Entry Costs and the Macroeconomy

by Germán Gutiérrez, Callum Jones and Thomas Philippon
IMF Working Paper

Western Hemisphere Department

Entry Costs and the Macroeconomy

Prepared by Germán Gutiérrez, Callum Jones and Thomas Philippon

Authorized for distribution by Luca Ricci

November 2019

Abstract

We combine a structural model with cross-sectional micro data to identify the causes and consequences of rising concentration in the US economy. Using asset prices and industry data, we estimate realized and anticipated shocks that drive entry and concentration. We validate our approach by showing that the model-implied entry shocks correlate with independently constructed measures of entry regulations and M&As. We conclude that entry costs have risen in the U.S. over the past 20 years and have depressed capital and consumption by about seven percent.

JEL Classification Numbers: E2, E4, E5, L4.

Keywords: Corporate Investment, Competition, Tobin's Q, Zero Lower Bound.

Author’s E-Mail Address: ggutierrez@stern.nyu.edu; cjones3@imf.org; tphilipp@stern.nyu.edu

---

1 We are grateful to Fabio Ghironi, Nic Kozeniauskas, William Lincoln for their feedback, and Susanto Basu, Martin Eichenbaum, Nicolas Crouzet, Emmanuel Farhi, Federico Diez, and seminar participants at the IMF and the NBER Summer Institute for comments. The Appendix is available online at https://callumjones.github.io.

2 New York University.

3 International Monetary Fund.

4 New York University, CEPR and NBER.
1 Introduction

Four stylized facts characterize the U.S. economy in recent decades. (i) U.S. industries have become more concentrated, while (ii) profit margins have increased; (iii) business dynamism – including firm entry rates and the share of young firms in economic activity – has fallen; and (iv) business investment has been low relative to measures of profitability, funding costs, and market values.\(^1\) While these stylized facts are well established (Decker, Haltiwanger, Jarmin and Miranda, 2014; Furman, 2015; Grullon, Larkin and Michaely, 2016; Gutiérrez and Philippon, 2017), their interpretation remains controversial. There is little agreement about the causes and consequences of these evolutions. For instance, Furman (2015) and CEA (2016) argue that the rise in concentration suggests “economic rents and barriers to competition”, while Autor et al. (2017b) argue almost exactly the opposite: that concentration reflects “a winner takes most feature” explained by the fact that “consumers have become more sensitive to price and quality due to greater product market competition.” The evolution of profits and investment could also be explained by intangible capital deepening, as discussed in Crouzet and Eberly (2018).\(^2\)

Several reasons explain why the literature has remained inconclusive. The first challenge is that entry, exit, concentration, investment, and markups are all jointly endogenous and it is difficult to find exogenous variation in any of these variables. The second challenge is that the macroeconomic implications of declining competition are difficult to analyze outside a fully specified model. As a result, both the empirical and the theoretical literature are limited: little has been done to identify the causes of the four trends empirically; and most macroeconomic models simply assume that markups have changed and study the implications without attempting to link them to independent measures of barriers to competition.

Our paper is a first attempt to address these issues. We propose a new approach to disentangle the various explanations using a structural model together with micro data. We build a fully specified macro model of the U.S. economy, featuring many industries and taking into account not only entry and investment, but also demand and interest rates.

At the industry level, the key identification issue is that entry and concentration are endogenous.

\(^1\)See Section 2 for additional details on these facts.

\(^2\)Finally trade and globalization can explain some of the same facts (Feenstra and Weinstein, 2017; Impullitti et al., 2017). Foreign competition can lead to an increase in domestic concentration and a decoupling of firm value from the localization of investment. We control for exports and imports in our analyses. Foreign competition is significant for about 3/4 of the manufacturing sector, or about 10% of the private economy. One could entertain other hypotheses – such as weak demand or credit constraints – but previous research has shown that they do not fit the facts. See Covarrubias et al. (2019) for detailed discussions and references.
To be concrete, consider an industry $j$ where firms operate competitively under decreasing returns to scale. Suppose industry $j$ receives the news at time $t$ that the demand for its products will increase at some time $t + \tau$ in the future. There would be immediate entry of new firms in the industry. As a result, we would measure a decrease in concentration followed and/or accompanied by an increase in output and investment. Anticipated demand (or productivity) shocks can thus explain why we see more activity in less concentrated industries even if it is not due to competition.

We make three contributions to the literature. Our first contribution is to use our model together with macro time series and panel data from firms and industries to address the identification issue. We specify a model with a rich set of demand and supply shocks, including shocks to investors’ expectations, and use the model’s joint restrictions to identify the shocks. Using current output together with forward looking asset prices, our model can recover the shocks to expected demand across industries. We find that these shocks are large in the late 1990s, and, exactly as theory would predict, they explain variation in entry rates. Instead of being an empirical roadblock, however, these large shocks become a useful way to estimate some important parameters of the model, such as the elasticity of entry to Tobin’s $Q$.

Our second contribution is to link the structurally estimated shocks to measures of regulations and antitrust enforcement, thereby providing the first direct structural evidence that policy is (partly) responsible for decreasing competition in the U.S. economy. This requires several steps, as we explain below, but the broad intuition is relatively simple. Using Bayesian estimation methods, our model recovers annual industry-level entry cost shocks – which can then be compared to independent measures of entry regulations and antitrust activities computed from the micro-data. We show that the model-implied entry cost shocks track rather closely our empirical measures of entry regulations, even though they come from entirely different data sources and methodologies. In the aggregate, entry shocks are large and important for matching time series. They explain 15% of the variation in investment, 29% of the variation in output and about half of the variation in concentration.

The last contribution of the paper is methodological but it turns out to be empirically important. We specify the likelihood function for the data panel and estimate the model taking into account the zero lower bound (ZLB). We solve for the path of the economy using the solution method and approach of Jones (2018), which uses a Kalman filter and information about the expected durations of the ZLB to back out the other shocks that drive the model (including productivity, discount rate and risk premia). While the ZLB – and monetary policy more generally – might seem like a separate issue,
we show that it plays an important role in our estimations for two reasons. First, the ZLB depresses the economy and therefore impacts entry and concentration. Failing to properly model the ZLB would over-estimate the magnitude and impact of aggregate entry costs. Second, because entry costs affect the natural rate of interest, the consequences of entry shocks are different with or without the ZLB.

To summarize, our main finding is that increasing barriers to entry have had a significant impact on macro-economic dynamics over the past 30 years. For instance, absent the decrease in competition since 2003, consumption and the capital stock would be about 7 percent higher.

**Literature.** Our approach introduces several new ways to examine the relationship between firm entry, competition, and the macroeconomy. This has been the subject of a long, but predominantly empirical literature in the U.S. Bernard et al. (2010) study the contribution of product creation and destruction to aggregate output. They estimate that product creation by both existing firms and new firms accounts for 47 percent of output growth in a 5-year period. Decker et al. (2015) argue that, whereas in the 1980’s and 1990’s declining dynamism was observed in selected sectors (notably retail), the decline was observed across all sectors in the 2000’s, including the traditionally high-growth information technology sector. Davis and Haltiwanger (2019) emphasize the role of the housing market for explaining the decline.

Furman (2015) shows that “the distribution of returns to capital has grown increasingly skewed and the high returns increasingly persistent” and argues that it “potentially reflects the rising influence of economic rents and barriers to competition.” CEA (2016) and Grullon et al. (2016) are the first papers to extensively document the broad increases in profits and concentration. Grullon et al. (2016) also show that firms in concentrating industries experience positive abnormal stock returns and more profitable M&A deals. Blonigen and Pierce (2016) find that M&As are associated with increases in average markups. Autor et al. (2017b) study the link between concentration and the labor share. An important issue in the literature is the measurement of markups and excess profits. De Loecker and Eeckhout (2017) estimate markups using the ratio of sales to costs-of-goods-sold and find a large increase in mark-ups. Barkai (2017), on the other hand, estimates the required return on capital directly and finds a moderate increase in excess profits. Both estimates are controversial, however, so we do not use them directly in our analyses (Basu, 2019; Syverson, 2019; Covarrubias et al., 2019).

---

3Furman (2015) also emphasizes the weakness of corporate fixed investment and points out that low investment has coincided with high private returns to capital, implying an increase in the payout rate (dividends and shares buyback).
Methodologically, our paper is related to general equilibrium models of imperfect competition. Bilbiie et al. (2012) study how entry affects the propagation of business cycles in a standard RBC model with technology shocks. Endogenous entry costs generate variations in the stock market price of investment, which in turn affect entry decisions. This is consistent with our model. We build on their approach, and use data on Tobin’s $Q$ to back out a time-series of entry costs. Relatedly, Cacciatore and Fiori (2016) estimate that reducing entry costs in Europe to the level observed in the U.S. in the late 1990s would have increased investment by 6%. Cacciatore et al. (2017) study the impact product market reforms at the ZLB. Lincoln and McCallum (2018) and Maggi and Felix (2019) study the effects of entry costs for international trade. Edmond et al. (2019) decompose the welfare costs of markups into misallocation across firms, inefficient entry, and an equivalent uniform output tax.

