The cumulative cost of regulations

Bentley Coffey [a], Patrick A. McLaughlin [b],[*], Pietro Peretto [c]

[a] *Department of Economics, University of South Carolina, United States of America*
[b] *Mercatus Center at George Mason University, United States of America*
[c] *Department of Economics, Duke University, United States of America*

A R T I C L E   I N F O

A B S T R A C T

We estimate the effects of federal regulation on the value added to GDP for a panel of 22 industries in the United States over a period of 35 years (1977–2012). The structure of our linear specification is explicitly derived from the closed-form solutions of a multisector Schumpeterian model of endogenous growth. We allow regulation to enter the specification in a fairly flexible manner. Our estimates of the model's parameters are then identified from covariation in some standard sector-specific data joined with RegData 2.2, which measures the incidence of regulations on industries based on a text analysis of federal regulatory code. With the model's parameters fitted to real data, we confidently conduct counterfactual experiments on alternative regulatory environments. Our results show that regulatory restrictions have had a net effect of dampening economic growth by approximately 0.8 percent per annum since 1980. Had regulation been held constant at levels observed in 1980, our model predicts that the economy would have been nearly 25 percent larger by 2012 (i.e., regulatory growth since 1980 cost GDP $4 trillion in 2012, or about $13,000 per capita).

1. Introduction

Regulations have long been the primary tool used by policymakers to advance their goals when free markets fail to deliver the outcomes that they desire. Even if we could thwart all attempts of rent-seeking forces to capture the regulatory process as described by Stigler (1971), skeptics have long suspected that the "law of unintended consequences" often undermines even otherwise virtuously conceived regulations in a complex and dynamic economy (Peltzman, 1975). Economists' efforts to gain an understanding of regulations' consequences, which can potentially be opposing in their effects on the economy's output, has laid the groundwork for a growing literature.

We add to this literature primarily by examining the cumulative effect of federal regulation on the United States economy, based upon the observation that the accumulation of an increasingly complex set of regulatory constraints is arguably the most striking characteristic of the history of regulations in the US. We model this cumulative effect in the context of endogenous growth. First, we build a formal model of economic growth from microeconomic foundations: growth is based on firms' decisions to invest in improving their productivity. By allowing our measure of regulations to flexibly enter into our econometric model, we rely on the data to tell us how much regulation distorts the investment decisions of firms and

thus hampers long-run economic growth. Hence, our model features enough complexity in detail and flexibility in specification to capture the aggregate effect of regulatory accumulation as a drag on the long-run macroeconomy. Simultaneously, our model accommodates a rich mix of short-run outcomes in output, (net) entry, and investment in particular industries, which can be spurred by some types of regulations.

We have constructed our theoretical model from a well-established model of endogenous growth (Peretto and Connolly, 2007).[1] This style of model has consistently received strong support in empirical investigations of competing models of economic growth (Ang and Madsen, 2011).[2] To conduct our analysis, we have extended the model from a single-sector economy to a truly multisector economy, where the growth process for each heterogeneous industry is governed by a set of linear equations that can be influenced by regulatory shocks. Ours is one of the first multisectoral endogenous growth model with closed-form solutions.[3]

We then estimate our model using industry data from the Bureau of Economic Analysis (BEA) and the US Census Bureau in combination with RegData 2.2, a publicly available text-based-quantification of regulation by industry over time from the RegData project (Al-Ubaydli and McLaughlin, 2015; McLaughlin and Sherouse, 2019). The available data permit us to estimate our model for a panel dataset of 22 industries observed annually from 1977 to 2012. Specifically, we examine not only the direct effect of regulation on the output per employer establishment and (net) entry of additional establishments but also the indirect effect of regulation on three different types of investment that in turn affect the output per establishment and (net) entry of additional establishments.

This combination of modeling and data enables us to estimate the effects of regulation on investment in an endogenous growth context. In endogenous growth theory, innovation is not an exogenous gift from the gods but rather the result of costly effort expended by firms to realize gains. The growth generated by that entrepreneurship can be stunted by misguided public policy. By deflecting firm resources away from the investments that maximize the stream of profits and toward regulatory compliance, regulations can theoretically slow the real growth of an industry. Indeed, each additional (binding) regulation should be seen as a (binding) constraint on the firm's profit maximization problem. Yet some regulations may yet succeed in enhancing growth. For instance, if a regulation does manage to efficiently correct a market failure, some deadweight loss in output will be recovered by that regulation. Alternatively, a regulation requiring the use of certain inputs like safety equipment can address a prisoner's dilemma in accumulating that capital which can enhance labor productivity. Naturally, these regulations can cause a select set of industries to grow faster, at least in the short run.[4]

These countervailing effects of regulation can now be handled in an empirical study because RegData provides separate observations for multiple sectors over a relatively long time frame. We show that the average industry does indeed experience a slower growth rate as a result of the accumulation of regulations that target the industry. However, a handful of industries do appear to grow faster because of regulation. Nonetheless, our results are dominated by the slower growth caused by regulations' distortion of firms' investment decisions. On net, cumulated regulations slow the growth of the entire economy by an average of 0.8 percent per annum. Because economic growth is an exponential process, this seemingly small figure grows powerfully over time into a truly dramatic difference in the level of GDP per capita. Had regulations been held constant at levels observed in 1980, our model predicts that the economy would be nearly 25 percent larger. In other words, the growth of regulation since 1980 cost the United States roughly $4 trillion in GDP (nearly $13,000 per person) in 2012 alone.

Of course, our estimate says little about the benefits of regulation, aside from those that are captured within GDP. Some regulations may lead to benefits, such as improvements in environmental quality, which are well known to be mainly missing from GDP measurements. Other regulations, such as those designed to prevent monopolistic practices, or even those designed to improve human health, may only be captured in GDP to a limited degree. For example, if regulations decrease employee absenteeism because they reduce asthma-inducing air pollution, we would expect that positive health effect to register as a marginal increase in GDP. Nonetheless, we caution that this study is not a weighing of the costs and benefits of regulation. It is an examination of how regulatory accumulation in specific sectors of the economy affects the growth path of those sectors, and a deduction from our findings of how the effects of regulation change the US economy's growth path.

The remainder of this study proceeds as follows. Section 2 expands on the background briefly described in this introduction. Section 3 describes how we measure regulation, as well as the other data used in our estimation. Our theoretical model is outlined in section 4. Section 5 details our estimation methodology. We explain our results in section 6. Section 7 concludes our discussion.

---

[1] Although it may be less apparent on the surface, the structure of the endogenous growth model in Peretto and Connolly (2007) is deeply connected to the structure in Peretto (2007), where the related model was built to study a more intimately related topic: the dynamic deadweight loss owing to corporate taxation. However, the structure in Peretto (2007) is less amendable to generalizing those closed-form solutions to a multisector structure.

[2] For more evidence, see Madsen (2008, 2010).

[3] For another approach at a closed-form solution to this type of model, see Cai and Li (2019).

[4] For example, the output of corn farmers arguably increased, at least over some period of time, in response to federal regulations requiring that gasoline contain a minimum percentage of ethanol, which is primarily derived from corn.

2. Background

Over the past two decades, multinational indexes, such as the World Bank's Doing Business project or the Organisation for Economic Co-operation and Development's (OECD's) Indicators of Product Market Regulation database have permitted first-generation estimates of the effect of regulation on economic growth, generally finding that macroeconomic growth can be considerably slowed by lower-quality regulatory regimes.[5] Djankov et al. (2006) use the World Bank's Doing Business index to examine a large panel of countries' regulations, finding that a country's improvement from the worst (first) to the best (fourth) quartile of business regulations leads to a 2.3 percentage point increase in annual GDP growth.[6]

While informative, these studies' reliance on broad regulatory indexes, such as the World Bank's Doing Business index and the OECD's Product Market Regulation database, necessitates tradeoffs. First, these indexes generally cover relatively short time spans, while changes in regulatory policy often require several years, if not decades, to implement. Second, they are not comprehensive. Instead, they focus on a few areas of regulation and then typically only on whether regulations exist in these areas, not how complex or burdensome they are. Finally, they often rely on the opinions of either the creators of the index (in the cases where the index incorporates only a few areas of regulation, the creators select the areas) or of "country experts" who rate how regulated a country is in questionnaires and surveys.

Subsequently, metrics based on actual regulatory text have permitted Dawson and Seater (2013) to use time series measures of US federal regulation in a single-country, macroeconomic growth setting. Using a before-and-after identification strategy on covariation between aggregate data and a page count of federal regulatory code, Dawson and Seater conclude that regulatory accumulation was responsible for slowed economic growth in the United States by an average of 2 percent per year between 1949 and 2005.

A key insight of endogenous growth models in general is that the effect of government intervention on economic growth is not simply the sum of static costs associated with individual interventions. One recent proposal regarding the phenomenon of regulatory accumulation reframed this insight with respect to regulation, pointing out that when regulations are created in reaction to major events, "new rules are [placed] on top of existing reporting, accounting, and underwriting requirements.... For each new regulation added to the existing pile, there is a greater possibility for interaction, for inefficient company resource allocation, and for reduced ability to invest in innovation. The negative effect on US industry of regulatory accumulation actually compounds on itself for every additional regulation added to the pile" (Mandel and Carew, 2013, p. 4).

For its part, the federal government requires regulatory agencies to perform regulatory impact analyses of significant new regulations. In theory, these analyses estimate the marginal impacts—positive or negative—that new regulations would precipitate. However, these regulatory impact analyses are performed for only a very small portion of new regulations, they rarely consider interactive or cumulative effects, and their overall quality falls far short of the best practices specified in the executive orders and the Office of Management and Budget guidance documents related to the process (Ellig and McLaughlin, 2011).

More rigorous studies of regulations sometimes originate in academia, where scholars tend to focus on the empirical effect of a particular regulation or particular regulations on a limited scope of industries. For example, researchers have given considerable attention to the employment effects on manufacturing industries of particular environmental regulations from the Clean Air Act.[7] Another case that has been heavily studied is deregulation of the transportation and telecommunications industries during the mid-1980s. Findings broadly concur that deregulation resulted in a significant surge in investment for the United States and the United Kingdom relative to Italy, France, and Germany (Alesina et al., 2005). However, compared to the myriad regulations that actually affect the economy, these and most other regulatory studies focus on interventions that are relatively limited in scope or on economic outcomes related to a narrowly defined sector. Such studies may not be able to see beyond the immediate negative (or positive) effects that regulation imposes on the decisions of firms within an industry.

Second, in addition to contributing to endogenous growth theory, this study is the first in this literature to use a novel panel dataset that measures sectoral regulation over time: RegData 2.2 (McLaughlin and Sherouse, 2019). This rich data

---

[5] For examples, see Djankov et al. (2002) and Botero et al. (2004).

[6] Several other studies use similar World Bank or OECD-produced panel data on regulations across countries. See Aghion et al. (2010); Loayza et al. (2005); Nicoletti and Scarpetta (2003); and Gørgens et al. (2003).

[7] See Greenstone (2002) or Morgenstern et al. (2002) for heavily cited examples or Walker (2013) for a more recent example.

source offers a multisector panel that quantifies federal regulation by industry from 1975 to 2014. RegData's metrics of regulation are created using text analysis programs run on the entire corpus of federal regulations in effect in each year. Its primary improvement over other datasets is its association of regulations with affected industries in an industry-specific data series spanning multiple decades. This industry-specific dataset allows us to use panel data methods to identify the effects of regulation more confidently than if we only had national data. By using these data, we also can produce industry-specific estimates of the effects of regulation on the key decisions of firms: investment, output, and (net) entry. We combine industry regulation data from RegData 2.2 with measures of industry inputs and outputs taken from other data sources to identify estimates for enough of the parameters in our model to conduct valid counterfactual experiments.

