Profit Incentives and Technological Change

by

Joshua Abraham Linn

B.A., Yale University, 2000

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2005

Signature of Author ....
Department of Economics
15 May 2005

Certified by ....
Daron Acemoglu
Charles P. Kindleberger Professor of Economics
Thesis Supervisor

Certified by ....
Michael Greenstone
3M Associate Professor of Economics
Thesis Supervisor

Accepted by ....
Peter Temin
Elisha Gray II Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Profit Incentives and Technological Change

by

Joshua Abraham Linn

Submitted to the Department of Economics
on 15 May 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis is a collection of three empirical essays on the effect of profit incentives on innovation and technology adoption.

Chapter 1, written with Daron Acemoglu, investigates the effect of (potential) market size on entry of new drugs and pharmaceutical innovation. Focusing on exogenous changes driven by U.S. demographic trends, we find a large effect of potential market size on the entry of non-generic drugs and new molecular entities. These effects are generally robust to controlling for a variety of supply-side factors and changes in the technology of pharmaceutical research.

Chapter 2 investigates the effect of price-induced technology adoption on energy demand in U.S. manufacturing. I use plant data from the Census of Manufactures, 1967-1997, and identify technology adoption by comparing the energy efficiency of entrants and incumbents. I find a statistically significant effect of technological change, though the magnitude is small relative to changes in energy use due to factor substitution. The results suggest that technological change can reduce the long run effect of energy prices on growth, but by significantly less than previous research has suggested.

Chapter 3 studies the response of the manufacturing sector to a carbon tax. I estimate long run price elasticities for fuels and electricity, exploiting the ability of entering plants to choose their technology in response to expected prices. A tax of $10 per metric ton of carbon would reduce emissions by 2 percent and raise operating costs by 8 percent in the short run. Emissions would be 5 percent lower in the long run, and costs would be 5 percent higher. The tax would make plants more vulnerable to subsequent natural gas and distillate oil price shocks, and less sensitive to coal, residual and electricity shocks. Exit would increase by 0.2 percentage points.

Thesis Supervisor: Daron Acemoglu
Title: Charles P. Kindleberger Professor of Economics

Thesis Supervisor: Michael Greenstone
Title: 3M Associate Professor of Economics

Acknowledgements

I have benefited tremendously from the advice and guidance of my two advisors, Daron Acemoglu and Michael Greenstone. Their patience and enthusiasm for teaching have been inspirational, and I hope to share these attributes with my advisees in the future. They have helped me in ways too numerous to list here, which together have made me a more mature economist.

This thesis has benefited from the suggestions of a number of other people. For the first chapter, Daron and I received helpful comments from Joshua Angrist, David Autor, Ernst Berndt, Olivier Blanchard, Alfonso Gambardella, Amy Finkelstein, Stanley Finkelstein, Jerry Hausman, Paul Joskow, Frank Lichtenberg, Lawrence Katz, Michael Kremer, David Popp, Jeffrey Wooldridge, two anonymous referees, and many seminar participants. The research was partially funded by the National Science Foundation and the Russell Sage Foundation.

For the second and third chapters, I wish to thank Paul Joskow, whose knowledge and experience with energy and industrial organization has forced me to consider many important issues I would not have otherwise. In addition to my advisors, I have benefited from comments by Joshua Angrist, David Autor, Ernie Berndt, Richard Eckaus, Denny Ellerman, David Popp and Ian Sue Wing, and numerous seminar participants. I am grateful for financial support from the Center for Energy and Environmental Policy Research, and the Joint Program on the Science and Policy of Global Change.

On a personal level, graduate school would have been a much more difficult experience had it not been for the many friends I have made, to share in the various triumphs and failures. The spirit of camaraderie and support among my classmates has been wonderful. I must thank my family, and especially my wife, Katherine, for all they have put up with, and all they have done to support me.