Eggertsson et al. (2018), Corhay et al. (2018) and Kozeniauskas (2018) are perhaps the closest papers to our work. Eggertsson et al. (2018) take entry as exogenous and model a time-varying elasticity of substitution between intermediate goods to study the ability of time-varying market power to explain a number of broad macroeconomic trends. Corhay et al. (2018) develop an innovation-based endogenous growth model with aggregate risk premia and endogenous markups; and use it to decompose the rise in $Q$ into revised growth expectations, rising market power, and changes in risk premia. Corhay et al. (2018) conclude that declines in competition explain a large portion of the increase in $Q$. Albeit with a different structure, our model also features endogenous entry decisions sensitive to future demand expectations. In contrast to Corhay et al. (2018), however, we model industries separately. This allows us to identify the key parameters of the model and to connect entry costs to explicit measures of entry regulations. Kozeniauskas (2018) takes a different approach. He uses a general equilibrium model of occupational choice to study the contribution of four explanations to the decline in entrepreneurship: changes in wages driven by skill-biased technical change; changes in technology facilitating the expansion of large firms; changes in fixed costs (which combine sunk entry costs and per-period operating costs); and changes in demographics. In line with our findings, he concludes that increasing fixed costs are the main explanation for the decline in entrepreneurship.

Our paper is also related to a long literature in IO that studies the evolution of industries when entry costs are endogenous and may be influenced by incumbents. Stigler (1971) focuses on regulation and argues that “as a rule, regulation is acquired by the industry and is designed and operated primarily for its benefit.” Sutton (1991, 1997) studies how incumbents use marketing and R&D to increase entry costs – and therefore limit the number of firms in an industry. Djankov et al. (2002) document large
differences in entry costs across countries, and link them to levels of corruption.

Following Eggertsson and Woodford (2003), a large and growing literature studies the consequences of a binding ZLB on the nominal rate of interest. The ZLB has been proposed as an explanation for the slow recovery of most major economies following the financial crisis of 2008-2009 (Summers, 2013). Eggertsson et al. (2019) propose a model of secular stagnation, including a study of the role of demographic changes. Swanson and Williams (2014) study the impact on long rates. Most studies of the liquidity trap are based on simple New-Keynesian models that abstract from capital accumulation (see Fernández-Villaverde et al. 2015 for the exact properties of the New Keynesian model around the ZLB). Capital accumulation complicates matters, however, as consumption and investment can move in opposite directions.

Section 2 presents the relevant facts about the U.S. economy in recent decades. Section 3 presents our benchmark model. We start from a standard DSGE model in which we allow for the possibility that the ZLB constraint on short term nominal rates binds. Section 4 discusses how we solve our model and form its likelihood function. Section 5 verifies that our estimated entry cost shocks correlate with independently constructed measures of entry regulation and M&A; and Section 6 presents the aggregate implications of our model.

2 Four Facts About Entry, Concentration, Profits and Investment

We begin with four stylized facts that guide our analyses.4

Fact 1: Profits and Concentration have Increased. Figure 1 shows the ratio of Corporate Profits to Value Added for the U.S. Non-Financial Corporate sector, along with the cumulative weighted average change in 8-firm concentration ratio in manufacturing and non-manufacturing industries. As shown, both series increased after 2000. These patterns are pervasive across industries as shown by Grullon et al. (2016).

Fact 2: Entry Rates have Fallen. Figure 2 plots aggregate entry and exit rates from the Census BDS. Entry rates began to fall in the 1980s and accelerated after 2000. Exit rates have remained stable. This is true at the aggregate and industry-level, and remain when controlling for profits or Q as shown in Gutiérrez and Philippon (2018).

4See Appendix A.2 for additional details on the construction of these results.
Fact 3: Investment is Low Relative to Profits and $Q$. The top chart in Figure 3 shows the ratio of aggregate net investment and net repurchases to net operating surplus for the non financial corporate sector, from 1960 to 2015. As shown, investment as a share of operating surplus has fallen, while buybacks have risen. The bottom chart shows the residuals (by year and cumulative) of a regression of net investment on (lagged) $Q$ from 1990 to 2001, illustrating that investment has been low relative to $Q$ since the early 2000’s. By 2015, the cumulative under-investment is large at around 10% of capital. The decline appears across all asset types, notably including intangible assets (Covarrubias et al., 2019).

Fact 4: The Lack of Investment Comes from Concentrating Industries. Figure 4 shows that the capital gap is coming from concentrating industries. The solid (dotted) line plots the implied capital gap relative to $Q$ for the top (bottom) 10 concentrating industries. For each group, the capital gap is calculated based on the cumulative residuals of separate industry-level regressions of net industry investment from the BEA on our measure of (lagged) industry $Q$ from Compustat. This result highlights why it is critical to consider investment alongside concentration.

3 Model

To explain the drivers behind these facts, we use a model with capital accumulation, nominal rigidities, and time-varying competition with firm entry. We organize firms into industries and, for simplicity, separate them into capital producers who lend their capital stock, and good producers who hire capital and labor to produce goods and services.\footnote{This assumption simply allows us to maintain the standard $Q$-equation and the standard Phillips curve.} We use data at the industry-level on concentration and profitability to estimate the elasticity of firm entry to changes in $Q$. Those estimates are then used to understand the aggregate consequences of changes in entry costs.

Many of the features of our model are standard to the New Keynesian literature (see for example Smets and Wouters, 2007; Gali, 2008), and we focus on the new and non-standard additions to the basic framework, namely: (i) firm entry, (ii) estimated industry-specific shocks to expected demand, and (iii) monetary policy at the ZLB. The Appendix describes the remaining features of our model.
3.1 Firm Entry

Consider an industry indexed by \( j \) where goods-producing firms hire capital, labor, and intermediate goods for production, and make pricing decisions. Potential entrants pay an entry cost to become active producers in the subsequent period. Let \( N_{j,t} \) be the number of firms. The number of firms active at time \( t + 1 \) is

\[
N_{j,t+1} = (1 - \delta_n)N_{j,t} + n_{j,t}.  
\]  

(1)

Each active firm disappears with probability \( \delta_n \), while \( n_{j,t} \) is the number of entrants that become active in period \( t + 1 \). An exogenous exit rate is consistent with the data, as reported by Lee and Mukoyama (2018). Entry requires a fixed input \( \kappa_{j,t} \) produced competitively in industry with a convex cost function, so that the input price \( p_{j,t}^E \) is

\[
p_{j,t}^E = (\kappa_{j,t}n_{j,t})^\phi_n.  
\]  

(2)

Free entry then requires that

\[
p_{j,t}^E\kappa_{j,t} \geq \mathbb{E}_t \Lambda_{t+1}V_{j,t+1},  
\]  

(3)

where \( \Lambda_t \) is the household’s pricing kernel and \( V_{j,t} \) is the value of the goods-producing firm given by

\[
V_{j,t} = \text{Div}_{j,t} + (1 - \delta_n)\mathbb{E}_t \Lambda_{t+1}V_{j,t+1},  
\]  

(4)

where \( \text{Div}_{j,t} \) are real dividends. Equation (3) holds with equality as long as \( n_{j,t} > 0 \), which is the case in our simulations. Our assumption of convex entry costs slows entry during booms, which helps match the volatility of entry rates and their relationship to asset prices. This convexity can have multiple interpretations, from diminishing quality in managerial ability (Bergin et al., 2017) to congestion effects at firm creation (Jaef and Lopez, 2014) – perhaps due to a limited supply of Venture Capital needed to finance and monitor entrants (Loualiche, 2016). The entry cost \( \kappa_{j,t} \) is subject to industry-specific and aggregate shocks:

\[
\kappa_{j,t} = \kappa + \zeta_{j,t}^\kappa + \zeta_t^\kappa,  
\]  

(5)
where the industry and aggregate-level shocks are autoregressive processes

\[ \zeta^\kappa_{j,t} = \tilde{\rho}_\kappa \zeta^\kappa_{j,t-1} + \tilde{\sigma}_\kappa \epsilon_{j,t} \]  
\[ \zeta^\kappa_t = \rho_\kappa \zeta^\kappa_{t-1} + \sigma_\kappa \epsilon_{t} . \]  

(6)  
(7)

In this model entry costs regulate the link between entry of new firms and the market value of incumbents, therefore they capture not only technological costs, but also administrative costs and regulatory barriers, and deterrence by incumbents.

Finally, when we map our model to the data we take into account that Tobin’s \( Q \) reflects not only the usual capital adjustment costs but also monopolistic rents. An empirical feature of the data is that capital is not perfectly mobile across industries. Formally, we assume that there are industry-specific capital providers, so that an industry’s total \( Q \) combines the rents of goods-producers and capital-producers \( Q^k_{j,t} \), all measured at the end of the period

\[ Q_{j,t} = Q^k_{j,t} + (1 - \delta_n) \mathbb{E}_t [\Lambda_{t+1} V_{j,t+1}] / P_t K_{j,t+1} . \]  

(8)

The elasticity of the number of entrants \( n_{j,t} \) to \( Q_{j,t} \) depends on the parameter \( \phi_n \). The cross-industry relationships between concentration, profits, and output will be key to determining this sensitivity, which is important for quantifying the aggregate effects of entry shocks.