3. Data

Because RegData 2.2 is fairly new, we devote some time to explaining it here. The *Code of Federal Regulations* (CFR) is published annually and contains all regulations issued at the federal level. Among other things, RegData 2.2 searches the entirety of the CFR in each year from 1975 to 2014 for a vocabulary of industry-specific terms and phrases that were developed with machine-learning algorithms and that indicate that an industry is targeted by regulatory text. We use RegData 2.2's industry regulation index as our metric of regulation.[8]

RegData 2.2 measures two essential elements of regulatory text. First, it quantifies regulatory restrictions relevant to specific industries in the CFR. Restrictions are those words used in legal language to either obligate or prohibit an action (e.g., "shall").[9] Second, RegData estimates the relevance of restrictions to each industry in the economy. Version 2.2 of RegData relies on machine-learning algorithms to classify chunks of text in the CFR according to their relevance to specific industries, as defined by the North American Industry Classification System (NAICS). The RegData 2.2 program was trained to identify the relevance of regulatory text to specific industries using select documents from the *Federal Register*.[10]

Documents were analyzed using a vocabulary of 10,000 words learned from the training documents. RegData 2.2 uses Logit-based classifications that were first made available in RegData 2.1 and that have outperformed competing classification schemes.[11] This classification methodology yields a set of probability scores ranging from 0 to 1 for each CFR part—a legal division of text that typically houses a regulatory program. Since the CFR is published annually, RegData offers probability scores for each CFR part in each year from 1970 to 2014. A probability score reflects the probability that a given part is relevant to a given industry.[12]

Fig. 1 compares the growth of the stock of federal regulations, in aggregate as measured by RegData as well as the next-best alternative that has been used in the literature: counting pages in the CFR (Dawson and Seater, 2013). The lighter lines in Fig. 2 graph the time paths of regulation for each of the 22 industries in our data, which reveal considerable within and between variation from trend. The thick black line shows the industry mean.

We also gather some fairly standard data from the BEA and the Census Bureau. All three data sources organize industries according to the NAICS. We assemble data at the three-digit NAICS level. The BEA data combine certain industries (e.g., industries 113, 114, and 115), so we correspondingly combine regulation data from each industry to match the BEA's system.[13] Once the full dataset is assembled, it contains 22 industries with data running from 1977 to 2012. The industries are listed in Table 1. The investment variables from the BEA are investment in equipment, investment in structures, and investment in intellectual property. Investment in equipment consists primarily of expenditures on equipment and machinery with service lives of one year or more that are normally capitalized in business accounting. Investment in structures consists primarily of expenditures on construction of new buildings or additions to existing structures (including well drilling and exploration). Investment in intellectual property includes primarily expenditures on software used for production and on product and process development (research and development). Data from the Census Bureau included employer establishments and number of employees for each industry. We calculate people per establishment for each industry as national

---

[8] See McLaughlin and Sherouse (2019) for a more detailed description of the RegData 2.2 dataset. The industry regulation index in RegData 2.2 is originally proposed in Al-Ubaydli and McLaughlin (2015).

[9] RegData specifically searches for a subset of all restrictions, consisting of the strings, "shall," "must," "may not," "prohibited," and "required." While this subset is not comprehensive regarding all the ways in which a restriction can be created with legal language, it is probably representative of the restrictiveness of regulatory text (Al-Ubaydli and McLaughlin, 2015).

[10] The *Federal Register*, a daily publication of the federal government, includes rules, proposed rules, presidential documents, and a variety of notices of current or planned government activity. Some of these documents are specifically labeled with relevant NAICS codes, and the language used is similar to that of the CFR. Training documents for each three-digit NAICS industry were obtained by searching the *Federal Register* Application Programming Interface for an exact match for the word "NAICS" and the three-digit code and each four-, five-, and six-digit code it contains. Additionally, the exact names of the three-digit industries and their children industries were used to identify documents. These searches yielded approximately 24,500 documents associated with at least one NAICS three-digit industry. Industries with fewer than 5 positive training documents were excluded from analysis.

[11] RegData 2.1—the iteration previous to the version we use—explores several well-established methods of classification: Support Vector Machines (SVMs) with a linear kernel, Logistic Regression (Logit), Random Forests, and K-Nearest Neighbors. Cross-validation tests and other evaluation scores indicate that the Logit model outperforms all other models, albeit by a relatively narrow margin (McLaughlin and Sherouse, 2019).

[12] As in the first version of RegData, each part's probability score is then multiplied by the number of restrictions contained in the same part and then summed across all parts for each agency (Al-Ubaydli and McLaughlin, 2015).

[13] There were similar concerns that we had to address for translating between NAICS and Standard Industrial Classification codes. Also, the Census Bureau data that we gathered did not contain the full set of industries covered by the BEA's data for value added to GDP.
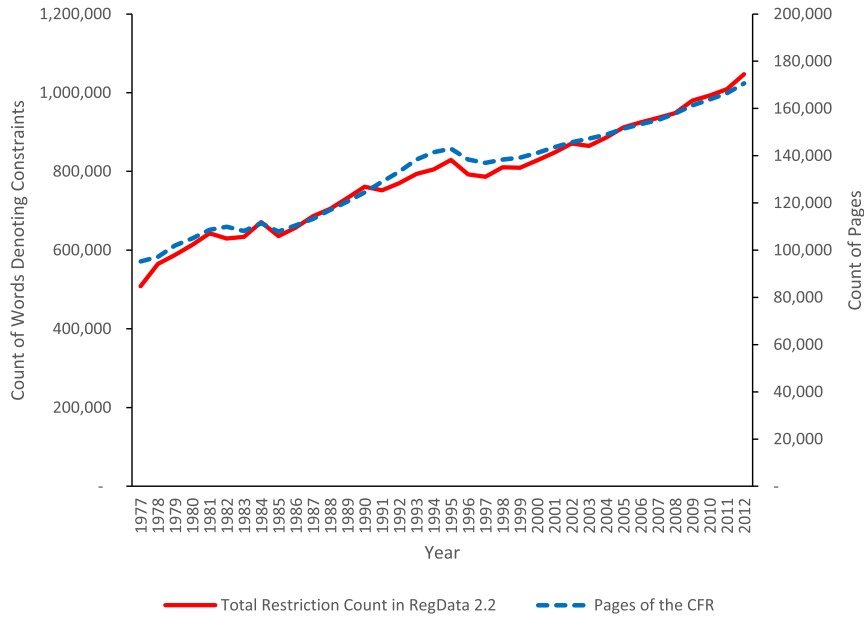
Note: CFR = *Code of Federal Regulations*.
Source: RegData 2.2 database, Mercatus Center at George Mason University, 2016, http://regdata.org/.

Fig. 1. Time paths of total pages in the code of federal regulations and total count of restrictions in RegData 2.2, 1977–2012.
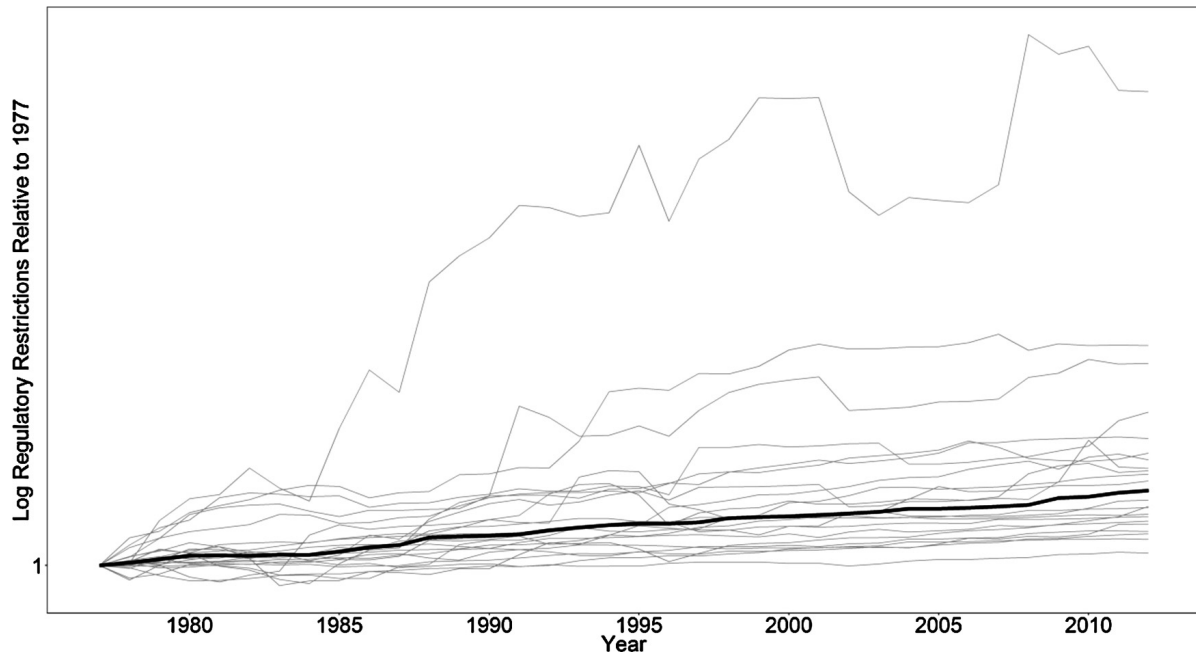
Source: RegData 2.2 database, Mercatus Center at George Mason University, 2016, http://regdata.org/.

Fig. 2. Time path of regulatory restrictions (in log scale) by industry, 1977–2012.

population divided by employer establishments. Likewise, real output per establishment is calculated as real value added to GDP divided by establishments for each industry.

Our assembled dataset is summarized in Table 2. All variables are reported in levels, except for our measure of regulation. As specified in our estimation equations, regulation is logged.[14]

---

[14] Given the panel nature of our dataset and the absence of a plausible instrument, we performed Granger causality testing focused on the relationship between regulation and each of the five different dependent variables that we ultimately use in our regressions: the three types of real investment (equipment, intellectual property, and structures), real value added to GDP per firm, and people per firm. The results are fairly inconclusive. For the average

Table 1
Identifiers for 22 industries examined in our study.

NAICS code	Full description (2012 NAICS code)	Abbreviated label
211	Oil and gas extraction	OIL_n_GAS
213	Support activities for mining	MINING_SUPPORT
322	Paper products manufacturing	PAPER
324	Petroleum and coal products manufacturing	PETROLEUM_n_COAL
325	Chemical products manufacturing	CHEMICALS
327	Nonmetallic mineral products manufacturing	NONMETALLICS
331	Primary metals manufacturing	PRIMARY_METALS
333	Machinery manufacturing	MACHINERY
334	Computer and electronic products manufacturing	COMPUTERS_n_ ELECTRONICS
335	Electrical equipment, appliances, and components manufacturing	ELECTRICAL_EQUIPMENT
339	Miscellaneous manufacturing	MISC_MANUFACTURING
481	Air transportation	AIR_TRANSPORT
483	Water transportation	WATER_TRANSPORT
484	Truck transportation	TRUCKING
486	Pipeline transportation	PIPELINE_TRANSPORT
493	Warehousing and storage	WAREHOUSING
523	Securities, commodity contracts, and investments	SECURITIES
524	Insurance carriers and related activities	INSURANCE
531	Real estate	REAL_ESTATE
562	Waste management and remediation services	WASTE
621	Ambulatory health care services	AMBULATORY CARE
713	Amusement, gambling, and recreation industries	AMUSEMENT_n_ RECREATION

Note: NAICS = North American Industry Classification System.

Table 2
Summary statistics.