Contents

Chapter 1

Innovation and Market Size: Theory and Evidence from the Pharmaceutical Industry

Summary 1 *This chapter, written with Daron Acemoglu, investigates the effect of (potential) market size on entry of new drugs and pharmaceutical innovation. Focusing on exogenous changes driven by U.S. demographic trends, we find a large effect of potential market size on the entry of non-generic drugs and new molecular entities. These effects are generally robust to controlling for a variety of supply-side factors and changes in the technology of pharmaceutical research.*

1.1 Introduction

This paper constructs a simple model linking innovation rates to current and future market size, and documents the empirical relationship between market size and innovation in the pharmaceutical industry. Our empirical work, which exploits changes in the market size for different drug categories driven by U.S. demographic trends, finds economically significant and relatively robust effects of market size on innovation.

Although many historical accounts of important innovations focus on the autonomous progress of science and on major breakthroughs that take place as scientists build on each other's work, economists typically emphasize profit incentives and the size of the target market. For example, in his seminal study, *Invention and Economic Growth,* Schmookler argued that: "...invention is largely an economic activity which, like other economic activities, is pursued for gain" (1966, p. 206). To emphasize the role of market size, Schmookler entitled two of his chapters "The amount of invention is governed by the extent of the market."

The role of profit incentives and market size in innovation is also important both for the recent endogenous technological change models, which make profit incentives the central driving force of the pace of aggregate technological progress (e.g., Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1992)), and for the induced innovation and directed technical change literatures, which investigate the influence of profit incentives on the types and biases of new technologies (see, e.g., Kennedy (1964), Drandkis and Phelps (1965), Samuelson (1965), Hayami and Ruttan (1970) and Acemoglu (1998, 2002, 2003)). A recent series of papers by Kremer, for example (2002), also builds on the notion that pharmaceutical research is driven by market size and argues that there is generally insufficient research to develop cures for third-world diseases such as malaria, because those who suffer from these diseases have a limited ability to pay.

In this paper, we investigate the effect of market size on drug entry and pharmaceutical innovation. A major difficulty in any investigation of the impact of market size on innovation is the endogeneity of market size—better products will have larger markets. Our strategy to overcome this problem is to exploit variations in market size driven by U.S. demographic changes, which should be exogenous to other, for example scientific, determinants of innovation and entry of new drugs.[1] To estimate potential market size, we construct age profiles of users for each drug category, and then compute the implied market size from aggregate demographic and income changes given these (time-invariant) age profiles.[2] We measure entry and innovation using the Food and Drug Administration's (FDA) approval of new drugs.[3]

[1] For many drugs non-U.S. markets may also be relevant. Nevertheless, the U.S. market is disproportionately important, constituting about 40 percent of the world market (IMS (2000)). Below, we report results using changes in OECD market size as well as U.S. market size.

[2] Loosely speaking, "market size" corresponds to the number of users times their marginal willingness to pay. Therefore, market size can increase both because the number of users increases and because their marginal willingness to pay changes. We focus on changes driven by demographics to isolate exogenous changes in market size.

[3] These data were previously used by Lichtenberg and Virahbak (2002), who obtained them under the Freedom

Our results show that there is an economically and statistically significant response of the entry of new drugs to market size. As the baby boom generation aged over the past 30 years, the markets for drugs mostly consumed by the young have declined and those for drugs consumed by the middle-aged have increased. The data show a corresponding decrease in the rate of entry of new drugs in categories mostly demanded by the young and an increase for drugs mostly consumed by the middle-aged. Our estimates suggest that a 1 percent increase in the size of the potential market for a drug category leads to a 6 percent increase in the total number of new drugs entering the U.S. market. Much of this response comes from the entry of generics, which are drugs that are identical or bioequivalent to an existing drug no longer under patent protection.

