under three versions of our model: (1) The benchmark model (solid blue line), (2) the model without fixed costs (i.e., $f = 0$) (the dashed red line), and (3) the model with fixed utilization (i.e., $\lambda \rightarrow +\infty$).