Variable	N	T	Mean	Standard deviation	Min.	Max.
Real_Investment_in_Equipment_per_Establishment	22	34	0.0203	0.0278	0.0007	0.2471
Real_Investment_in_Structures_per_Establishment	22	34	0.0125	0.0315	0	0.2714
Real_Investment_in_IP_per_Establishment	22	34	0.0092	0.0178	0	0.1018
Real_Interest_Rate	1	34	0.0258	0.0225	−0.0075	0.0749
Real_Output-per_Establishment	22	34	151.9	162.3	10.6	1279.5
People_per_Establishment	22	34	38968.7	47742.7	543.6	230455.5
Log(Regulatory Constraints)	22	34	3.4928	0.6061	1.5867	4.8233
Population	1	34	266M	29M	220M	314M
GDP	1	34	8.3T	4.4T	2.1T	16.2T

4. Model

This section constructs our multisector endogenous growth model from its firm microeconomic foundation.

4.1. Overview

We consider an economy populated by identical individuals who supply labor services from their unit of time endowment in a competitive labor market and spend their earnings on consuming final goods. Their decisions yield an optimal path of expenditures and savings via freely borrowing and lending in a competitive market for financial assets at the prevailing interest rate. The households' income consists of wage income and returns on asset holdings. Government policy (i.e., regulation and spending) can convey benefits to the representative household but we restrict those benefits to be completely separable in the household's preferences.

The production side of the economy consists of a final-goods sector, commodity producers in *J* industries, and the intermediate-good suppliers in those *J* industries. The final-goods sector has a representative firm that produces final goods from the *J* commodity outputs from the *J* different industries. Likewise, each industry's commodities are produced by a representative firm that aggregates intermediate goods. Also, in each industry, there is a continuum of identical firms that make differentiated intermediate goods from labor and also invest in gaining additional knowledge to reduce their production costs. As each firm invests in knowledge, it contributes to the pool of public knowledge that benefits the quality of all intermediate-goods firms in the industry. The engine for endogenous growth in this economy is the increasing returns

---

industry, regulation Granger causes 2.8 of the five dependent variables, while 2.4 of the dependent variables Granger cause regulation. Detailed results are available upon request.

from this public knowledge and the variety of intermediate goods that grow with the proliferation of intermediate-goods firms.

4.2. Households' primitives

The population, $P(t)$ of identical individuals is exponentially growing at rate ω from an initial level of P_0. At time t, each individual maximizes the discounted stream of dynastic utility:

$$\int_t^\infty e^{-(\rho-\omega)(s-t)} \left[\ln c + \lambda \ln (1-l) + u(G,R) \right] ds, \tag{1}$$

where the time arguments to functions have been suppressed for brevity, ρ is the subjective discount rate, c is per capita consumption, l is the fraction of time allocated to work ($1-l$ is leisure), λ governs preference for leisure, and u represents any utility received from government spending (G) and regulations (R), such as environmental protection and safety precautions. The individual faces the following flow budget constraint:

$$\dot{a} = [r(1-\tau_A) - \omega] a + (1-\tau_L) Wl - (1+\tau_C) p_Y c, \tag{2}$$

where a is the individual's asset holdings, r is the rate of return to savings (and the after-tax rate of return to equity), τ_A is the tax rate on assets, W is the wage rate, τ_L is the tax rate on labor income, τ_C is the tax rate on consumption, and p_Y is the price of the final goods the individual consumes. From this objective and budget constraint, we derive the household's labor supply and the Euler equation, as shown in appendix A.

4.3. Final-goods producers' primitives

The representative firm produces final goods (Y) with a constant elasticity of substitution technology:

$$Y = \left[\sum_{j=1}^J \psi_j Y_j^{\frac{\xi-1}{\xi}} \right]^{\frac{\xi}{\xi-1}}. \tag{3}$$

Exhibiting constant returns to scale (i.e., $\sum \psi_j = 1$), this final-goods sector is perfectly competitive and generates no profits. It serves only to assemble final goods from the commodity outputs of J different industries.

4.4. Industry commodities producers' primitives

Each industry has a representative firm that produces its commodity (Y_j) from a mass of differentiated intermediate goods (X_{ij}):

$$Y_j = N_j \left[\int_0^{N_j} \frac{1}{N_j} X_{ij}^{\frac{\chi_j-1}{\chi_j}} di \right]^{\frac{\chi_j}{\chi_j-1}}, \tag{4}$$

where N_j is the mass of intermediate-goods firms for industry j, and χ_j is the elasticity of substitution for industry j. The representative commodity producer for industry j maximizes its profits:

$$\Pi_j = p_{Yj} Y_j - \int_0^{N_j} p_{Xij} X_{ij} di. \tag{5}$$

Because this sector is perfectly competitive, the industry commodity producer earns zero economic profit. The industry commodity producer's demand for the intermediate good is implicitly given by:

$$p_{xij} X_{ij} = p_{Yj} Y_j \left[\frac{p_{xij}^{1-\chi_j}}{\int_0^{N_j} p_{xij}^{1-\chi_j} di} \right]. \tag{6}$$

4.5. Intermediate-goods producers' primitives

The typical intermediate firm produces its differentiated good with the following technology:

$$X_{ij} = Z_{ij}^{\zeta_j(R_j)} \left[L_{Xij} - \phi_j\left(R_j\right) \right], \tag{7}$$

where $Z_{ij}^{\zeta_j(R_j)}$ is the total factor productivity, Z_{ij} is the labor-enhancing knowledge specific to the firm, ζ_j is the elasticity of the intermediate firm's output to knowledge, L_{Xij} is the labor employed in producing X_{ij}, and ϕ_j is a fixed labor cost. Appearing in two locations in this production function, R_j represents regulatory constraints. Both parameters in this production function may depend on regulations because, fundamentally, the purpose of regulations is to alter firms' behavior by constraining the decisions of firms. Regulations that require unproductive expenditures, such as safety precautions and regulatory compliance officers, would increase the firm's fixed labor cost. Regulations that prevent the use of production knowledge that is more productive, such as the knowledge that freely venting emissions into the environment is more efficient (from the firm's perspective) than reducing such emissions, would decrease the firm's total factor productivity. The firm accumulates knowledge according to the technology:

$$\dot{Z}_{ij} = \kappa\left(R_j\right) K_j L_{Zij}, \tag{8}$$

where L_{Zij} is the quantity of labor invested in knowledge accumulation, K_j is the stock of public knowledge in the industry, and $\kappa\left(R_j\right)$ governs how much knowledge is generated by the firm's investment. The parameter in this production function of knowledge may also depend on regulations to the extent that regulations may result in a firm's investments generating less knowledge, such as a constraint requiring a fraction of the time that research and development workers use for safety precautions or environmental protection. Public knowledge, which is nonrival and nonexcludable, accumulates via spillovers:

$$K_j = \max_i Z_{ij}. \tag{9}$$

In equilibrium, under the assumption of within-industry symmetry of identical intermediate-goods firms, the stock of firm-specific knowledge will end up equaling the stock of public knowledge over the industry (i.e., $Z_{ij} = K_j \forall i$). The firm's pretax flow of profits is its flow of revenue net contemporaneous production costs, knowledge accumulation investment (I_{Zij}), and fixed cost (F_{ij}):

$$\Pi_{ij} = p_{Xij} X_{ij} - W Z_{ij}^{-\zeta(R_j)} X_{ij} - \underbrace{W L_{Zij}}_{I_{Zij}} - \underbrace{W \phi_j}_{F_{ij}}. \tag{10}$$

The firm maximizes its value, which equals the present value of its stream of after-tax profits:

$$V_{ij}(t) = \int_t^\infty e^{-\int_t^s [r(v)+\delta_j]dv} (1 - \tau_\pi) \, \Pi_{ij}(s)\, ds, \tag{11}$$

where τ_π is the tax rate on corporate income, r is the interest rate (return to savings), and δ_j parameterizes the hazard of a death shock. The firm maximizes V_{ij} subject to the knowledge accumulation process [Eq. (8)], a given initial stock of knowledge $Z_{ij}(0) > 0$, and a nonnegativity constraint on the growth of knowledge. The solution to this problem, detailed in appendix A, yields the (maximized) value of the firm given exogenous variables.

4.6. Government primitives

The federal government sets tax rates, determines expenditures, and regulates. Because the government's behavior is purely exogenous, we have not engaged in modeling it.[15] Nonetheless, it is noteworthy that our limited model of the government does not necessarily reduce it to a useless drain on the economy. Government spending and regulation can be net beneficial at some level depending on its magnitude in the household's separable utility function where it directly enters. Alternatively, we could model government activities as entirely wasteful and our results would still hold.

---

[15] We could describe a budget constraint for the government that is consistent with our model. Suppose that any deficit gets supplied by net imports (M) and accumulates into debt (D) held by foreign investors willing to inject real wealth into the economy in return for that principal wealth plus interest (at the market rate) financed by future tax receipts:

$$D = [rD + Mp_Y] + [G - M]\, p_Y - \left[\tau_L W L + \tau_C C + \tau_A r Pa + \sum_{j=1}^{J} N_j \tau_\pi \Pi_{ij} \right],$$

where G is the quantity of final goods purchased by the government and rD is the repayment to the (foreign) owners of the debt.

4.7. Solution for equilibrium

The model primitives result in supply and demand equations for the goods markets, the labor market, and the asset market. Bringing in market-clearing conditions, we attain a solution for the equilibrium as detailed in appendix A. In this section, we simply describe the three equations in the solution that motivate our empirical specification.

Real expenditure by firms on investment per establishment, equal to the labor allocated to knowledge accumulation times the wage rate, is a linear function of the real interest rate (which we define as the nominal interest rate on long-term US Treasury bills net the growth in nominal wages) and real output per establishment:

$$
\frac{I_j}{N_j W} = \left[-\frac{\delta_j}{\kappa_j(R_j)} \right] + \left[-\frac{1}{\kappa_j(R_j)} \right] \left(r_{Zj} - \frac{\dot{W}}{W} \right) + \left[\zeta_j(R_j) \left(\frac{\chi_j - 1}{\chi_j} \right) \right] \left(\frac{p_{Yj} Y_j}{N_j W} \right). \tag{12}
$$

Note that each of the bracketed terms depends on regulation. The state equation that drives the dynamics of the model is fairly simple: the change in real output per establishment is a linear function of its level, as well as real investment per establishment and an interaction between the real interest rate with the level of real output per establishment:

$$
\left(\frac{\dot{p_{Yj} Y_j}}{N_j W} \right) = \left[-\psi_j \left(\frac{1 - \tau_\pi}{\theta W^*} \right) \left(\frac{1}{\chi_j} - \zeta_j(R_j) \left(\frac{\chi_j - 1}{\chi_j} \right) \right) \right] + \left[-\psi_j \kappa_j(R_j) \right] \left(\frac{I_j}{N_j W} \right)
$$
$$
+ \left[-\psi_j \zeta_j(R_j) \left(\frac{\chi_j - 1}{\chi_j} \right) - \frac{(1 - \tau_\pi) W^*}{\theta \kappa_j(R_j) \psi_j} \left(\delta_j - \kappa_j(R_j) \phi_j(R_j) \right) \right] \left(\frac{p_{Yj} Y_j}{N_j W} \right)
$$
$$
+ \left[-\frac{(1 - \tau_\pi) W^*}{\theta \kappa_j(R_j) \psi_j} \right] \left(\left(r - \frac{\dot{W}}{W} \right) \times \left[\frac{p_{Yj} Y_j}{N_j W} \right] \right). \tag{13}
$$

Note that each of the bracketed terms depends on regulation. According to the model, as detailed in appendix A, the number of people per establishment should follow the exact same process. These dynamic processes naturally lend themselves to an autoregressive (AR(1)) econometric structure.