### 3.2 Industry-Specific Demand Shocks

**Demand System**  We use a standard nested CES demand system. The final good is a composite of industry-level outputs \( Y_{j,t} \) aggregated by a perfectly competitive final goods firm:

\[ Y_t = \left[ \int_0^1 (D_{j,t} Y_{j,t})^\frac{\sigma-1}{\sigma} d j \right]^\frac{-\sigma}{\sigma-1}, \]  

(9)

where \( \sigma \) is the elasticity of demand across industry-level goods, and \( D_{j,t} \) is an industry-level demand shifter described in more detail below. The price index for \( Y_t \) is \( P_t \), defined as \( P_t = \left( \int_0^1 P_{j,t}^{\alpha-\sigma} d j \right)^{\frac{1}{\alpha-\sigma}} \), where \( P_{j,t} \) is the price index of industry \( j \). As a rule, we define real variables (i.e., scaled by the GDP deflator \( P_t \)) unless there are nominal rigidities. So, for instance, \( R^k_{j,t} \) is the real rental rate of capital in industry \( j \), while \( W_t \) is the nominal wage and \( p_{i,j,t} \) is the nominal price set by firm \( i \) in industry.
Under (9), the demand curve faced by firms in an industry is

\[ Y_{j,t} = D_{j,t} \left( \frac{P_{j,t}}{P_t} \right)^{-\sigma} Y_t. \]  

(10)

Firms’ output (indexed by \(i\)) is aggregated into the industry output

\[ Y_{j,t} = \left( \int_0^{N_{j,t}} \frac{y_{i,j,t}^{\epsilon_j-1}}{y_{i,j,t}^{\epsilon_j}} \, di \right)^{\frac{\epsilon_j}{\epsilon_j-1}} = y_{j,t} \left( N_{j,t} \right)^{\frac{\epsilon_j}{\epsilon_j-1}}, \]

where, with some abuse of notation, we denote by \(y_{j,t}\) the average firm output in industry \(j\), \(N_{j,t}\) is the number of active (producing) firms in industry \(j\) at time period \(t\) and \(\epsilon_j\) is the elasticity of substitution across firms within the industry. We see here the impact of product variety on productivity. The industry price index is an aggregate of firm level price choices:

\[ P_{j,t} = \left( \int_0^{N_{j,t}} \frac{1}{p_{i,j,t}^{\epsilon_j}} \, di \right)^{\frac{1}{\epsilon_j}}. \]

(12)

The term \(D_{j,t}\) is an industry-level demand shifter, with \(d_{j,t} = \log D_{j,t}\) following the process:

\[ d_{j,t} = (1 - \tilde{\rho}_d)d_{j} + \tilde{\rho}_d d_{j,t-1} + \tilde{\sigma}_d \epsilon_{j,t}^d. \]  

(13)

We estimate transitory shocks to \(d_{j,t}\), or \(\epsilon_{j,t}^d\), to help account for variation in relative industry output over time.

**Beliefs and Expected Demand** We also estimate shocks to beliefs about \(d_j\), the “average” value of demand, which can differ from fundamentals. In our baseline specification, we will estimate changes to steady-state beliefs between 1995Q1 and 1999Q4. This specification is chosen to account for excess entry and valuations observed in a number of industries before 2000, and is a useful instrument for identifying the responsiveness of firm entry to profitability, as parameterized by \(\phi_n\). The presence of noisy entry is documented in several studies. Doms (2004), for example, looks at IT investment and firm entry during the 1990s. He concludes that a “reason for the high growth rates in IT investment was that expectations were too high, especially in two sectors of the economy, telecommunications services and the dot-com sector,” where dot-com covers a wide range of traditional sectors, from retail trade to business services.\(^6\)

\(^6\)See also Hogendorn (2011).
One explanation for noisy entry is that there is variation in the willingness of investors (venture capitalists, or market participants in general) to fund risky ventures. This is particularly important during the 1990s given the large inflows into Venture Capital (VC).\footnote{According to the National Venture Capital Association, annual VC commitments surged during the bubble period, growing from about $10 billion in 1995 to more than $100 billion in 2000. They then receded to about $30 billion per year for the next decade (NVCA (2010)). According to Gompers and Lerner (2001), about 60 percent of VC funding in 1999 went to information technology industries, especially communications and networking, software, and information services. About 10 percent went into life sciences and medical companies, and the rest is spread over all other types of companies.} Clearly, not all entry is funded by VC firms so this can only explain a portion of the variation in entry rates, but the wide dispersion and strong industry focus highlights the differential impact of the dot-com bubble across industries. Another explanation is the presence of large stock market variations across industries, as documented by Anderson et al. (2010). These extreme valuations may translate into excess investment and excess entry, especially because firm entry increases precisely during periods of high-growth such as the late 1990’s (Asturias et al., 2017).

3.3 Monetary Policy at the ZLB

To close the model, we specify a policy rule for the central bank, taking into account the ZLB on nominal interest rates. We assume that monetary policy follows a standard Taylor rule for the nominal interest rate

$$\tilde{r}_t^* = -\log (\beta) + \phi_i \tilde{r}_{t-1}^* + (1 - \phi_i) (\phi_p \pi^p_t + \phi_y \ln Y_t - \ln Y^F_t) + \phi_y \ln \left( \frac{Y_t/Y^F_t}{Y^F_{t-1}/Y_{t-1}} \right) + \phi_y \ln \left( \frac{Y_t/Y^F_t}{Y^F_{t-1}/Y_{t-1}} \right) + \sigma_i e^i_t,$$

(14)

where $\pi^p_t$ is price-level inflation, $Y^F_t$ is the flexible price level of output, $e^i_t$ is a monetary policy shock, and the actual (log) short rate is constrained by the ZLB

$$\tilde{r}_t = \max (0; \tilde{r}_t^*).$$

(15)

At the ZLB, we allow for forward guidance as an extension of the ZLB duration beyond that implied by fundamentals and the shocks. That is, we allow, but do not impose, that the policy rate be extended beyond the duration implied by shocks, in line with the optimal policy prescription of Eggertsson and Woodford (2003). We discipline the expected lower bound durations with empirical measures, as discussed in the estimation section.
3.4 Remaining Model Elements

The rest of the model is standard and the equations are in the Appendix. The model features capital producers in each industry who accumulate capital subject to convex adjustment costs to maximize market value. Their problem gives rise to the $Q$-theory of investment, with net investment rising in line with Tobin’s $Q$ – the market value of the firm relative to the size of the capital stock. Goods-producers in each industry operate a Cobb-Douglas production function and face a price-setting problem under Calvo nominal rigidities, giving rise to industry-specific Phillips curves. The household sector is standard, with workers belonging to unions that face Calvo-style nominal wage rigidities.

In addition to industry-specific shocks to demand beliefs, we also model, at the industry-level, transitory shocks to demand, productivity, the valuation of corporate assets, the inflation equation, and entry costs. At the aggregate-level, we model shocks to productivity, the valuation of corporate assets, the Phillips curve, the household’s discount factor, the monetary policy rule, and entry costs. The rich set of shocks will help account for the industry and aggregate data.

4 Estimation

We next discuss the parameterization of the model for the quantitative analysis. We first calibrate a set of parameters to those commonly used in the literature and to moments in the data. We then estimate with Bayesian methods a small set of key structural parameters, beliefs about demand, the persistence and size of transitory shocks, as well as the parameters of the monetary policy rule.

The estimation is conducted in two stages. In the first stage, the industry-level data is used to estimate $\sigma$ and $\phi_n$, along with the parameters of the industry-level shock processes. In the second stage, the estimated value of $\phi_n$ is used in an aggregated version of the model with a single sector, and the parameters of the aggregate-level shock processes and the monetary policy rule are estimated. We then use the estimated aggregate model to conduct our aggregate experiments on the role of barriers to entry.

4.1 Calibrated Parameters

Table 1 presents the assigned and calibrated parameters for our quarterly model. These estimates are based on 43 industries that cover the US Business sector.\footnote{Investment and output data are available for 63 granular industry groupings from the BEA. We omit 7 industries in the Finance, Insurance and Real Estate sectors; as well as the ‘Management of companies and enterprises’ industry} We set $\delta_n$, the exogenous firm exit rate, to
0.09/4 to match the average annual exit rate of Compustat firms.\textsuperscript{9} We calibrate the quarterly capital adjustment cost $\phi_k$ to a value of 20, in line with a regression across industries of net investment on Tobin’s $Q$, with a full set of time and industry fixed effects. Next, we calibrate the within-industry, across-firm elasticities of substitution parameters, $\epsilon_j$, to match the gross operating surplus to output ratio in 1993 from the BEA industry series.\textsuperscript{10} Most values are centered around 5, the standard calibration in New Keynesian models, while some industries have higher elasticities of substitution, which reflect relatively low profit levels in those industries.\textsuperscript{11} We follow Edmond et al. (2019) and set $\rho$, the elasticity of substitution between labor and intermediate goods, to 0.5.