More important, there is a statistically significant response of the entry of non-generic drugs, which more closely correspond to new products and "innovation": a 1 percent increase in potential market size leads to approximately a 4 percent increase in the entry of new non-generic drugs. We also look at the relationship between market size and entry of new molecular entities. These drugs, which contain active ingredients that have not been previously marketed in the United States, provide a measure of more radical innovations (there are 442 new molecular entities compared to 2,203 new non-generics during our sample period). We find that a 1 percent increase in potential market size is associated with a 4-6 percent increase in the entry of new molecular entities. These results together show an important effect of potential market size on pharmaceutical innovation.

The effect of market size on the entry of new drugs is generally robust. We obtain similar results when we use different measures of market size, when we exploit changes in OECD market size, and when we control for a variety of supply-side factors including advances in biotechnology.

We also investigate whether it is current market size or past or future market sizes that have the largest effect on entry of new drugs. On the one hand, because changes in demographics can be anticipated in advance, drug entry may respond to future market size. On the other hand, because there is typically a 10-15 year gap between research and FDA approval (e.g., DiMasi et al. (1991)), entry may respond to past market size. We find that all non-generics respond to

of Information Act. We thank Frank Lichtenberg for sharing these data with us.

current market size, while current and five-year lead of market size have the strongest effects on new molecular entities and generics. These results suggest that pharmaceutical research responds to anticipated changes in market size with a lead of 10-20 years.

The magnitude of the effect of potential market size on drug entry is quite large. This may be partly because our key variable measures potential market size rather than actual market size. (i.e., what the market size would be if the number and incomes of individuals in a particular age group change without a change in the age profile of use and expenditure). Results using another data source suggest that a 1 percent increase in potential market size is associated with approximately a 4 percent increase in actual market size, so the estimates for non-genetics and new molecular entities are consistent with a proportional effect of actual market size on innovation as predicted by our theoretical model.[4]

There are a number of other studies related to our work. First, Schmookler (1966) documents a correlation between sales and innovation, and argues that the causality ran largely from the former to the latter. The classic study by Griliches (1957) on the spread of hybrid seed corn in U.S. agriculture also provides evidence consistent with the view that technological change and technology adoption are closely linked to profitability and market size. Pakes and Schankerman (1984) investigate this issue using a more structural approach, linking R&D intensity at the industry level to factor demands and to growth of output. In more recent research, Scott Morton (1999) and Reiffen and Ward (2004) study the decision of firms to introduce a generic drug and find a positive relationship between entry and expected revenues in the target market. None of these studies exploit a potentially exogenous source of variation in market size, however.

Second, some recent research has investigated the response of innovation to changes in energy prices. Most notably, Newell, Jaffee and Stavins (1999) show that between 1960 and 1980, the typical air-conditioner sold at Sears became significantly cheaper, but not much more energy-efficient. On the other hand, between 1980 and 1990, there was little change in costs, but air-conditioners became much more energy-efficient, which, they argue, was a response to higher energy prices. In a related study, Popp (2002) finds a strong positive relationship

[4]It is also possible that the marginal innovations induced by an increase in market size are less productive, so a 4-6 percent increase in the number of new drugs may correspond to a smaller increase when weighted by effectiveness or other measures of productivity.

between patents for energy-saving technologies and energy prices.

Third, there is substantial research focusing on innovation in the pharmaceutical industry. Henderson and Cockburn (1996), Cockburn and Henderson (2001), and Danzon, Nichelson and Sousa Pereira (2003) study the determinants of success in clinical trials, focusing mainly on firm and project size. Galambos and Sturchio (1998), Cockburn, Henderson and Stern (1999), Gambardella (2000), Malerba and Orsenigo (2000), and Ling, Berndt and Frank (2003) discuss various aspects of the recent technological developments in the pharmaceutical industry.