5. Estimation methodology

Our structural model can be estimated by a few simple linear regressions, where the specification for each is explicitly derived from the model. To operationalize eq. (12) into a regression, we replace the bracketed terms with composite parameters and include a disturbance term. In our data, we actually have three different measures of investment: equipment, structures, and intellectual property. Some of that investment would map to the knowledge accumulation investment in the model, while other investment would map to the creation of additional establishments (for which the model specifies a different equation governing that effort). Rather than trying to reproduce that mapping in an ad hoc fashion, we have simply used all three of these measures and let the data identify the contribution of each. Hence, this yields three different investment regressions that have essentially the same form.

Taking an agnostic approach about exactly how investment changes with regulation, we want a specification that can flexibly capture how regulations tend to affect the parameters governing firms' investment decisions, whatever the deeper underlying mechanisms at work. Hence, we include regulation in the most flexible form that is practical for our data: we allow each composite parameter to be linear in both the log of the current level of regulation and the lag of the regulation, which has the effect of including the current stock regulation and its lag as independent regressors that each interact with each of the other regressors specified by the theoretical model[16]:

$$
\left[\frac{I_{jst}}{N_{jt} W_t} \right] = \beta_{0js} + \beta_{1js} \left[r_t - \frac{W_{t+1} - W_t}{W_t} \right] + \beta_{2js} \left[\frac{p_{Yjt} Y_{jt}}{N_{jt} W_t} \right] + \beta_{3js} \log R_{jt}
$$
$$
+ \beta_{4js} \log R_{jt-1} + \beta_{5js} \left[\left(r_t - \frac{W_{t+1} - W_t}{W_t} \right) \times \log R_{jt} \right]
$$
$$
+ \beta_{6js} \left[\left(r_t - \frac{W_{t+1} - W_t}{W_t} \right) \times \log R_{jt-1} \right]
$$
$$
+ \beta_{7js} \left[\left(\frac{p_{Yjt} Y_{jt}}{N_{jt} W_t} \right) \times \log R_{jt} \right] + \beta_{8js} \left[\left(\frac{p_{Yjt} Y_{jt}}{N_{jt} W_t} \right) \times \log R_{jt-1} \right] + \varepsilon_{jst}. \tag{14}
$$

---

[16] Recall that including both the log and its lag in the specification is equivalent to including the level (in log scale) and the growth rate (as the first difference in logs).

The t subscript, indicating the year and previously suppressed to keep our theory model's expressions compact, has been included here. We also introduce a new subscript, s, to indicate each of the three types of investment. Note that the econometric model is specified in real terms (units of labor) from the model, which does not include inflation. However, by normalizing all prices by the average nominal wage rate for the nation (i.e., total annual labor earnings divided by number of full-time and part-time workers), all measures are now in real terms, both in units of labor and with inflation effectively removed. Our estimations and counterfactuals are then conducted by using these real units, but the results can then easily be transformed into nominal terms and then into the more conventional measure of real terms as desired (e.g., using the BEA's GDP deflator).

As mentioned in the modeling section, the state equation that drives the dynamics of the model naturally lends itself to an AR(1) in the real output per firm. We proceed in operationalizing eq. (15) in a manner similar to the investment equation. We (a) operationalize this into a regression by replacing the square bracketed terms into composite parameters, (b) include a disturbance term, (c) include each of the investment measures, and (d) interact each term with both the log of regulations and its lag:

$$
\begin{aligned}
\left[\frac{p_{Yjt+1}Y_{jt+1}}{N_{jt+1}W_{t+1}} \right] = {} & \alpha_{0j} + \sum_s \alpha_{1js}\left[\frac{I_{jst}}{N_{jt}W_t} \right] + \alpha_{2j}\left[\frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \right] \\
& + \alpha_{3j}\left[\left(r_t - \frac{W_{t+1}-W_t}{W_t}\right) \times \frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \right] + \alpha_{4j}\log R_{jt} + \alpha_{5j}\log R_{jt-1} \\
& + \sum_s \alpha_{6js}\left[\left(\frac{I_{jst}}{N_{jt}W_t}\right) \times \log R_{jt} \right] + \sum_s \alpha_{7js}\left[\left(\frac{I_{jst}}{N_{jt}W_t}\right) \times \log R_{jt-1} \right] \\
& + \alpha_{8j}\left[\frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \times \log R_{jt} \right] + \alpha_{9j}\left[\frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \times \log R_{jt-1} \right] \\
& + \alpha_{10j}\left[\left(r_t - \frac{W_{t+1}-W_t}{W_t}\right) \times \frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \times \log R_{jt} \right] \\
& + \alpha_{11j}\left[\left(r_t - \frac{W_{t+1}-W_t}{W_t}\right) \times \frac{p_{Yjt}Y_{jt}}{N_{jt}W_t} \times \log R_{jt-1} \right] + \varepsilon_{0jt}.
\end{aligned}
\tag{15}
$$

As previously mentioned, the number of people per establishment follows the same process. Hence, we estimate a nearly identical AR(1), with these same covariates, for people per establishment (where real GDP per establishment on the left-hand side, as well as its appearance on the right-hand side, gets replaced by people per establishment and its lag).

A key element of our approach is our treatment of the different types of investment. While the specification is symmetric with respect to how the various types of investment variables enter into the regressions, their estimated parameters are not symmetric. Thus, if any type of investment (e.g., equipment) does not actually contribute to observed growth, then its specific parameter would be estimated as zero or close to zero. Similarly, in our counterfactual simulations (presented in the next section), we do not force each type of investment to symmetrically affect growth via a common parameter but instead use the corresponding parameter estimate specific to each type of investment. In theory, each of these industry-specific regressions could simply be estimated by ordinary least squares (OLS). However, because we have only 36 data points for each regression, OLS runs the risk of overfitting the data given the number of parameters in our regression equations. The risk with such overfitting is that out-of-sample predictions on these linear equations can explode beyond reasonable bounds. Indeed, we have included simple OLS results in our plots (as red lines), and we do witness a few such explosions. A simple pooling of all the data would yield more degrees of freedom but would be pooling industries that are very heterogeneous, as well as somewhat diluting our inter-industry identification strategy.

Instead, we make the assumption that each industry's parameter is drawn from a normal distribution across all industries (called a hyperdistribution). This sort of hierarchical model structure represents a compromise between completely pooling the data and running strictly separate regressions. Hierarchical models enable a partial pooling of information across industries. When an industry's own data are too noisy, the parameter values to which OLS would have overfit will shrink back toward the grand mean (a parameter of the hyperdistribution—a so-called hyperparameter), so outlying values for a parameter become unlikely.

The easiest way to fit such a hierarchical model is with the Bayesian Markov Chain Monte Carlo method, with the hyperdistribution moved from specification of the likelihood to specification of the prior. Draws from a posterior distribution of uncertainty over parameter values, given the data, are generated from sampling the product of the likelihood of observing the data given the parameter values and the prior distribution of uncertainty over parameter values.[17] We begin with flat

---

[17] Bayesian estimation is equivalent to classical estimation when the priors are flat—meaning that the prior is a constant, so that the posterior is just the likelihood. (Maximum likelihood estimation finds the maximum of that likelihood and invokes a central limit theorem to get a normal distribution of uncertainty around that point estimate, while Bayesians would use the likelihood itself as the uncertainty distribution with the mode, mean, or median serving as equally good point estimates.)
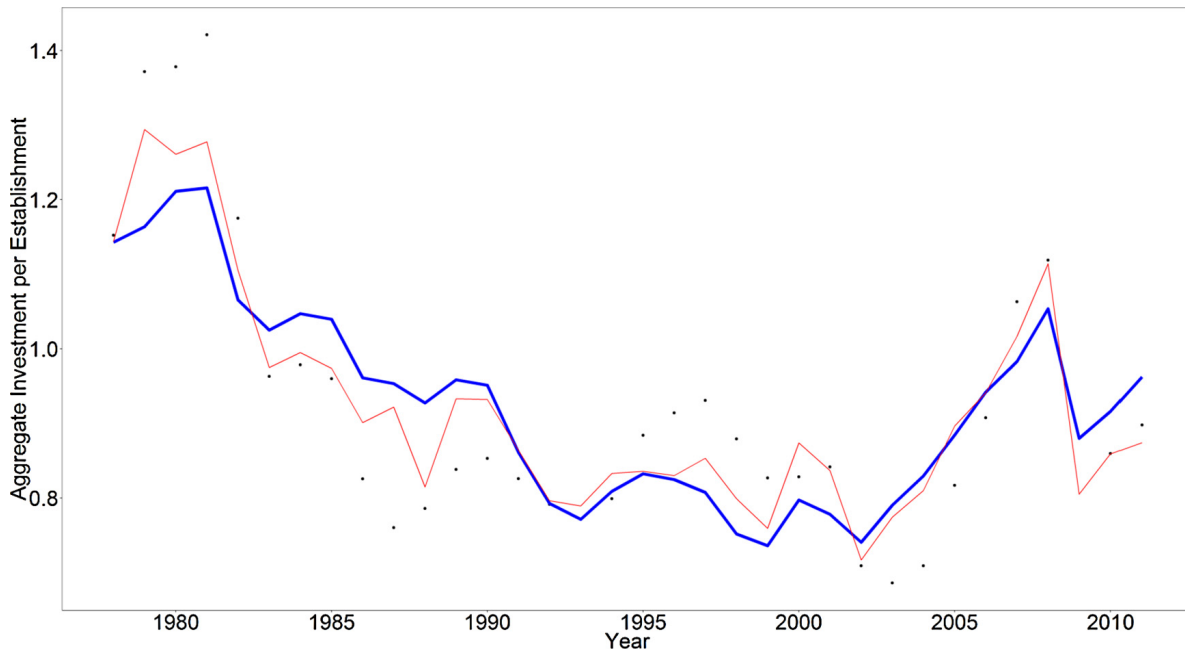
Fig. 3. Goodness of fit for real investment (equipment, intellectual property, and structures aggregated) using Bayesian estimates (blue) and ordinary least squares (red) as a basis for comparison. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

priors so that the only information we impose on this estimation is that the industry-specific parameters are drawn from a common normal hyperdistribution. Likewise, the hyperparameters are also given flat priors, so we are completely agnostic on the degree of similarity between these industry parameter values.

Once we have estimated these equations, we can make predictions on GDP by simply dividing real GDP per establishment by the number of people per establishment and then scaling the resulting rational function by the population (and the nominal wage if we desire to have GDP in nominal terms). Because the residuals are orthogonal to the regressors by construction, presumably those unobserved shocks would be unchanged by the counterfactual and can be added to the resulting estimate to retain those features in the data (e.g., the financial crisis) when comparing the factual to the counterfactual.

6. Results

One of our primary goals is to use the model to produce a counterfactual of a different regulatory regime, one that "freezes" regulations at some point in the past. Hence, what matters most is not hypothesis testing on individual parameters and providing an insightful interpretation of such results; indeed, there are too many parameters to realistically review within the confines of a single academic article.[18] Instead, what matters for our purposes is a good fit to the within-sample data and a reasonable argument for the external validity of our out-of-sample forecasting. The latter is justified by our theoretical model, while the former is justified below, where we summarize the performance of the model by examining goodness of fit on within-sample data.

6.1. Performance of the model

The goodness of fit is apparent in Figs. 3, 4, and 5. Eqs. (14) and (15) indicate five goodness-of-fit tests: one equation for each of the three types of investment outcomes, one AR(1) equation for real output per establishment as the outcome variable, and one AR(1) equation for persons per establishment as the outcome variable.[19] For easier reporting, we aggregate the three types of investment across all individual industries into a single measure of total investment.[20] Fig. 3 shows the observed values of total investment, alongside predicted values from our preferred Bayesian estimates and a benchmark set of OLS estimates.