In the second stage of the estimation with a single intermediate-goods sector, we set the elasticity of substitution across varieties to $\epsilon = 5$, which is around the average of our calibrated $\epsilon_j$ parameters across industries, and which is in line with a standard calibration of the elasticity of substitution in the New Keynesian literature, implying a steady-state markup of 25%.

\subsection*{4.2 Estimated Parameters}

With industry-level data, we estimate $\sigma$, $\phi_n$, and the persistence and variance of the industry-level shocks. We also estimate the beliefs $d_j$ about average demand in industry $j$. We normalized the “true” values to $d_j = 0$, but we allow agents to have different beliefs during the internet bubble, from 1995 to 1999. With aggregate-level data, we estimate the parameters of the monetary policy rule, and the persistence and variance of aggregate shocks.

\subsection*{4.3 Data}

We use cross-sectional data in our identification, based on heterogeneity across industries (and firms). At the industry level, we use annual data on concentration ratios, $Q$, nominal output, capital, and prices, from 1989 to 2015. The Appendix provides a complete description of the data and definitions.

- We measure the concentration ratio as the share of sales by the top 8 firms in each BEA industry using Compustat. To account for time-varying coverage in Compustat as well as foreign

\textsuperscript{9} Because no data is available in Compustat for it. We then group some of the remaining industries due to missing data at the most granular-level (Hospitals and Nursing and residential care facilities), or to ensure that all groupings have material investment; good Compustat coverage; and reasonably stable investment and concentration time series.

\textsuperscript{10} We use Compustat firms to focus on the exit of large firms.

\textsuperscript{11} Given that the model’s implied steady-state gross operating surplus to output ratio changes as $\sigma$ changes, we recalibrate the values of $\epsilon_j$ accordingly. Figure A.5 in the Appendix plots the distribution of $\epsilon_j$ across industries in our baseline calibration.

\textsuperscript{12} For most industries, the gross operating surplus to output ratios are stable and do not change much over time, as we show in the Appendix.
competition, we adjust for total industry-level gross output using the BEA GDP by Industry accounts, and for imports using the data of Pierce and Schott (2012).\textsuperscript{12} In the model, all firms are identical, so that the Herfindahl index is \( \int (y_{i,j,t}/y_{j,t})^2 \, di = \int (y_{i,j,t}/N_{j,t}y_{j,t})^2 \, di = 1/N_{j,t} \) and the top \( k \) firms have a share of \( k/N_{j,t} \).

- We measure industry \( Q \) as the ratio of market value to total assets across all firms in Compustat that belong to a given BEA industry.\textsuperscript{13} In our baseline, we match the observed values of \( Q \) to the \( Q \) of goods-producers.\textsuperscript{14}

- We measure nominal output and prices at the industry-level using the BEA’s GDP-By-Industry accounts, and investment and capital stocks at the industry-level using the BEA’s Fixed Assets Tables.

At the aggregate level, our data is quarterly from 1989Q1 to 2015Q1, and includes the Fed Funds rate, the change in real consumption per capita, the net investment rate, inflation, and employment. We also link observed changes in the aggregate concentration ratio to changes in the model’s aggregate Herfindahl index. To discipline the expected durations of the ZLB between 2009Q1 and 2015Q1, we use data from the New York Federal Reserve Survey of Primary Dealers, following Kulish et al. (2017).

### 4.4 Solution Method

The first challenge in our estimation is to model time-varying beliefs about the steady-state of industry-level demand that differs from fundamentals, for each industry. The second challenge is to simultaneously account for the nonlinearities caused by the ZLB. We discuss each challenge in this section. The next section discusses how we use our approximation below to form the likelihood function for estimation.

**Solution Method for Subjective Beliefs.** We first discuss the linearized solution under time-varying beliefs for a single industry denoted by \( j \), abstracting from industry \( j \)'s dependence on aggregate

\textsuperscript{12}We use Compustat as opposed to Census concentration because it is available over a longer period under a consistent segmentation. One downside of using Compustat is that all of the activity of multi-industry firms is mapped to each firm’s primary industry, as opposed to the corresponding industries. This is not an issue for Census concentration measures, where each establishment is mapped to the corresponding industry. To validate our use of Compustat, we validate that Compustat and Census concentration measures exhibit similar behavior in unreported tests: they have 65 to 70\% correlation in levels and 40 to 50\% in 5-year changes.

\textsuperscript{13}Gutiérrez and Philippon (2017) compare alternate measures of \( Q \) used in the literature and conclude that market-to-book is the most robust and stable definition

\textsuperscript{14}We explore robustness to matching \( Q \) in the data to the aggregate sector level \( Q \), which combines both the rents of the goods-producers and the capital-producers. We discuss these robustness exercises below.
variables. We follow Kulish and Gibbs (2017) in specifying two regimes: one under the industry’s “true”
parameters, and another where beliefs about demand differ from the truth. We use the adjective “true”
to simplify the discussion but this is of course a semantic distinction. Our model is entirely consistent
with agents receiving unobserved noisy signals about future demand. Under this interpretation the
late 1990s was simply a period of high optimism. Denote the true demand regime which is driving the
observables as
\[ \mathbf{A} \mathbf{x}_t^j = \mathbf{C} + \mathbf{B} \mathbf{x}_{t-1}^j + \mathbf{D} \mathbf{E}_t \mathbf{x}_{t+1}^j + \mathbf{F} \mathbf{e}_t^j, \]  
(16)
where \( \mathbf{x}_t^j \) is the vector of state variables for an industry \( j \) and \( \mathbf{e}_t^j \) collects the shocks for industry \( j \). Agents in the model, however, believe in an alternative demand regime. They think that the industry’s
law of motion follows the * matrices
\[ \mathbf{A}^* \mathbf{x}_t^j = \mathbf{C}^* + \mathbf{B}^* \mathbf{x}_{t-1}^j + \mathbf{D}^* \mathbf{E}_t \mathbf{x}_{t+1}^j + \mathbf{F}^* \mathbf{e}_t^j. \]  
(17)
Our goal is to construct the reduced-form VAR approximation for industry \( j \) of the form
\[ \mathbf{x}_t^j = \mathbf{J}_t + \mathbf{Q}_t \mathbf{x}_{t-1}^j + \mathbf{G}_t \mathbf{e}_t^j. \]  
(18)
In periods when beliefs \( \mathbf{E}_t \mathbf{x}_{t+1}^j \) accord with regime (16), the solution is the standard time-invariant solution
\[ \mathbf{x}_t^j = \mathbf{J} + \mathbf{Q} \mathbf{x}_{t-1}^j + \mathbf{G} \mathbf{e}_t^j. \]  
(19)
Instead, in periods when beliefs \( \mathbf{E}_t \mathbf{x}_{t+1}^j \) are formed with (17) then \( \mathbf{E}_t \mathbf{x}_{t+1}^j = \mathbf{J}^* + \mathbf{Q}^* \mathbf{x}_t^j \), where \( \mathbf{J}^* \) and \( \mathbf{Q}^* \) are the matrices of the reduced form solution corresponding to the system (17). Substituting these beliefs \( \mathbf{E}_t \mathbf{x}_{t+1}^j \) into (16) and rearranging gives
\[ \mathbf{\tilde{Q}} = [\mathbf{A} - \mathbf{D} \mathbf{Q}^*]^{-1} \mathbf{B} \]  
(20)
\[ \mathbf{\tilde{G}} = [\mathbf{A} - \mathbf{D} \mathbf{Q}^*]^{-1} \mathbf{F} \]  
(21)
\[ \mathbf{\tilde{J}} = [\mathbf{A} - \mathbf{D} \mathbf{Q}^*]^{-1} [\mathbf{C} + \mathbf{D} \mathbf{J}^*] \]  
(22)
For our time-varying representation (18), we therefore set \( \mathbf{Q}_t = \mathbf{Q} \), \( \mathbf{G}_t = \mathbf{G} \), and \( \mathbf{J}_t = \mathbf{J} \) in periods when beliefs align with the truth, and \( \mathbf{Q}_t = \mathbf{\tilde{Q}} \), \( \mathbf{G}_t = \mathbf{\tilde{G}} \), and \( \mathbf{J}_t = \mathbf{\tilde{J}} \) in periods when beliefs differ from the truth.
Zero Lower Bound and Forward Guidance. Our second computational challenge is to approximate the dynamics of our model where the policy rate is subject to the ZLB. We do so following the approach of Guerrieri and Iacoviello (2015) and Jones (2017). The logic of the solution is similar to the time-varying approximation (18) that we use for estimating demand beliefs. We define two additional regimes, one for when the ZLB does not bind, and one for when the ZLB binds. At each point in time the ZLB is observed, we assume that agents believe no shocks will occur in the future and iterate backwards through our model’s equilibrium conditions from the date that the ZLB is conjectured to stop binding. We then iterate on the periods that the interest rate is conjectured to be in effect until it converges, after which the solution is that in (18).