Most closely related to this study are Lichtenberg and Waldfogel (2003), Finkelstein (2004), Cerda (2003), and DellaVigna and Pollet (2004). Lichtenberg and Waldfogel (2003) document a relative decline in mortality among individuals with rare diseases following the Orphan Drug Act, and argue that this is related to the incentives created by the Act to develop drugs for these conditions. Finkelstein (2004) exploits three different policy changes affecting the profitability of developing new vaccines against 6 infectious diseases: the 1991 Center for Disease Control recommendation that all infants be vaccinated against hepatitis B, the 1993 decision of Medicare to cover the costs of influenza vaccinations, and the 1986 introduction of funds to insure manufacturers against product liability lawsuits for certain kinds of vaccines. She finds that increases in vaccine profitability resulting from these policy changes are associated with a significant increase in the number of clinical trials to develop new vaccines against the relevant diseases.[5] Cerda's (2003) Ph.D. dissertation at Chicago is an independent study of the effect of demographics on innovation in the pharmaceutical sector. Although Cerda uses a somewhat different empirical methodology, he reaches similar conclusions to our study. Finally, DellaVigna and Pollet (2004) investigate whether the stock market responds to demographics-driven changes in the size of the market for a number of products.

The rest of the paper is organized as follows. We outline a simple model linking innovation to market size in the next section. Section 1.3 briefly explains our empirical strategy, and Section 1.4 describes the basic data sources and the construction of the key variables. Section 1.5 presents the empirical results and a variety of robustness checks. Section 1.6 contains some concluding remarks, and the Appendix gives further data details.

[5] Lichtenberg (2003) also presents evidence suggesting that the types of new drugs changed towards drugs more useful for the elderly after Medicare was established.

1.2 Theory

This section outlines a simple model that illustrates the impact of market size on innovation. We consider a small open economy consisting of a set I of infinitely-lived individuals. Time is continuous $t \in [0, \infty)$. There are two types of goods in this economy. First, there is a basic good, y, which can be consumed, used for the production of other goods, or for research expenditure. Individual i has an exogenously given endowment $y_i(t)$ at time t. Second, there are J drugs, $x_1, \ldots, x_J$, each with a potentially time-varying "quality", $q_1(t), \ldots, q_J(t)$. Each individual demands only one type of drug. Hence, we partition the set I of individuals into J disjoint groups, $G_1, \ldots, G_J$ with $G_1 \cup G_2 \cup \ldots \cup G_J = I$, such that if $i \in G_j$, then individual i demands drug j. More specifically, if $i \in G_j$, then his preferences are given by

$$\int_0^\infty \exp(-rt) \left[c_i(t)^{1-\gamma} \left(q_j(t) \, x_{ji}(t) \right)^\gamma \right] dt, \tag{1.1}$$

where r is the discount rate of the consumers and the interest rate faced by the economy, $\gamma \in (0,1)$, $c_i(t)$ is the consumption of individual i of the basic good at time t, and $x_{ji}(t)$ is the consumption of drug j. This Cobb-Douglas functional form and the assumption that each individual only consumes one type of drug are for simplicity and do not affect the main results.[6]

Normalizing the price of the basic good to 1 in all periods, and denoting the price of drug j at time t by $p_j(t)$, the demand of individual $i \in I$ for drug j is $x_{ij}(t) = \gamma y_i(t) / p_j(t)$ if $i \in G_j$, and $x_{ij}(t) = 0$ if $i \notin G_j$. Summing across individuals, total demand for drug j is

$$X_j(t) = \frac{\gamma Y_j(t)}{p_j(t)}, \tag{1.2}$$

where $Y_j(t) \equiv \sum_{i \in G_j} y_i(t)$ is the total income of the group of individuals consuming drug j.

At any point in time, there is one firm with the best-practice technology for producing each type of drug, and it can produce one unit of this drug with quality $q_j(t)$ using one unit of the basic good. If there is an innovation for drug line j currently with quality $q_j(t)$, this leads to the discovery of a new drug of quality $\lambda q_j(t)$ where $\lambda > 1$. For the purposes of the model,

[6] The Cobb-Douglas assumption implies that the share of income spent on drugs is constant. This assumption can easily be relaxed by considering a utility function with an elasticity of substitution different from 1, as in the factor market models with directed technical change (see, e.g., Acemoglu (1998, 2002)).

we think that any new innovation is approved (for example by the FDA) and can be sold to consumers immediately (and is under patent protection indefinitely).