We ran 1,000 simulations for each tuple of parameters for each industry, drawing from the posterior distribution. Figs. 4 shows goodness of fit for GDP per establishment, while Fig. 5 shows goodness of fit for people per establishment. In both figures, the blue line plots the mean prediction for the outcome variable across those 1,000 uncertainty draws for our

---

[18] See appendix C for summary statistics across the Monte Carlo uncertainty draws on the hyperparameters as our attempt at concisely reporting our parameter estimates.

[19] Figs. B1–B5 (appendix B) show additional goodness-of-fit figures for each individual industry for all five of these estimation equations.

[20] Appendix B reports goodness-of-fit figures for each type of investment and industry, rather than the aggregated result.
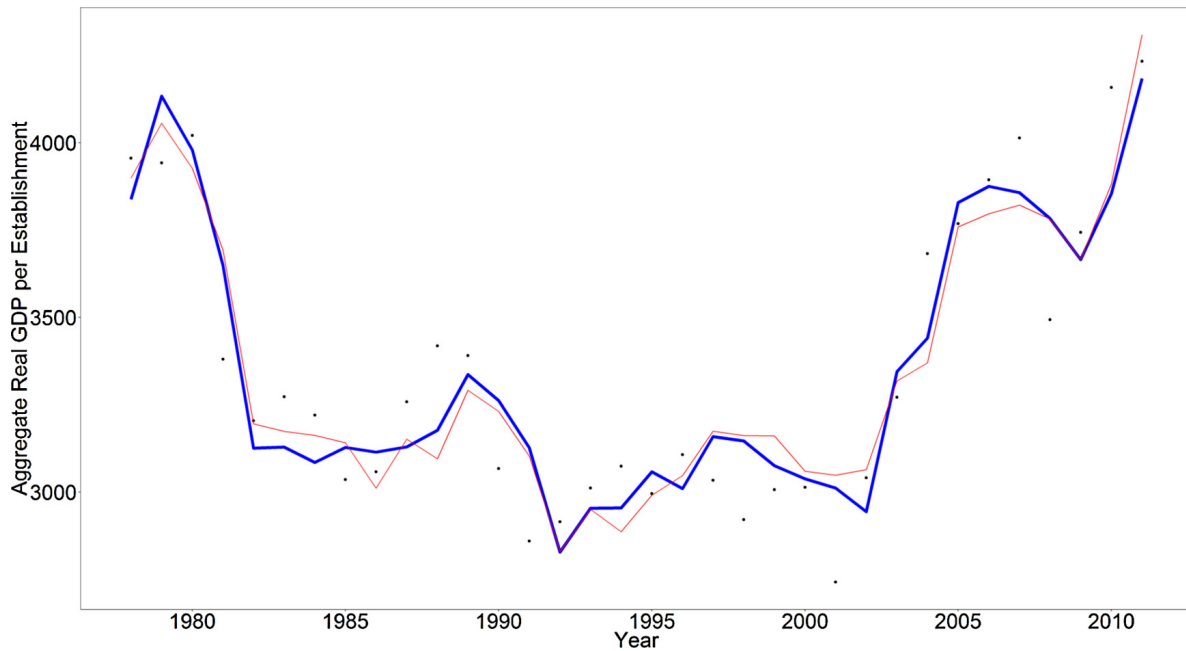
Fig. 4. Goodness of fit for nominal GDP per establishment using Bayesian estimates (blue) and ordinary least squares (red) as a basis for comparison.
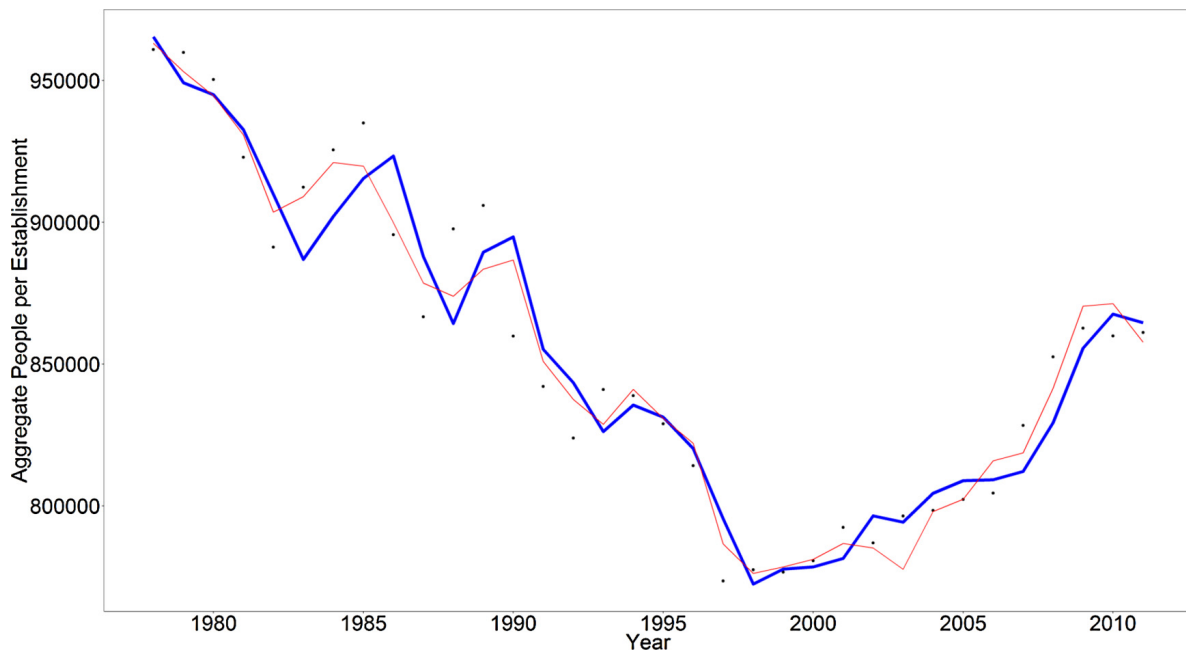

Fig. 5. Goodness of fit for people per establishment using Bayesian estimates (blue) and ordinary least squares (red) as a basis for comparison.

Bayesian hierarchical model. The red line plots the same prediction when OLS is used to estimate the parameters (without a hyperdistribution in a hierarchical structure) as a basis for comparison. These are also shown at the industry level in appendix B, where we report the R^2 summary statistic in color corresponding to the respective method used to generate the predictions for each industry.

In summary, Figs. 3 through 5 show goodness of fit plots. Each figure has observed data plotted as black dots, Bayesian predictions plotted as the blue line, and OLS predictions plotted as the red line.[21]

---

[21] We are not particularly concerned with nonstationarity here. Even within our 36 years of data for each industry, our variables of interest do not appear to be too trendy (as can be gleaned from the graphs). Moreover, nonstationarity is a greater concern when performing hypothesis tests on some parameters, where the appropriate distribution under the null hypothesis can change considerably in the presence of nonstationarity. In our context of prediction, the main concern is avoiding spurious correlation, which again is less of a concern because our series do not appear to be following a common trend.
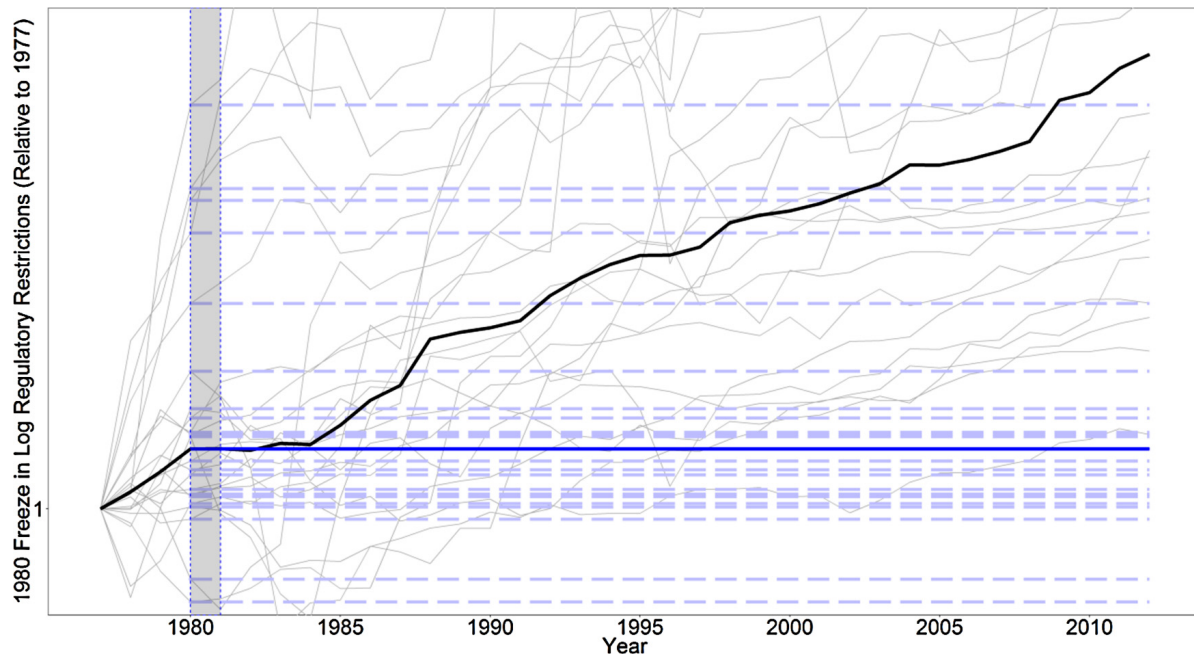
Note: Thick lines show averages.

Fig. 6. Factual (solid) and counterfactual (dotted) regulation time paths.

We find the fits to be acceptable across the board. In general, the blue fits of our Bayesian model appear less noisy, which is attributable to the hierarchical structure that effectively downweights the temptation to overfit when the underlying signal in the data is less informative and instead shrinks the prediction toward the grand mean. The R^2 often gets higher for OLS because it is not constrained by the joint distribution of the industry-specific parameters and hence is free to overfit the data. The fits are generally highest when examining the dynamic estimation equations for real output per establishment and people per establishment. This finding should be expected because these dynamic estimation equations have an AR(1) structure and most economists' ex ante expectation here would be for considerable persistence in these outcome variables. Nonetheless, we are pleased to have such tight fits because our counterfactual forecast ultimately produces predictions for GDP as a rational function that puts real output per establishment in the numerator and people per firm in the denominator. With stochastic disturbance terms that are normal and mean zero, this implies that the noise in their ratio will have the much fatter tails of the Cauchy distribution, so obtaining a low variance of the stochastic terms in these dynamic equations is highly desirable.

6.2. Counterfactual simulation

Satisfied with our model's performance, as demonstrated by its goodness of fit, we proceed to our counterfactual simulation. For our counterfactual experiment on regulation, we freeze regulations at the levels observed in the year 1980 (i.e., regulation in 1981 is fixed at 1980 levels) and sustain that freeze through the last year in our study, 2012. This counterfactual regulatory regime is depicted in Fig. 6, where the solid thin lines show the observed regulation levels of industries, the thin dashed lines show the counterfactual levels of industries, and the thick lines show the average observed and the average counterfactual industry.

We have elected to freeze regulations over so broad a window in our study in an effort to be more comparable with the leading estimate in the literature, where that counterfactual experiment was a regulatory freeze from the late 1940s until the early 21st century (Dawson and Seater, 2013).