4.5 The Likelihood Function

The direct approach to estimating the parameters of the model and demand beliefs would be to form the likelihood function using the solution (18) and the industry and aggregate data together. However, the nonlinearities induced by the ZLB together with the large number of industries makes this approach computationally infeasible. As a result, we follow Jones, Midrigan and Philippon (2018) and construct the likelihood function differently and exploit the relative variation across industry outcomes for identification. This approach allows us to separate the likelihood into an industry-level component and an aggregate component and conduct the estimation in two stages, which we describe here.

Let $x^j_t$ denote the vector of variables for each industry $j$, expressed in log-deviations from the steady state. Under a piece-wise linear approximation and an assumption that aggregate shocks propagate to each industry in the same way we can write the evolution of $x^j_t$ as the sum of two components:

$$x^j_t = J + Qx^j_{t-1} + G\epsilon^j_t + J^a_t + Q^a_t x^*_t - 1 + G^a_t \epsilon^*_t.$$  \hspace{1cm} (23)

Here, the first set of matrices, $J$, $Q$ and $G$, account for how an industry’s variables depend on its own state variables and industry-specific shocks $\epsilon^j_t$, while the vector $x^*_t$ collects the aggregate variables and $\epsilon^*_t$ collects the aggregate shocks, and the matrices $J^a_t$, $Q^a_t$ and $G^a_t$ express how the industries’ variables depend on the aggregate variables, with the aggregate variables evolving according to:

$$x^*_t = J^* + Q^*_t x^*_{t-1} + G^*_t \epsilon^*_t.$$  \hspace{1cm} (24)
The matrices multiplying the aggregate variables and shocks are time-varying because of the nonlinearities caused by the ZLB. In contrast, the matrix of coefficients $J$, $Q$ and $G$ multiplying the industry-level variables is time-invariant.

Intuitively, under \((23)\), for an industry $j$, aggregate shocks and the ZLB do not change the response of firms in that industry to its own history of idiosyncratic shocks. Under this, letting $x_t = \int x_t^j \text{d}j$ denote the economy-wide average of the industry-level variables, the deviation of industry-level variables from their economy-wide averages,
\[
\hat{x}_t^j = x_t^j - x_t,
\]
(25)
can be written as a time-invariant function of industry-level variables alone:
\[
\hat{x}_t^j = J + Q\hat{x}_{t-1}^j + G\epsilon_t^j,
\]
(26)
where we use the assumption $\int \epsilon_t^j \text{d}j = 0$, that industry-level shocks have zero mean in the aggregate. We then use the representation in (24) and (26) to estimate the model using industry-level and aggregate U.S. data, separately. Because the industry-level outcomes are independent of each other, the likelihood contribution of industry-level data as a whole is the sum of each industry’s likelihood contribution. We then use standard Bayesian methods to characterize the posterior distribution of the model’s parameters.

To ensure that the industry data is consistent with (25), we express the industry-level data series relative to their respective aggregate series, by subtracting a full set of time effects, one for each year and each variable. We also subtract an industry-specific fixed effect. The resulting series for each industry are plotted in the Appendix.

4.6 Estimates

Table 2 presents moments of the prior and posterior distributions of the estimated parameters, for both the industry and aggregate-level parameters.

Industry Estimates. Panel A shows the estimates of $\sigma$ and $\phi_n$, and Panel B shows moments of the posterior distributions of the persistence and standard errors of the industry specific shocks.\(^{15}\) We choose wide priors for the parameters. The elasticity across industry-level goods, $\sigma$, is estimated to be

\(^{15}\)We compute two independent chains of length 500,000 and discard the first 20% of each chain. The convergence of the posterior distributions is analyzed in the Appendix.
around 0.4, which is reasonable for broad classes of goods, and consistent with the trade literature. The value of $\phi_n$ is around 1.6, with a 10th and 90th percentile of 1.2 and 2.1, respectively. The implications of these estimates for the speed of firm entry are discussed in the next section. The persistence of entry shocks is low, being centered around 0.1. The persistence of demand shocks is high, close to 0.99, while the persistence of the technology shocks is around 0.95.

For the demand beliefs, we choose a diffuse prior for each estimated $D_j$ which is an equal mixture of an inverse gamma distribution with a mode of 1 and a uniform distribution between 0 and 100. This prior has a 10th percentile around 1, a mode slightly above 1, and a 90th percentile around 80. We assume that from 2000 on, beliefs revert and align with their true assumed values. Most estimates of $D_j = \log d_j$ are around their fundamental value of $d_j = 0$. Some industries are estimated to have low expected demand; for example, rail transport has an estimated expected steady-state demand of around -0.4. In addition, some durable manufacturing industries are estimated to have low beliefs about demand, such as durable non-metal, durable primary metals, and durable miscellaneous manufacturing, which have estimated steady-state demands of around -0.15. Agriculture, waste management, and healthcare-related industries are also estimated to have comparatively low beliefs about steady-state demand between 1995 and 2000.

By contrast, technology industries are estimated to have high beliefs about their steady-state levels of demand between 1995Q1 and 1999Q4, in line with the dot-com exuberance before 2000. The information data industry is estimated to have beliefs about demand of around 1.4, significantly higher than its true value of 0. This industry includes IBM, Google and Facebook. Durable computing and professional and administrative services (which includes computer systems design and related services) also exhibit high estimated beliefs about demand. The former includes Dell, Hewlett-Packard, and Motorola; while the latter includes large IT service providers such as NEC Corp, Fujitsu and Accenture. We explore the implications of these estimated demand parameters in the next section.

**Aggregate Estimates.** The estimates of the monetary policy rule are presented in Panel A of Table 2. The values of the coefficients are similar in magnitude to those estimated in other studies (see for example Justiniano et al., 2010). Panel C of Table 2 presents estimates of the persistence and size of the aggregate shock processes. To interpret these, we show the unconditional forecast error variance decompositions of a set of aggregate variables in Table 3. We find that the aggregate shocks to entry

---

The full estimates of beliefs about demand that are held between 1995 and 2000 are shown in Figure A.3 in the Appendix.
costs, TFP and to the valuation of corporate assets – risk premia shocks – are key drivers of aggregate variables. In reduced-form, the shock to the valuation of corporate assets has similar implications as the marginal efficiency of investment shocks that are found to be key drivers of business cycles in Justiniano et al. (2010). As discussed in that paper, shocks to the marginal efficiency of investment, in the presence of frictions that drive an endogenous wedge between the marginal product of labor and the marginal rate of substitution, are able to generate the comovement of hours and consumption observed in the data.

Aggregate entry cost shocks are found to explain a significant amount of the variation of hours (about 19%), the natural rate (about 55%), and most of the Herfindahl index (about 97%) at business cycle frequencies. Unconditionally, the Herfindahl and the number of firms in the economy in our model is largely explained by technology shocks (46%), entry cost shocks (46%), and risk premia shocks (7%). We also find that entry cost shocks explain a large fraction of the variation in hours (21%), aggregate output (13%), consumption (10%), and the natural rate (50%), at the infinite horizon. As shown in counterfactual simulations in the final section, during our sample period 1989 to 2015, we find an important role for firm entry cost shocks in explaining investment, consumption, and the natural interest rate. Intuitively, similar to technology shocks, entry cost shocks can generate the comovement between consumption, hours, and investment present in the data, as well as comovement between inflation and the Fed Funds rate, while the use of data on concentration is a powerful way to identify the shock in the data (Figure A.1 in the Appendix plots the aggregate-level impulse responses).

4.7 Industry Implications

We next examine the implications of our estimates for industry-level variables. We show that about 10% of the variation in relative industry concentration ratios between 1995 and 2000 can be explained by firm entry driven by firms’ estimated beliefs about an industry’s long-run demand. The remainder of the variation in relative industry concentration is largely accounted for by transitory entry-cost shocks. We also find that about 30% of the relative variation in the capital stock between 1995 and 2000 is accounted for by demand belief shocks, with the remainder of the variation mostly accounted for by risk premia shocks.

We first present the industry-level impulse responses to industry-level shocks. We focus on the average industry with an elasticity of substitution between firm-level goods of $\epsilon = 5$, when the remaining model parameters are set to the mode of their estimated posterior distributions. Figure 5a plots
the response of goods-producers’ $Q_t$, industry-level concentration, and real output following a one standard deviation transitory demand shock. Following the demand shock, industry-level real output rises, goods-producers’ profits increase, and new firms enter, lowering the Herfindahl and the level of concentration in the industry. Our estimate of $\phi_n = 1.6$ implies that entry into the industry is fairly gradual. The impulse response implies that, following a demand shock that raises goods-producers’ $Q_t$ to 10% above steady-state after one year, the number of firms increases by 1.4% after two years. This is consistent with the evidence in Gutiérrez and Philippon (2018). This illustrates one of the main identification issues in the literature: demand shocks create negative correlations between output and investment on the one hand, and concentration on the other even though concentration is not causing the changes in real activity.

Figure 5b plots the impulse response of industry-level observables following a one standard deviation shock to entry costs. Profits rise, entry drops and concentration increases. The Herfindahl index rises by about 1%, while industry-level real output falls by just over 0.1%. Figure 5c plots the impulse response to a one standard deviation productivity shock, which we find temporarily lowers prices and goods-producers’ $Q_t$, reduces firm entry, and raises real output. These responses to productivity shocks of course depend on the elasticity of substitution $\sigma$, which is less than one in our estimation across broad industry categories.