There is free entry into R&D and each firm has access to an R&D technology that generates a flow rate δ_j of innovation for every dollar spent for research on drug j. So if R&D expenditure at time t is $z_j(t)$, the flow rate of innovation (and of entry of new drugs) for drug j is

$$n_j(t) = \delta_j z_j(t). \tag{1.3}$$

Differences in δ_j's introduce the possibility that technological progress is scientifically more difficult in some lines than others.

A key feature of this R&D technology for our focus is that research is *directed* in the sense that firms can devote their R&D to developing particular types of drugs. The pharmaceutical industry, especially in the recent past, is a prime example of an industry where companies with fairly sophisticated R&D divisions or specialized R&D firms can undertake research for specific drug lines (e.g., Gambardella (2000) and Malerba and Orsenigo (2000)).[7]

The demand curves in (1.2) have an elasticity equal to 1, so an unconstrained monopolist would charge an arbitrarily high price. However, the firm with the best drug in line j is competing with the next best drug in that line. An arbitrarily high price would allow the next best firm to capture the entire market. Therefore, the firm with the best drug sets a *limit* price to exclude the next best firm—i.e., to ensure that consumers prefer its product rather than the next best drug supplied at the lowest possible price (i.e. equal to marginal cost, which is 1). If a consumer buys from the best-practice firm with quality $q_j(t)$ and price $p_j(t)$ and chooses her optimal consumption as given by (1.2), her instantaneous utility at time t is $(q_j(t))^\gamma (1-\gamma)^{1-\gamma} \gamma^\gamma (p_j(t))^{-\gamma} y_i(t)$; and if she purchases from the next best firm, at quality $q_j(t)/\lambda$ and price equal to marginal cost, 1, she obtains utility $\lambda^{-\gamma} (q_j(t))^\gamma (1-\gamma)^{1-\gamma} \gamma^\gamma y_i(t)$.

[7]Naturally, there exist examples of research directed at a specific drug type leading to the discovery of a different product, such as the well-known example of Viagra, which resulted from research on hypertension and angina, and was partly accidentally discovered from the detection of side effects in a clinical study (see, e.g., Kling (1998)). The working paper version, Acemoglu and Linn (2003) shows that the results here generalize even when there is a large component of random R&D, whereby research directed at drug j can result in the discovery of other drugs.

The limit price, which equalizes these two expressions, is

$$p_j(t) = \lambda \text{ for all } j \text{ and } t. \tag{1.4}$$

The profits of the firm with the best product of quality $q_j(t)$ in line j at time t are

$$\pi_j(q_j(t)) = (\lambda - 1)\gamma Y_j(t). \tag{1.5}$$

Here $\lambda\gamma Y_j(t)$ corresponds to the market size (total sales) for drug j. Notice that profits of drug companies are independent from quality, $q_j(t)$, which is a feature of the Cobb-Douglas utility.

Firms are forward-looking, and discount future profits at the rate r. The discounted value of profits for firms can be expressed by a standard dynamic programming recursion. $V_j(t \mid q_j)$, the value of a firm that owns the most advanced drug of quality q_j in line j at time t, is[8]

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j(t)) - \delta_j z_j(t) V_j(t \mid q_j), \tag{1.6}$$

where $\pi_j(q_j(t))$ is the flow profits given by (1.5), and $z_j(t)$ is R&D effort at time t on this line by other firms.[9] Intuitively, the value of owning the best technology in line j, $rV_j(t \mid q_j)$, is equal to the flow profits, $\pi_j(q_j(t))$, plus the potential appreciation of the value, $\dot{V}_j(t \mid q_j)$, and takes into account that at the flow rate $n_j(t) = \delta_j z_j(t)$ there will be a new innovation, causing the current firm to lose its leading position and to make zero profits thereafter.