Our counterfactual simulation then proceeds one year at a time. First, we predict the amount of each of the three types of investment, inputting into the model that year's real output per establishment,[22] the real interest rate (shocks to which have been assumed to be strictly exogenous), and our measure of the amount of regulation (which, for the counterfactual, was frozen at the level observed in 1980). To each prediction we add in the shocks (i.e., the residuals from the regressions described in our goodness of fit tests), which are orthogonal by construction but help avoid false comparisons between our

---

[22] As a robustness test, we performed common correlation group mean estimates of the investment equations given in eq. (14) and predicted the counterfactual investments. Although these techniques could not be adapted to our hierarchical model to produce GDP or people-per-establishment counterfactuals, the counterfactual investment time path, as reported in Fig. B13, is substantively similar to the time path constructed with our Bayesian approach, which is shown in Fig. 3. Disaggregated or other additional output is available on request.
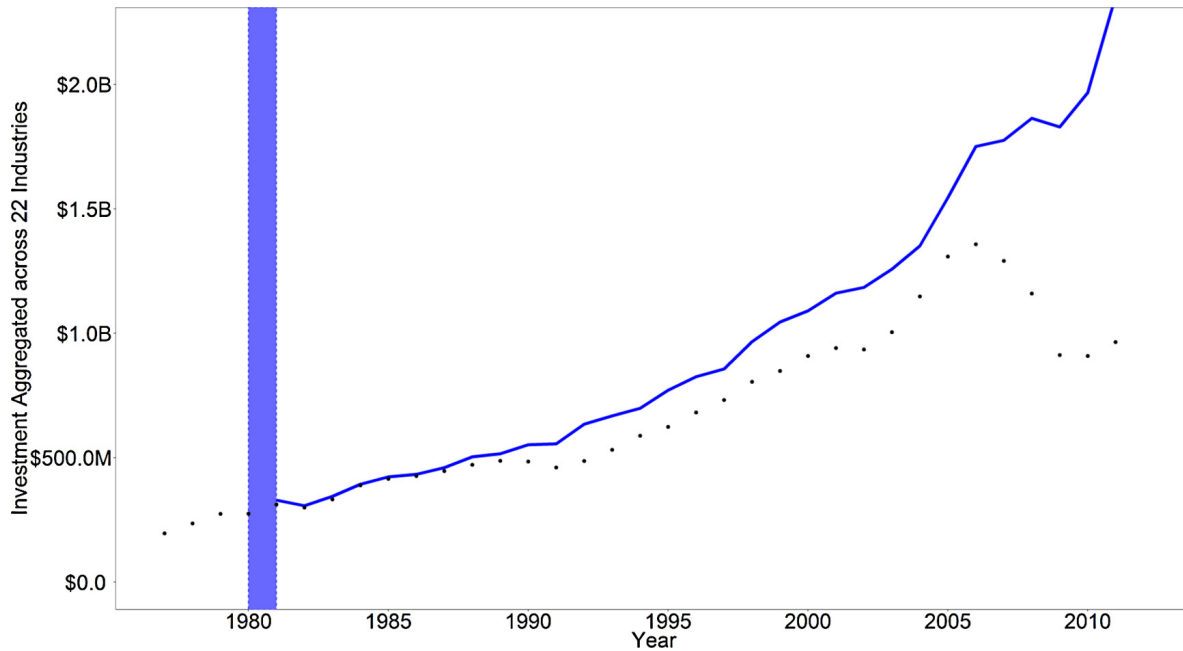
Fig. 7. Factual (dots) and counterfactual (solid) total investment, aggregated across all 22 industries.

counterfactual and the observed data by capturing major business cycle events (such as the Great Recession) that would not be produced by the model. After obtaining predictions for investment under the regulatory freeze counterfactual, we turn our attention to how these alternate levels of investment would change our primary economic outcomes of interest. To do this, we predict the next year's real output per capita and persons per establishment from the current year's real output per capita and persons per establishment, as well as the current year's investment (all three types), regulation, and real interest rate. Once we have made all the predictions of outcome variables in an industry in a year, we repeat the two-step process for the next year, using the counterfactual prediction for real output per capita or persons per establishment as the inputs in the next year's first predictions.

This exercise results in the counterfactual prediction for each of the three types of investment, for real output per establishment, and for persons per establishment. Fig. 7 shows the former—investment—in the form of the total of all three types of investment aggregated across the 22 industries. The counterfactual predictions are performed at the disaggregated level, as explained above. (Figures for each industry and type of investment are given in appendix B, Tables B8–B10.)

In general, the counterfactual predictions appear reasonable with the Bayesian predictions tending to be milder than the OLS predictions (shown in appendix B). However, there are a few industries that have a few variables that explode for the OLS forecasting—either implausibly strong exponential growth that dwarfs the observed data into a flat line or an exponential dive below zero. We interpret these findings as the hazards of using OLS with so few degrees of freedom in a given industry. Essentially, this sort of undesirable behavior justifies our use of the more complicated Bayesian estimation technique because estimates for individual industries will be tied to the grand mean across industries, making it unlikely for an individual industry's predictions to dramatically diverge from the rest of the pack.

Similar behavior occurs in our prediction for value added to GDP for each industry that results from our dividing of the counterfactual prediction for real output per establishment by establishments per person (and then scaled up by population, which is assumed to be exogenous). Again, the results appear reasonable except for those few industries that suffer from the OLS predictions exploding outside the reasonable range. To get the aggregate effect across these 22 industries, we can simply add up our counterfactual predictions, as depicted in Fig. 8.

Fig. 8 shows observed value added to GDP (represented by dots), summed across the 22 industries in our study, compared to the counterfactual prediction from our model, shown as a blue line. The shaded area to the top and bottom of the blue solid line is the 90 percent confidence band. With compounding growth, as well as other recursive models, confidence tends to decrease after many successive iterations. The confidence band on our counterfactual GDP prediction, however, remains fairly small until about the year 2000. The area below the confidence band remains small until the Great Recession, when it finally begins to expand. On the other side of our prediction, the confidence interval began expanding a few years earlier, and by 2004 or so it had grown past the $10 trillion point. Given our model's focus on knowledge creation and the growth that innovation can spur, it seems appropriate that our confidence interval generally exhibits more explosive growth on the upper bound of our prediction.

In aggregate, we find about 0.8 of a percentage point in the (real) growth rate is lost from the additional regulations promulgated since 1981. To the extent that we are comfortable treating these 22 industries as representative of the rest of the economy, our result would scale up to a similar loss in aggregate GDP (0.8 of a percentage point). Over the course of 30 years, the difference between our counterfactual simulation and observed GDP grew to about $4 trillion in year 2012, or
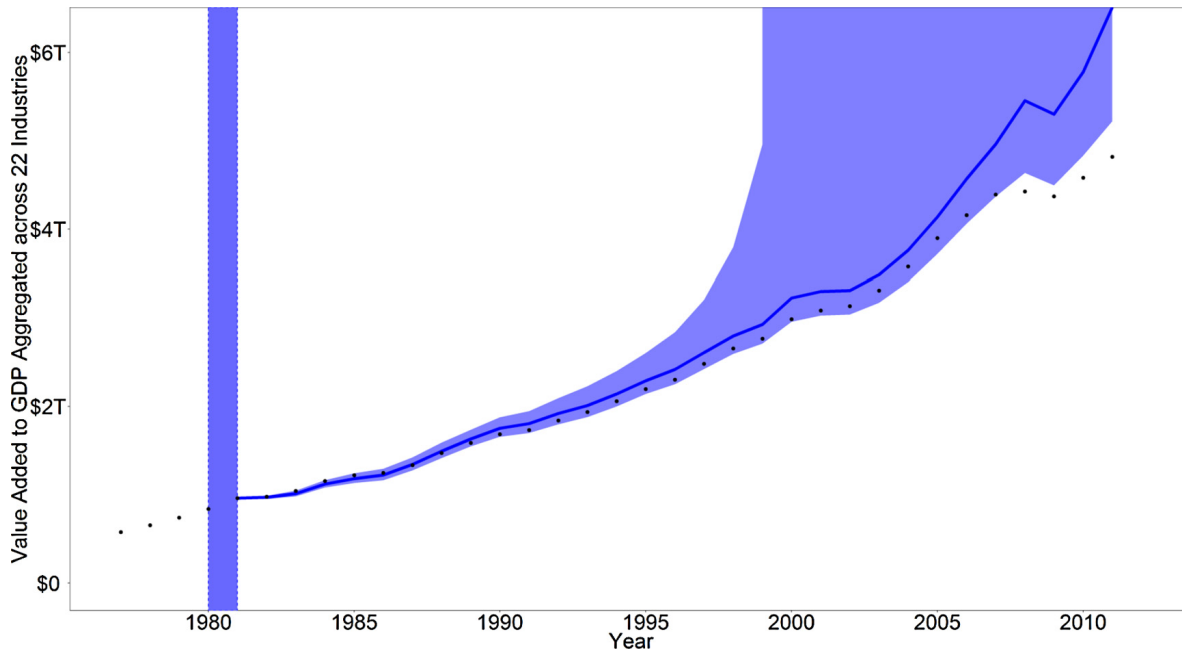
Fig. 8. Factual (dots) and counterfactual (blue line) value added to GDP, with 90 percent confidence interval.

nearly $13,000 per capita. Although our finding is slightly less than one-half of the result in Dawson and Seater, 2013, it is still within a reasonable range, especially since their study covers a longer time horizon. We caution that our results are for the costs of regulation to the measurable economy (net any benefits to the measurable economy), but that does not imply that none of the regulations promulgated since 1981 have been net beneficial. Indeed, many regulations exist to generate nonmarket benefits that would only indirectly affect the measurable economy. Nevertheless, this suggests that a widescale review of regulations—for example, a commission focused on eliminating redundant or obsolete regulations and supplanting command-and-control regulations with simpler market-based mechanisms—would deliver not only lower compliance costs but also a substantially higher economic growth rate.

7. Conclusion

Though the topic of regulation and economic growth has been widely studied, most studies focus on a narrow set of regulations, industries, or both. Such designs cannot estimate the cumulative effect of regulation, even though accumulation of an increasingly complex set of regulatory constraints is the most striking characteristic of the history of regulations in the United States. A single regulation may appear net beneficial when examined on its own—indeed, government agencies typically claim that all, or nearly all, their regulations create positive net benefits—but may still have a net negative effect on economic growth by virtue of being piled on top of (and possibly interacting with) other regulations. Although government agencies do not have much of an incentive to count these cumulative effects in their costs, it is only fair to acknowledge that economists have not provided these agencies with the tools they would need to quantify all the costs of their regulations.

We model this cumulative effect in the context of endogenous growth. Our model considers the effect of regulation on a firm's investment choices, which are factors that lead to innovation and efficiency. We have dramatically extended the model of Peretto and Connolly (2007) to a truly multisector economy, where each heterogeneous industry's growth is governed by a set of linear equations that can be influenced by regulatory shocks. To our knowledge, ours is one of the first multisectoral endogenous growth model with closed-form solutions.[23] We then estimate our model using industry data from the BEA in combination with RegData 2.2.

We are able to estimate our model with a panel dataset of 22 industries observed annually from 1977 to 2012. Specifically, we examine not only the direct effect of regulation on the output per establishment and (net) entry of additional establishments but also the indirect effect of regulation on three different types of investment that in turn affect the output per establishment and (net) entry of additional establishments. Our dataset uses BEA and Census Bureau measures of these industries, combined with the measures of regulation by industry provided by RegData's text-analysis-based quantification of regulations impacting industries over time.

A key insight of endogenous growth models in general is that the effect of government intervention on economic growth is not simply the sum of static costs associated with individual interventions. Instead, the deterrent effect that intervention can have on knowledge growth and accumulation can induce considerable deceleration to an economy's growth rate. Our

---

[23] See also Cai and Li (2019).

results suggest that regulation has been a considerable drag on economic growth in the United States, on the order of 0.8 percentage points per year. Our counterfactual simulation predicts that the economy would have been about 25 percent larger than it was in 2012 if regulations had been frozen at levels observed in 1980. The difference between observed and counterfactually simulated GDP in 2012 is about $4 trillion, or $13,000 per capita.