We next explore what the estimated shocks imply for the industry-level variables. In Figure 6 we focus on the information data industry. Panel A plots the path of the Herfindahl index used in estimation and the path under the estimated demand beliefs only (without stochastic shocks). Panel B plots the path of the capital stock used in estimation under demand beliefs only. As discussed earlier, beliefs about steady-state demand are estimated to be strongly positive for the information data industry. Under these optimistic beliefs about long-run demand, firms enter and the Herfindahl falls by about 5%, which is about half of the observed decline in the relative Herfindahl index between 1995 and 2000. Following the reversion in beliefs back to their true values in 2000, firms exit the industry and the Herfindahl increases. As shown in Panel B, about a quarter of the increase in the relative capital stock observed in the information data industry is accounted for by the shock to beliefs about steady-state demand.

For all industries, the expected demand shock explains about 10% of the change in firm concentration. Figure 7 shows the impact of expected demand shocks on concentration and investment. We switch off all other shocks and simulate industry dynamics with only expected demand shocks. Panel
A shows the predicted change in Herfindahl against the observed change in Herfindahl from 1995 to 2000. Across all industries, the slope of the regression line is 0.1 and the correlation is 0.4. For the capital stock, the slope is 0.3 and the correlation is 0.85. There are two take-aways from this figure. First, expected demand shocks are large and volatile in the late 1990s, and this gives us much more power to identify the model’s parameters than any time series evidence. Second, empirical models that do not take these shocks into account could be significantly biased.

5 Explaining Entry Cost Shocks

Our estimation results suggest that entry costs play an important role in explaining the behavior of key aggregate and industry-level variables. In this section, we seek to validate our estimates. We show that the entry cost shocks recovered from the model are correlated with empirical proxies of barriers to entry not used in the estimation – namely regulations and M&A activity.\(^\text{17}\)

The link between regulation and entry costs is the subject of a long literature. Davis (2017) discusses recent evolutions for the U.S. and Gutiérrez and Philippon (2018) study the relationship empirically. Easing of M&A restrictions – as documented by Kwoka (2015) – allow incumbents to consolidate and potentially increase barriers to entry. M&A may increase for other reasons, including demand shocks and technological change (Andrade et al., 2001) so we would not expect our estimates to explain all the time variation in M&A activities.

We use Thomson Reuters SDC to measure M&A activity and RegData 3.1 (Al-Ubaydli and McLaughlin, 2015) for regulations.\(^\text{18}\) RegData is a substantial improvement relative to simple page

\(^\text{17}\)It is worth noting that regulation and M&A affect more than entry costs: regulations likely increase fixed costs in addition to entry costs, while M&A likely affects competitive dynamics for incumbents as well as entrants. Our tests, therefore, seek to establish a relationship between these measures and entry costs, not to comprehensively study the impact of regulation or M&A on firm dynamics. More broadly, one might worry that the estimated entry costs are driven by technological change. Figure A.8 plots the average entry cost shock by industry against the industry’s intangible intensity. There does not appear to be a positive relationship between the two.

\(^\text{18}\)RegData relies on machine learning and natural language processing techniques to construct measures of regulatory stringency at the industry level. It counts the number of restrictive words or phrases such as ‘shall’, ‘must’ and ‘may not’ in each section of the Code of Federal Regulations and assigns them to industries. Goldschlag and Tabarrok (2018) provide several validation analyses for RegData, including comparisons of regulation indices to the size of regulatory agencies and the employment share of lawyers in each industry. They conclude that “the relative values of the regulatory stringency index capture well the differences in regulation over time, across industries, and across agencies.” We use log-changes in regulation throughout our analyses. Gutiérrez and Philippon (2018) suggest using absolute changes when considering a long history because regulation increased rapidly from a low initial level in the 1970s, which exaggerates log-changes early in the sample. Our sample period is more recent and log-changes appear well-behaved. Last, we focus on national regulations given the use of national data, even though State and Local government also have regulatory responsibilities. It is hard to summarize the scale of state and local government regulation or its growth over time, but anecdotal evidence suggests a similar increase. Occupational licensing, for example, increased from less than 5% in the 1950s to 29% in 2008 (Kleiner and Krueger, 2013) – in large part because of greater prevalence of licensing requirements at the State-level.
counts but, given the sheer scale of regulation, measuring regulatory stringency at the industry level is a challenging task. To control for measurement error, we complement RegData with measures of regulatory employment from the Census’ Occupational Employment Statistics in some of our tests. Figure 8 compares the median rise in regulation (across industries) with the decline in firm entry rates. Regulatory restrictions increased relatively slowly until 1995, when the growth accelerated. The timing of the increase in regulation is consistent with the hypothesis that regulation hurts entry, but the trends could also be explained by some common factor. We therefore look across industries, which experienced widely different paths of regulation. Figure 10, for example, shows that restrictions nearly tripled in Chemical and Non-metal Manufacturing (the two industries with the fastest increase), yet remained largely stable in Food Manufacturing and Mining (the two industries with the lowest increase).

To test whether regulation and M&A explain the rise in entry costs, we compare the model-implied structural shocks with our empirical measures of Regulation and M&A – as illustrated in Figure 10 for two industries where the entry shocks are large. The top panel shows the growth of regulatory restrictions in the Nondurable Paper Manufacturing industry. The model captures well the broad cyclical variations but not always the timing. For instance, the large increase in regulatory restrictions happens around 1997 but entry declines most strongly in 1999. Unfortunately the data does not allow us to build proxies for implementation lags. The bottom panel of figure 10 shows M&A activity in the Air Transportation industry. Once again the model captures well the broad cyclical variations but not always the timing. The pattern aligns with a controversial merger wave that included Delta-Northwest (2008), United-Continental (2010), Southwest-AirTran (2011) and American-US Airways (2014). As expected the M&A series seems to lead the model-implied shocks. The timing and persistence of the shocks is sometimes difficult to compare between the model and the data. For instance, a large merger can increase concentration for several years if there is no subsequent entry. In our model, this shows up as a persistent entry shock, but in the data it is a one time merger.

In figure 10 we selected two industries with large shocks to illustrate how we can compare the model and the data and to emphasize important caveats. Let us now turn to a more systematic analysis. Table 4 confirms the significant connection between our model-implied shocks and our direct empirical proxies across all industries. We regress $\zeta_{j,t}$ on measures of regulation and M&A at the industry-level. All regressions include industry and year fixed effects. Column 1 includes the full sample 1989-2015.

---

19 In particular, we consider changes in the number of employees in Legal and Compliance occupations (SOC codes 23-0000 and 13-1040, respectively), by industry. Data following the NAICS hierarchy is available only after 2002, which limits our sample – but we still find a robust relationship.
and uses the regulation indices from RegData. Columns 2 and 3 consider our alternate measure of regulation over the industry-years when both are available (after 2002). As shown, both measures are positively correlated with entry cost shocks. To control for measurement error, column 4 takes the average between both measures. As expected, the t-statistic in column 4 increases substantially.\footnote{In unreported tests, we instrument changes in the regulation index with changes in regulatory employment. We find consistent albeit somewhat noisy results given the short time-period.}

Columns 5 and 6 replicate column 1 and 4, adding a measure of M&A activity.\footnote{Figure A.2 in the Appendix plots the aggregate entry-cost shock $\zeta_t$ against measures of regulation and M&A activity.}

Overall, we find support for the key predictions of our model across industries and over time. Our stylized model cannot explain all the variation in the micro data but the results allow us to validate our approach and give us confidence for the macro simulations that follow.

6 Firm Entry and the Decline in Investment

In this section, we use our estimated model to study the macroeconomic consequences of entry costs. We focus in particular on investment, output and monetary policy. In our main counterfactual, we set entry costs to zero from 2003 onwards and we use the model to simulate the economy. Our findings suggest that entry cost shocks account for much of the increase in aggregate concentration and that they have large effects on aggregate investment, the natural interest rate, and the stance of monetary policy. In our counterfactual exercise, we find that absent entry cost shocks, the aggregate Herfindahl index would have been about 15 percentage points lower by 2015 and the capital stock would have been about 7% higher.

The first step in our approach is to obtain the smoothed shocks that generate the aggregate data.\footnote{For this experiment, we keep the Herfindahl fixed at its 2012Q1 level from 2012Q1 on. This ensures our Herfindahl series is consistent with the patterns observed in Census data (available only until 2012), and mitigates the issues with relative prices and weights during the financial crisis, as documented in Figure A.4 in the Appendix. We also show in the Appendix that our implied series for entry rates matches the decline in entry rates observed in Census data and documented by a number of papers discussed in the Introduction. Furthermore, to obtain the model’s estimated shocks, we use the sequence of expected ZLB durations that are used in the estimation.}

With those shocks, we construct counterfactual series by setting the entry cost shocks to zero from 2003Q1 on. The presence of the occasionally binding ZLB during 2009 to 2015 complicates the interpretation of this counterfactual because, in practice, the ZLB can be binding because of the shocks themselves (including the entry cost shock) or because of monetary policy actions taken during this period (which, in principle, could be in response to the effects of entry cost shocks). For this reason, in assessing the effect of entry cost shocks, we construct two comparable counterfactual series: (i) a counterfactual where we remove the contribution of stimulatory forward guidance during the ZLB period,
and (ii) a counterfactual with entry cost shocks set to zero from 2003 onwards and the nominal interest rate subject to the occasionally binding ZLB. Our approach to removing the contribution of forward guidance is to allow the ZLB durations to react endogenously to the identified shocks; the difference between these endogenous durations and the durations used in the estimation (and in deriving the shocks) quantifies the extent of lower-for-longer forward guidance at each point in time.