The empirical investigation of the relationship between all federal regulations and a broad panel of industries is unprecedented and has only recently been made possible by the advent of RegData. Further research along these lines is also possible with RegData. For example, RegData permits the user to separate regulations according to the agency that created them. Researchers could therefore examine the effects—positive or negative—of regulations created by labor, environmental, or financial regulatory bodies in isolation or in comparison to other sorts of regulation.

Acknowledgments

Appendix A

The household saves according to the following Euler equation:

$$\frac{\dot{c}}{c} + \frac{\dot{p}_Y}{p_Y} = (1 - \tau_A) r - \rho. \tag{16}$$

The individual's optimal tradeoff of labor and leisure results in the following aggregate supply of labor:

$$L = Pl = P - \frac{(1 + \tau_C) \lambda p_Y c P}{(1 - \tau_L) W}. \tag{17}$$

The representative firm for the final-goods sector must have a unit cost equal to p_Y for there to be zero economic profit:

$$p_Y = \left(\sum_{j=1}^{J} \psi_j^{\xi} p_{Yj}^{1-\xi} \right)^{\frac{1}{1-\xi}}. \tag{18}$$

Its conditional demands for the commodity outputs of industry j, Y_j, are implicitly given by

$$p_{Yj} Y_j = \psi_j^{\xi} p_{Yj}^{1-\xi} p_Y^{\xi-1} p_Y Y. \tag{19}$$

Summing over all J industries gives

$$\sum_{j=1}^{J} p_{Yj} Y_j = p_Y Y. \tag{20}$$

To characterize the behavior of the typical intermediate-goods firm, consider the following current value Hamiltonian:

$$H_{ij} = (1 - \tau_\pi) \Pi_{ij} + q_{ij} \kappa_j (R_j) K_j L_{Zij}, \tag{21}$$

where the costate variable, q_{ij}, is the value of the marginal unit of knowledge. With X_{ij} pinned down by market demand given $p_{X_{ij}}$, the control variables are the price of the intermediate good ($p_{X_{ij}}$) and the effort in knowledge accumulation ($L_{Z_{ij}}$). Note that the production and investment decisions are separable in the Hamiltonian. The first-order condition on $p_{X_{ij}}$ yields the intermediate-goods firm's pricing strategy as a markup (depending on the price elasticity of demand) over marginal cost:

$$p_{X_{ij}} = \left(\frac{\chi_j}{\chi_j - 1} \right) \left[Z_{ij}^{-\zeta_j(R_j)} W \right]. \tag{22}$$

By substituting into the firm's pretax flow of profits this pricing strategy and the implicit equation for the conditional factor demand under within-industry symmetry across firms:

$$\Pi_{ij} = \left(\frac{1}{\chi_j} \right) \left[\frac{p_{Yj} Y_j}{N_j} \right] - W \phi_j (R_j) - L_{Zij} W. \tag{23}$$

Since the Hamiltonian is linear in L_{Zij}, its partial derivative yields three cases: (1) The after-tax cost exceeds the value of the marginal unit of knowledge. The firm then does not invest. (2) The after-tax cost is less than the value of the marginal

unit of knowledge. Since the firm would then demand an infinite amount of labor to invest in knowledge accumulation, this case violates the general equilibrium conditions and is ruled out. (3) The first-order conditions for the interior solution are given by equality between marginal revenue and marginal cost of knowledge:

$$\kappa_j\left(R_j\right)K_j q_{ij} = (1 - \tau_\pi)\,W \Leftrightarrow L_{Zij} > 0. \tag{24}$$

This first-order condition is paired with the constraint on the state variable and a terminal condition:

$$\lim_{s \to \infty} e^{-\int_t^s [r(v)+\delta_j]dv} q_{ij}(s)\,Z_{ij}(s) = 0. \tag{25}$$

The partial derivative with respect to the state, the remaining part of the Hamiltonian's first-order conditions (which gets equated to costate times the discount rate less the time derivative), yields a differential equation in the costate variable that amounts to an arbitrage condition:

$$r + \delta_j = \frac{\dot{q}_{ij}}{q_{ij}} + \frac{(1 - \tau_\pi)\frac{\partial \Pi_{ij}}{\partial Z_{ij}}}{q_{ij}}, \tag{26}$$

which defines the (after-tax) rate of return to investment in knowledge as the ratio between additional profit from the knowledge and its shadow price plus (minus) the appreciation (depreciation) in the value of knowledge. Substituting the condition for an interior solution for investment (in knowledge accumulation) into the arbitrage equation, including in time derivatives, the corporate income tax rate drops out of the equation governing optimal investment in quality improvement because investments in research and development get expensed:

$$r + \delta_j = \frac{\dot{W}}{W} - \frac{\dot{K}_j}{K_j} + \frac{\zeta_j\left(R_j\right)WK_j^{\zeta_j(R_j)-1}p_{Xij}X_{ij}/p_{Xij}}{\left(\frac{W}{\kappa_j(R_j)K_j}\right)}. \tag{27}$$

We can further simplify this equation by substituting in the knowledge accumulation technology, the pricing strategy, and conditional demands under within-industry symmetry of identical intermediate-goods firms:

$$r_{Zj} = \frac{\kappa_j\left(R_j\right)}{W}\left[\zeta_j\left(R_j\right)\left(\frac{\chi_j - 1}{\chi_j}\right)\frac{p_{Yj}Y_j}{N_j} - \frac{L_{Zj}W}{N_j}\right] + \frac{\dot{W}}{W} - \delta_j. \tag{28}$$

Entrepreneurs create new entrants in industry j by making an investment that requires their labor:

$$I_{Nj} = WL_{Nj}. \tag{29}$$

Entrepreneurs target only new product lines because entering an existing product line in Bertrand competition with the existing supplier leads to losses. With the nonrival, nonexcludable nature of knowledge, any new entrant starts in a state that is identical to the industry's incumbents. Because the new entrant solves a problem that is identical to an incumbent firm, the present value of profits from entry is V_{ij}. Entrepreneurs compare that present value of profits from introducing a new good to the entry cost. If the entry cost is higher, then there is no new entry. If the entry cost is less, entrepreneurs would demand an infinite quantity of resources, which would violate resource constraints. Therefore, any entry indicates that the value of firms equals the entry cost, given in the following equation:

$$V_{ij} = W\theta\left[\frac{p_{Yj}Y_j}{N_j}\right] \Leftrightarrow \dot{N}_j > 0. \tag{30}$$

Given this expression for the cost of entry, we can divide the investment in new entry by this entry cost to obtain the differential equation that describes the growth in the number of firms as a function of the labor expended on forming new entrants given the scale of the industry and the firm failure hazard rate:

$$\frac{\dot{N}_j}{N_j} = \frac{L_{Nj}}{\theta\left[p_{Yj}Y_j\right]} - \delta_j. \tag{31}$$

The rate of return on an investment in a new entrant to industry j is governed by an equation easily derived from the logs and time derivatives for the value of incumbent firms:

$$r + \delta_j = \frac{\dot{V}_{ij}}{V_{ij}} + \frac{(1 - \tau_\pi)\Pi_{ij}}{V_{ij}}. \tag{32}$$

Analogous to the other equation describing the rate of return on an investment in capital accumulation, this equation states that the rate of return to the entrepreneurial activity of creating new firms is the capital gain/loss (i.e., the growth

rate of the market price of the firm, V_{ij}) and the dividend-price ratio [i.e., the ratio of after-tax profits, $(1 - \tau_\pi)\Pi_{ij}$, to the market price of the firm, V_{ij}]. Substituting in for V_{ij} and its time derivative using our equation for entry costs, as well as profits, we obtain the following:

$$r_{Nj} = \frac{(1 - \tau_\pi) N_j}{\theta W [p_{Yj} Y_j]} \left(\frac{[p_{Yj} Y_j]}{\chi_j N_j} - W\phi_j(R_j) - \frac{L_{Zj} W}{N_j} \right) + \frac{\dot{W}}{W} + \frac{[\dot{p_{Yj} Y_j}]}{[p_{Yj} Y_j]} - \frac{\dot{N_j}}{N_j} - \delta_j. \tag{33}$$

We equate the quantity of assets held by individuals to the value of intermediate-goods firms and substitute in our expression for the value of firms to obtain the following:

$$Pa = \sum_{j=1}^{J} N_j V_{ij} = \sum_{j=1}^{J} N_j \theta W \left[\frac{p_{Yj} Y_j}{N_j} \right] = \theta W \sum_{j=1}^{J} p_{Yj} Y_j = \theta W p_Y Y. \tag{34}$$

Taking logs and time derivatives as follows:

$$\frac{\dot{a}}{a} = \frac{\dot{Y}}{Y} + \frac{\dot{p_Y}}{p_Y} + \frac{\dot{W}}{W} - \frac{\dot{P}}{P}. \tag{35}$$

Dividing both sides of the household's budget constraint by an individual's assets and substituting in the previous two equations yields the following:

$$\frac{\dot{a}}{a} = \frac{\dot{Y}}{Y} + \frac{\dot{p_Y}}{p_Y} + \frac{\dot{W}}{W} - \underbrace{\frac{\dot{P}}{P}}_{\omega} = [r(1 - \tau_A) - \omega] + \frac{(1 - \tau_L) Wl}{\frac{\theta W p_Y Y}{P}} - \frac{(1 + \tau_C) p_Y c}{\frac{\theta W p_Y Y}{P}}. \tag{36}$$

To solve for the interest rate in the model, we turn to the Euler equation, which describes the after-tax rate of return required by households to equal their subjective rate of time preference plus the growth in the individual's expenditure on consumption:

$$r_A = \left(\frac{1}{1 - \tau_A} \right) \left[\frac{\dot{c}}{c} + \frac{\dot{p_Y}}{p_Y} + \rho \right] = \left(\frac{1}{1 - \tau_A} \right) \left[\frac{\dot{C}}{C} - \frac{\dot{P}}{P} + \frac{\dot{p_Y}}{p_Y} + \rho \right], \tag{37}$$

where C is aggregate consumption. Aggregate consumption can be eliminated using the goods market-clearing condition (which takes the familiar form except without any investment term because investment in this model, (i.e., $L_{Zj} + I_{Nj}$) is denominated in units of labor instead of final goods):

$$Y = C + [G_I + G_C - M]. \tag{38}$$

If the share of final goods (net of net imports) controlled by the government is constant, this reduces to a simple solution for aggregate consumption equaling a fraction of final goods, which can be rewritten as follows:

$$C = (1 - g) Y, \tag{39}$$

where g is the government's share of the final goods. This immediately implies that the growth rate of aggregate consumption equals the growth rate of output. Hence, the Euler equation becomes the following:

$$r_A = \left(\frac{1}{1 - \tau_A} \right) \left[\frac{\dot{Y}}{Y} + \frac{\dot{p_Y}}{p_Y} + \rho - \omega \right]. \tag{40}$$

In equilibrium, the same interest rate holds throughout the economy: $r = r_A = r_{Nj} = r_{Zj} \forall j$. Hence, we can substitute this into the household's budget constraint (without assets already solved) to obtain the following:

$$\frac{\dot{W}}{W} = [\rho - \omega] + \frac{(1 - \tau_L) L}{\theta p_Y Y} - \frac{(1 + \tau_C)(1 - g)}{\theta W}. \tag{41}$$

Substituting the individual's labor supply into L provides an equation relating the value of final-goods output to the wage rate:

$$\frac{\dot{W}}{W} = [\rho - \omega] + \frac{(1 - \tau_L) P}{\theta p_Y Y} - \frac{(1 + \tau_C)(1 - g)(1 + \lambda)}{\theta W}. \tag{42}$$

Because this is a general equilibrium model where only relative prices matter, we can normalize any single price to a positive constant. The obvious choices are the prices on final goods (p_Y) and time sold as labor (W). The former is more useful for our empirical analysis, while the latter is more useful for solving the model. Therefore, we set the numeraire to

the time sold as labor to solve the model, after which we can renormalize final goods to be the numeraire for conducting our empirical analysis. The most convenient value with which to set the numeraire's price is not actually 1 but rather the following:

$$W^* = \frac{(1+\tau_C)(1-g)(1+\lambda)}{\theta(\rho-\omega)+(1-\tau_L)}. \tag{43}$$

Substituting this into the previous equation reveals that, in these units, the value of output per capita equals 1:

$$\frac{p_Y Y}{P} = 1. \tag{44}$$

Hence, this measure of the value of output grows at the same rate as the population, which reduces the equilibrium interest rate to a constant:

$$r = \left(\frac{1}{1-\tau_A}\right)\rho. \tag{45}$$

Substituting this expression into the rate of return to knowledge accumulation and using the technology for knowledge accumulation, we can solve for the growth rate of knowledge (given positive investment in knowledge accumulation) as a function of the value of the industry's output and the number of firms in the industry:

$$\frac{\dot{K}_j}{K_j} = \frac{\kappa_j(R_j)L_{Zj}}{N_j} = \left[\frac{\kappa_j(R_j)\zeta_j(R_j)}{W^*\chi_j/(\chi_j-1)}\right]\frac{p_{Yj}Y_j}{N_j} - \frac{\delta_j(1-\tau_A)+\rho}{1-\tau_A}. \tag{46}$$

At this point, we invoke an additional assumption to diagonalize the system of differential equations that describe the model's growth dynamics. This effectively restricts the solution to the set of balanced growth paths because, otherwise, one particular industry could grow to a scale that dwarfs all others. By assuming that $\psi=1$, which is equivalent to a Cobb-Douglas technology in the aggregation of industry commodity outputs into final goods, the value of industry j's output is a constant fraction of the value of final goods:

$$p_{Yj}Y_j = \psi_j p_Y Y = \psi_j P. \tag{47}$$

Given this simplifying assumption, it is most convenient to work in the number of firms per capita (n_j) as the state variable:

$$n_j = \frac{N_j}{P}. \tag{48}$$

Using this simplifying assumption and state variable, as well as accounting for the nonnegativity constraint in the growth of knowledge, we can write the growth rate of industry j's stock of knowledge as a function of the number of firms per capita in industry j:

$$\frac{\dot{K}_j}{K_j} = \begin{cases} \left[\frac{\psi_j\kappa_j(R_j)\zeta_j(R_j)}{W^*\chi_j/(\chi_j-1)}\right]\left(\frac{1}{n_j}\right) - \left[\frac{\delta_j(1-\tau_A)+\rho}{(1-\tau_A)}\right] & n_j < \tilde{n}_j \\ 0 & n_j \geq \tilde{n}_j, \end{cases} \tag{49}$$

where $\tilde{n}_j$ is the critical mass of firms per capita, above which firms do not invest in innovation:

$$\tilde{n}_j = \frac{\psi_j\kappa_j(R_j)\zeta_j(R_j)(\chi_j-1)(1-\tau_A)}{\delta_j(1-\tau_A)\chi_j W^* + \rho\chi_j W^*}. \tag{50}$$

Substituting these terms into the rate of return to new entry and solving for the growth rate of n_j:

$$\frac{\dot{n}_j}{n_j} = \gamma_{j0}\left(1\{n_j<\tilde{n}_j\}\right) + \gamma_{j1}\left(1\{n_j<\tilde{n}_j\}\right)n_j, \tag{51}$$

where

$$\gamma_{j1}\left(1\{n_j<\tilde{n}_j\}\right) = \frac{(1-\tau_\pi)}{\theta\kappa_j(R_j)\psi_j}\left[1\{n_j<\tilde{n}_j\}\left[\frac{\rho+\delta_j(1-\tau_A)}{(1-\tau_A)}\right] - \kappa_j(R_j)\phi_j(R_j)\right]$$

$$\gamma_{j0}\left(1\{n_j<\tilde{n}_j\}\right) = \left(\frac{1-\tau_\pi}{\theta W^*}\right)\left[\frac{1}{\chi_j} - \frac{1\{n_j<\tilde{n}_j\}\zeta(R_j)}{\chi_j/(\chi_j-1)}\right] - \left[\frac{\rho+\delta_j(1-\tau_A)}{(1-\tau_A)}\right]. \tag{52}$$

This piecewise ordinary differential equation is of the Bernoulli class with a closed-form solution following logistic growth:

$$n_j = \frac{\gamma_{j0}\left(n_j\right) n_j\left(0\right)}{\left[\gamma_{j1}\left(n_j\right) n_j\left(0\right) + \gamma_{j0}\left(n_j\right)\right] e^{-\gamma_{j0} t} - \gamma_{j1}\left(n_j\right) n_j\left(0\right)}, \tag{53}$$

which ultimately converges to the following steady-state number of firms per capita:

$$n_j^* = \frac{-\gamma_{j0}\left(n_j\right)}{\gamma_{j1}\left(n_j\right)} = \frac{\psi_j \kappa_j\left(R_j\right) \left[\frac{1}{\chi_j} - \frac{1\{n_j < \tilde{n}_j\}\zeta(R_j)}{\chi_j/(\chi_j-1)}\right] - \left[\frac{\rho+(1-\tau_A)\delta_j}{(1-\tau_A)}\right]}{W^* \left[1\{n_j < \tilde{n}_j\}\left[\frac{\rho+\delta_j(1-\tau_A)}{(1-\tau_A)}\right] - \kappa_j\left(R_j\right)\phi_j\left(R_j\right)\right]}. \tag{54}$$

Assuming $n_j^* < \tilde{n}_j$ as a regularity condition on the parameters, which is consistent with us observing firms still investing in productivity enhancements when their industry is undergoing consolidation (i.e., reducing the number of firms per capita), the steady-state growth rate in knowledge accumulation is given by the following expression:

$$\frac{\dot{K}_j}{K_j}^* = \left[\frac{\psi_j \kappa_j\left(R_j\right) \zeta_j\left(R_j\right)}{W^* \chi_j/\left(\chi_j - 1\right)}\right]\left(\frac{1}{n_j^*}\right) - \left[\frac{\delta_j\left(1 - \tau_A\right) + \rho}{\left(1 - \tau_A\right)}\right]. \tag{55}$$

This solution uses an industry's number of firms per capita as the state variable in that industry's dynamics. By a simple transformation, we can recast the state variable as a scalar multiple of the per-firm revenue:

$$\left(\frac{1}{\psi_j}\right) \times \frac{p_{Yj} Y_j}{N_j} = \left(\frac{1}{\psi_j}\right) \times \frac{\psi_j p_Y Y}{N_j} = \frac{1}{N_j} \times \frac{p_Y Y}{P/P} = \frac{1}{N_j/P} \times \underbrace{\frac{p_Y Y}{P}}_{1} = \frac{1}{n_j}. \tag{56}$$

Although this transformed state variable may be less intuitive for some, it further simplifies the dynamics from logistic growth to the exponential growth exhibited by a piecewise linear differential equation:

$$\left[\frac{\dot{p_{Yj} Y_j}}{N_j}\right] = -\psi_j \times \gamma_{j1}\left(1\{n_j < \tilde{n}_j\}\right) - \gamma_{j0}\left(1\{n_j < \tilde{n}_j\}\right)\left[\frac{p_{Yj} Y_j}{N_j}\right]. \tag{57}$$

Regardless of our choice of state variable, the resulting solution is unchanged.

Appendix B. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.red.2020.03.004.

References

Aghion, Philippe, Algan, Yann, Cahuc, Pierre, Shleifer, Andrei, 2010. Regulation and distrust. The Quarterly Journal of Economics 125 (3), 1015–1049.

Alesina, Alberto, Ardagna, Silvia, Nicoletti, Giuseppe, Schiantarelli, Fabio, 2005. Regulation and investment. Journal of the European Economic Association 3 (4), 791–825.

Al-Ubaydli, Omar, McLaughlin, Patrick A., 2015. RegData: a numerical database on industry-specific regulations for all United States industries and federal regulations, 1997–2012. Regulation and Governance. https://doi.org/10.1111/rego.12107.

Ang, James B., Madsen, Jakob B., 2011. Can second-generation endogenous growth models explain the productivity trends and knowledge production in the Asian miracle economies? Review of Economics and Statistics 93 (4), 1360–1373.

Botero, Juan C., Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, Schleifer, Andrei, 2004. The regulation of labor. The Quarterly Journal of Economics 119 (4), 1339–1382.

Cai, J., Li, N., 2019. Growth through inter-sectoral knowledge linkages. The Review of Economic Studies 86 (5), 1827–1866.

Dawson, John W., Seater, John J., 2013. Federal regulation and aggregate economic growth. Journal of Economic Growth 18 (2), 137–177.

Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, 2002. The regulation of entry. The Quarterly Journal of Economics 117 (1), 1–37.

Djankov, Simeon, McLiesh, Caralee, Ramalho, Rita Maria, 2006. Regulation and growth. Economics Letters 92 (3), 395–401.

Ellig, Jerry, McLaughlin, Patrick A., 2011. The quality and use of economic analysis in 2008. Risk Analysis 32 (5), 855–880.

Gørgens, Tue, Paldam, Martin, Würtz, Allan, 2003. How Does Public Regulation Affect Growth? Working Paper No. 2003-14. Department of Economics, University of Aarhus, Aarus, Denmark.

Greenstone, Michael, 2002. The impacts of environmental regulations on industrial activity: evidence from the 1970 and 1977 Clean Air Act amendments and the census of manufactures. Journal of Political Economy 110 (6), 1175–1219.

Loayza, Norman V., Oviedo, Ana Maria, Servén, Luis, 2005. The Impact of Regulation on Growth and Informality: Cross-Country Evidence. Policy Research Working Paper No. 3623. World Bank, Washington, DC.

Madsen, Jakob B., 2008. Semi-endogenous versus Schumpeterian growth models: testing the knowledge production function using international data. Journal of Economic Growth 13 (1), 1–26.

Madsen, Jakob B., 2010. The anatomy of growth in the OECD since 1870. Journal of Monetary Economics 57 (6), 753–767.

Mandel, Michael, Carew, Diana G., 2013. Regulatory Improvement Commission: A Politically-Viable Approach to U.S. Regulatory Reform. Policy Memo. Progressive Policy Institute, Washington, DC.

McLaughlin, Patrick A., Sherouse, Oliver, 2019. RegData 2.2: a panel dataset on US federal regulations. Public Choice 180, 43–55.

Morgenstern, Richard D., Pizer, William A., Shih, Jhih-Shyang, 2002. Jobs versus the environment: an industry-level perspective. Journal of Environmental Economics and Management 43 (3), 412–436.

Nicoletti, Giuseppe, Scarpetta, Stefano, 2003. Regulation, productivity and growth: OECD evidence. Economic Policy 18, 9–72.

Peltzman, Sam, 1975. The effects of automobile safety regulation. Journal of Political Economy 83 (4), 677–725.

Peretto, Pietro F., 2007. Corporate taxes, growth and welfare in a Schumpeterian economy. Journal of Economic Theory 137, 353–382.

Peretto, Pietro F., Connolly, Michelle, 2007. The Manhattan metaphor. Journal of Economic Growth 12 (4), 329–350.

Stigler, George J., 1971. The theory of economic regulation. The Bell Journal of Economics and Management Science 2 (1), 3–21.

Walker, W. Reed, 2013. The transitional costs of sectoral reallocation: evidence from the Clean Air Act and the workforce. The Quarterly Journal of Economics 128 (4), 1787–1835.