Figure 11 plots the simulated paths of the Herfindahl and the Fed Funds rate without entry cost shocks. Panel A plots the Herfindahl index in the data and in the counterfactual. There is substantially more entry in the counterfactual and the simulated Herfindahl is about 15 points lower without entry cost shocks by the end of the sample. Panel B shows that the Fed Funds rate would have lifted off by the second half of 2013 without entry costs. Panel B also shows the path of the Fed Funds rate if we take away the ZLB constraint. The ZLB seems to be a significant constraint on monetary policy, particularly around 2013 where we estimate that the Fed would have lowered the rate by almost 2 percentage points in 2013. This observation is consistent with estimates of the shadow interest rate (see Wu and Xia, 2016).

Next, we explore what our model predicts for investment and consumption. Panel A of Figure 12 plots the log of the capital stock, and panel B the log of consumption both in the data and in our two simulations. Without entry costs consumption and the capital stock would be almost 7.5% higher by 2015. We conclude that entry cost shocks have a significant effect on aggregate quantities and that modeling monetary policy during the ZLB is crucial to determine the aggregate effects of these shocks.  

Entry costs also have a large impact on labor income and on the labor share. In Panel A of Figure 13, we plot the filtered series for labor income against the counterfactual without entry shocks from 2003 onwards. By the end of 2015, labor income is almost 20 percent higher in the counterfactual. For the labor share, we construct the contribution of entry cost shocks by looking at the difference between our counterfactual without forward guidance and with all shocks, and our counterfactual without entry cost shocks and forward guidance. We plot this contribution in Panel B of Figure 13 against the filtered measure of the labor share implied by our data series. We find that entry cost shocks explain about half of the decline in the labor share.

The aggregate effect of the identified decline in firm entry can be separated into the direct channel

---

23Removing forward guidance causes the level of capital to fall by 0.6 percentage points by 2012 and by 1.1 percentage points by 2015. For consumption, removing forward guidance causes consumption to fall by 1.6 percentage points by 2012, and by 0.7 percentage points by 2015 as the forward guidance stimulus unwinds.
caused by fewer firms investing, and an indirect channel caused by the binding ZLB and the inability of monetary policy to accommodate further declines in the natural rate. To illustrate the interaction between entry cost shocks and a binding ZLB, we plot in Panel C of Figure 13 the natural interest rate implied by the data and the Kalman filter, against the natural interest rate computed in a simulation where entry cost shocks are removed. The simulation illustrates the important role that firm entry has had in explaining movements in the natural rate. Comparing the counterfactual against the data-implied natural rate, our estimates imply that the positive entry cost shocks caused the annualized natural rate to fall by about an additional 3 percentage points by the end of 2010.

7 Conclusions

Entry has decreased in the U.S. economy, and markets have become more concentrated. We find that entry costs shocks have played an important role and that they are related to entry regulations. The methodology we use in this paper, mixing a structural model with cross-sectional evidence, can usefully be applied to other contexts.
References


Jones, Callum, “Unanticipated Shocks and Forward Guidance at the Zero Lower Bound,” 2017. mimeo NYU.


Table 1: Assigned Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
<th>Source/Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$</td>
<td>2</td>
<td>Inverse labor supply elasticity</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.97^{1/4}</td>
<td>Discount factor</td>
<td></td>
</tr>
<tr>
<td>$\theta_p$</td>
<td>2/3</td>
<td>Price setting Calvo probability</td>
<td>Average price contract of 3Q</td>
</tr>
<tr>
<td>$\theta_p$</td>
<td>3/4</td>
<td>Wage setting Calvo probability</td>
<td>Average wage contract of 4Q</td>
</tr>
<tr>
<td>$\phi_k$</td>
<td>20</td>
<td>Capital adjustment cost</td>
<td>Industry regression of $x_{j,t}$ and $Q_{j,t}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>1/3</td>
<td>Capital share</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.025</td>
<td>Capital depreciation rate</td>
<td></td>
</tr>
<tr>
<td>$\delta_n$</td>
<td>0.09/4</td>
<td>Exogenous firm exit rate</td>
<td>Average annual % firm exit</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.5</td>
<td>EOS between $\ell_{i,j,t}$ and $m_{i,j,t}$</td>
<td>Edmond et al. (2019)</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0.6</td>
<td>Weight on labor in composite $h_{i,j,t}$</td>
<td></td>
</tr>
<tr>
<td>$\epsilon_j$</td>
<td>2.5 to 14.3</td>
<td>Industry substitution elasticity</td>
<td>GOS$_j$/Nominal Output$_j$ in 1993</td>
</tr>
</tbody>
</table>

Table 2: Estimated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dist</th>
<th>Median</th>
<th>10%</th>
<th>90%</th>
<th>Mode</th>
<th>Median</th>
<th>10%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Structural Parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>N</td>
<td>1.0</td>
<td>0.3</td>
<td>1.6</td>
<td>0.424</td>
<td>0.423</td>
<td>0.398</td>
<td>0.447</td>
</tr>
<tr>
<td>$\phi_{n}$</td>
<td>N</td>
<td>1.5</td>
<td>0.8</td>
<td>2.1</td>
<td>1.513</td>
<td>1.552</td>
<td>1.192</td>
<td>2.027</td>
</tr>
<tr>
<td>$\phi_{r}$</td>
<td>B</td>
<td>0.7</td>
<td>0.6</td>
<td>0.9</td>
<td>0.782</td>
<td>0.786</td>
<td>0.755</td>
<td>0.815</td>
</tr>
<tr>
<td>$\phi_{p}$</td>
<td>N</td>
<td>2.0</td>
<td>1.7</td>
<td>2.3</td>
<td>1.512</td>
<td>1.559</td>
<td>1.344</td>
<td>1.802</td>
</tr>
<tr>
<td>$\phi_{g}$</td>
<td>N</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.175</td>
<td>0.170</td>
<td>0.066</td>
<td>0.281</td>
</tr>
<tr>
<td>$\phi_{y}$</td>
<td>N</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.260</td>
<td>0.260</td>
<td>0.198</td>
<td>0.332</td>
</tr>
</tbody>
</table>

B. Industry Shock Processes

| $\tilde{\rho}_q$ | B    | 0.5    | 0.2 | 0.8 | 0.002 | 0.002  | 0.001 | 0.005 |
| $\tilde{\rho}_z$ | B    | 0.5    | 0.2 | 0.8 | 0.083 | 0.099  | 0.043 | 0.170 |
| $\tilde{\rho}_d$ | B    | 0.5    | 0.2 | 0.8 | 0.991 | 0.990  | 0.985 | 0.993 |
| $\tilde{\rho}_a$ | B    | 0.5    | 0.2 | 0.8 | 0.947 | 0.946  | 0.939 | 0.953 |
| $\tilde{\rho}_e$ | B    | 0.5    | 0.2 | 0.8 | 0.054 | 0.079  | 0.030 | 0.163 |

| 10 × $\tilde{\sigma}_q$ | IG   | 0.6    | 0.3 | 1.9 | 3.078 | 3.076  | 3.015 | 3.138 |
| 10 × $\tilde{\sigma}_z$ | IG   | 0.6    | 0.3 | 1.9 | 0.316 | 0.314  | 0.277 | 0.350 |
| 10 × $\tilde{\sigma}_d$ | IG   | 0.6    | 0.3 | 1.9 | 0.957 | 0.957  | 0.927 | 0.956 |
| 10 × $\tilde{\sigma}_a$ | IG   | 0.6    | 0.3 | 1.9 | 0.064 | 0.065  | 0.060 | 0.070 |
| 10 × $\tilde{\sigma}_e$ | IG   | 0.6    | 0.3 | 1.9 | 0.431 | 0.429  | 0.397 | 0.457 |

C. Aggregate Shock Processes

| $\rho_z$ | B    | 0.5    | 0.2 | 0.8 | 0.984 | 0.984  | 0.978 | 0.989 |
| $\rho_b$ | B    | 0.5    | 0.2 | 0.8 | 0.990 | 0.998  | 0.883 | 0.932 |
| $\rho_c$ | B    | 0.5    | 0.2 | 0.8 | 0.985 | 0.983  | 0.777 | 0.857 |
| $\rho_q$ | B    | 0.5    | 0.2 | 0.8 | 0.938 | 0.938  | 0.923 | 0.952 |
| $\rho_e$ | B    | 0.5    | 0.2 | 0.8 | 0.599 | 0.592  | 0.529 | 0.644 |
| 100 × $\sigma_z$ | IG   | 0.6    | 0.3 | 1.9 | 1.038 | 1.037  | 0.954 | 1.138 |
| 100 × $\sigma_b$ | IG   | 0.6    | 0.3 | 1.9 | 1.250 | 1.257  | 0.872 | 1.138 |
| 100 × $\sigma_c$ | IG   | 0.6    | 0.3 | 1.9 | 0.177 | 0.181  | 0.106 | 0.130 |
| 100 × $\sigma_q$ | IG   | 0.6    | 0.3 | 1.9 | 0.120 | 0.122  | 0.102 | 0.147 |
| 100 × $\sigma_i$ | IG   | 0.6    | 0.3 | 1.9 | 0.158 | 0.155  | 0.137 | 0.178 |
| 10 × $\sigma_e$ | IG   | 0.6    | 0.3 | 1.9 | 0.960 | 0.970  | 0.882 | 1.076 |
### Table 3: Variance Decomposition of Aggregate Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shock A. 8 Quarter Horizon</th>
<th>Technology</th>
<th>Preference</th>
<th>Markup</th>
<th>Risk Premia</th>
<th>Policy</th>
<th>Entry Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Funds Rate</td>
<td></td>
<td>0.4</td>
<td>14.1</td>
<td>34.3</td>
<td>22.3</td>
<td>28.1</td>
<td>0.8</td>
</tr>
<tr>
<td>Output</td>
<td></td>
<td>76.6</td>
<td>0.3</td>
<td>0.8</td>
<td>1.5</td>
<td>0.2</td>
<td>20.6</td>
</tr>
<tr>
<td>Consumption</td>
<td></td>
<td>70.0</td>
<td>7.1</td>
<td>5.6</td>
<td>3.0</td>
<td>1.0</td>
<td>13.2</td>
</tr>
<tr>
<td>Net Investment</td>
<td></td>
<td>44.6</td>
<td>10.5</td>
<td>1.7</td>
<td>33.6</td>
<td>1.8</td>
<td>7.8</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td>16.7</td>
<td>3.1</td>
<td>43.9</td>
<td>7.2</td>
<td>10.5</td>
<td>18.6</td>
</tr>
<tr>
<td>Inflation</td>
<td></td>
<td>1.8</td>
<td>16.8</td>
<td>15.4</td>
<td>32.8</td>
<td>26.0</td>
<td>7.1</td>
</tr>
<tr>
<td>Herfindahl</td>
<td></td>
<td>0.8</td>
<td>0.1</td>
<td>0.0</td>
<td>2.0</td>
<td>0.0</td>
<td>97.1</td>
</tr>
<tr>
<td>Natural Rate</td>
<td></td>
<td>0.2</td>
<td>18.8</td>
<td>0.0</td>
<td>25.6</td>
<td>0.0</td>
<td>55.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Shock B. Unconditional</th>
<th>Technology</th>
<th>Preference</th>
<th>Markup</th>
<th>Risk Premia</th>
<th>Policy</th>
<th>Entry Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fed Funds Rate</td>
<td></td>
<td>3.9</td>
<td>13.8</td>
<td>31.6</td>
<td>23.1</td>
<td>26.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Output</td>
<td></td>
<td>81.7</td>
<td>0.7</td>
<td>0.1</td>
<td>4.2</td>
<td>0.0</td>
<td>13.2</td>
</tr>
<tr>
<td>Consumption</td>
<td></td>
<td>83.9</td>
<td>1.5</td>
<td>0.7</td>
<td>3.4</td>
<td>0.1</td>
<td>10.4</td>
</tr>
<tr>
<td>Net Investment</td>
<td></td>
<td>54.6</td>
<td>7.4</td>
<td>1.2</td>
<td>26.2</td>
<td>1.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Employment</td>
<td></td>
<td>16.0</td>
<td>3.0</td>
<td>41.1</td>
<td>9.5</td>
<td>9.8</td>
<td>20.7</td>
</tr>
<tr>
<td>Inflation</td>
<td></td>
<td>1.7</td>
<td>17.3</td>
<td>15.1</td>
<td>34.6</td>
<td>24.6</td>
<td>6.6</td>
</tr>
<tr>
<td>Herfindahl</td>
<td></td>
<td>46.0</td>
<td>1.1</td>
<td>0.1</td>
<td>6.8</td>
<td>0.0</td>
<td>46.0</td>
</tr>
<tr>
<td>Natural Rate</td>
<td></td>
<td>4.2</td>
<td>19.0</td>
<td>0.0</td>
<td>27.3</td>
<td>0.0</td>
<td>49.6</td>
</tr>
</tbody>
</table>

### Table 4: Regression of Entry Cost Shocks vs. Regulation and M&A

This table reports regression results of industry-level entry cost shocks on measures of regulation and M&A activity. Measures of regulation are standardized to ensure comparability. Entry cost shocks estimated using the model. Regulation indices from RegData. Changes in regulatory employment based on the Census’ OES. M&A activity from Thomson Reuters SDC. Standard errors clustered at the industry-level in brackets. + p<0.10, * p<0.05, ** p<.01.

<table>
<thead>
<tr>
<th>ζj,t</th>
<th>(1) All</th>
<th>(2) Post-02</th>
<th>(3) Post-02</th>
<th>(4) Post-02</th>
<th>(5) All</th>
<th>(6) Post-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ log(Reg Indexj,t-2,t-1)</td>
<td>0.044**</td>
<td>0.047*</td>
<td>0.044**</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Δ log(Reg Empj,t,t+1)</td>
<td>0.031*</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean(L·RegIndex,F·RegEmp)</td>
<td>0.038**</td>
<td>0.033**</td>
<td>(0.009)</td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>log(M&amp;A,j,t)(2Y MA)</td>
<td>0.047*</td>
<td>0.087*</td>
<td>(0.021)</td>
<td>(0.037)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ind FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Year FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>R2</td>
<td>.051</td>
<td>.091</td>
<td>.085</td>
<td>.095</td>
<td>.057</td>
<td>.11</td>
</tr>
<tr>
<td>Observations</td>
<td>837</td>
<td>358</td>
<td>358</td>
<td>358</td>
<td>837</td>
<td>358</td>
</tr>
</tbody>
</table>
Figure 1: Concentration and Profits

Notes: Solid line plots ratio of After Tax Corporate Profits with inventory valuation adjustment and capital consumption adjustment to Value Added for the U.S. Non-Financial Corporate sector (series W328RC1A027NBEA and NCBGVAA027S, respectively). Annual data from the Financial Accounts of the United States, via FRED. Dotted lines show the cumulated sales-weighted average change in 8-firm Concentration Ratio (CR8). Data from the U.S. Economic Census based on SIC codes before 1992 and NAICS codes after 1997. We include only those industries that are consistently defined over each 5-year period. Change from 1992 to 1997 imputed from Autor et al. (2017a) given the change in industry classification from SIC to NAICS. When multiple tax groups are reported, only taxable firms are included. CR8 equals the market share (by sales) of the 8 largest firms in each industry.

Figure 2: Firm Entry and Exit Rates

Source: Census BDS
Figure 3: Net Investment, Profits and Q-Residuals

Net Investment and Net Buybacks to Net Operating Surplus (NFCB)

Annualized Prediction Residuals (by period and cumulative)

Notes: Quarterly data from the Financial Accounts of the United States, via FRED. Top plot shows the ratio of net investment and net buybacks to net operating surplus for U.S. Non Financial Corporate sector. Bottom plot shows the per-period and cumulative residuals of a regression of net investment for the U.S. Non Financial Business sector on $Q$ for Non Financial Corporate sector. We use the 1990 to 2001 period as a training sample and use the estimated coefficients to forecast out-of-sample after 2001. See Appendix A for additional details.
Figure 4: Cumulative Capital Gap for Concentrating and Non-Concentrating Industries

Notes: Annual data. Top plot shows the weighted average import adjusted 8-firm Concentration Ratio (CR8) for the 10 industries with the largest and smallest log-change in import-adjusted CR8 between 2000 and 2017. Bottom plot shows the cumulative implied capital gap (as a percent of capital stock) for the corresponding industries. See text for details.
Figure 5: Industry-Level Impulse Response Functions

(a) Industry-Level Impulse Responses to Demand Shock

(b) Industry-Level Impulse Responses to Entry-Cost Shock

(c) Industry-Level Impulse Responses to Productivity Shock
Figure 8: Regulation Index and Firm Entry Rate


Figure 9: Changes in Industry-level Regulatory Restrictions

Notes: Annual data from RegData. Figure plots industries with the two largest and smallest changes in restrictions.
Figure 10: Industry Entry Cost Shocks vs. Regulation and M&A

Notes: Annual data. Entry cost shocks estimated by the model. Regulation indices from RegData. M&A activity from Thomson Reuters SDC.
Figure 11: Aggregate Counterfactual, Entry and Monetary Policy

A. Herfindahl, Deviation

B. Fed Funds Rate
Figure 12: Log Capital and Consumption, Without ZLB and Without Entry Cost Shocks

A. Log Capital

B. Log Consumption
Figure 13: Labor Income, Labor Share, and Natural Interest Rate