Free entry into R&D to develop better quality drugs implies zero profits; i.e.,

$$\text{if } z_j(t) > 0, \text{ then } \delta_j V_j(t \mid q_j) = 1 \text{ for all } j \text{ and } t \tag{1.7}$$

(and if $z_j(t) = 0$, $\delta_j V_j(t \mid q_j) \leq 1$ and there will be no equilibrium R&D for this drug).

An equilibrium in this economy is given by sequences of prices $p_j(t)|_{j=1,..J}$ that satisfy (1.4), consumer demands for drugs $x_i(t)|_{i\in I}$ that satisfy (1.2) and R&D levels $z_j(t)|_{j=1,..J}$ that satisfy (1.7) with $V_j(\cdot)$ given by (1.6).

[8] Throughout, we assume that the relevant transversality conditions hold and discounted values are finite.

[9] Because of the standard replacement effect first emphasized by Arrow (1963), the firm with the best technology does not undertake any R&D itself (see, for example, Aghion and Howitt (1992)).

An equilibrium is straightforward to characterize. Differentiating equation (1.7) with respect to time implies $\dot{V}_j(t \mid q_j) = 0$ for all j and t, as long as $z_j(t) > 0$. Substituting this equation and (1.7) into (1.6) yields the levels of R&D effort in the unique equilibrium as

$$z_j(t) = \max\left\{\frac{\delta_j(\lambda - 1)\gamma Y_j(t) - r}{\delta_j}; 0\right\} \text{ for all } j \text{ and } t. \tag{1.8}$$

Equation (1.8) highlights the market size effect in innovation: the greater is $Y_j(t)$, i.e., the greater is the market size for a particular drug, the more profitable it is to be the supplier of that drug, and consequently, there will be greater research effort to acquire this position. In addition, a higher productivity of R&D as captured by δ_j also increases R&D, and a higher interest rate reduces R&D since current R&D expenditures are rewarded by future revenues.

Another important implication of this equation is that there are no transitional dynamics. At any point in time, R&D for a particular drug line is determined by the current market size—past and anticipated future market sizes do not affect current research effort. This is an implication of the linear R&D technology, which ensures that whenever there are profit opportunities, there will immediately be sufficient R&D to arbitrage them, ensuring $\dot{V}_j(t \mid q_j)=0$. The intuition for the lack of response to anticipated changes in future market size highlights an important effect in quality ladder models of technological progress: firms would like to own the best-practice product at the time the market size actually becomes larger. Investing in R&D far in advance of the increase in market size is not profitable, since another firm would improve over this innovation by the time the larger market size materializes. In fact, with the linear model here, z_j can change discontinuously, so investing even a little in advance of the actual increase in the size of the market is not profitable.

Combining equations (1.3) and (1.8) gives entry of new drugs as

$$n_j(t) = \max\left\{\delta_j(\lambda - 1)\gamma Y_j(t) - r; 0\right\}. \tag{1.9}$$

This equation relates innovation or entry of new products to market size (total expenditure of consumers in this line of drug). It also encompasses the alternative view of the determinants of innovation discussed in the Introduction, that the cross-drug distribution of R&D is determined by technological research opportunities or perhaps by other motives unrelated to profits. If there are large and potentially time-varying differences in δ_j's, then these may be the primary factor determining variation in R&D across drug lines, and market size may have only a small effect. Whether or not this is so is an empirical question.

The working paper version of our paper, Acemoglu and Linn (2003), presented a number of generalizations of this framework. First and most importantly, we modified the R&D technology captured in equation (1.3) to allow for within-period decreasing returns, so that

$$n_j(t) = \delta_j z_j(t) \phi(z_j(t)),$$

where $\phi'(z) \leq 0$ (the model studied above is the special case with $\phi'(z) \equiv 0$). Most of the results here generalize, but the model also implies a potential response to anticipated changes in market size. In particular, let us assume that $Y_j(t) = Y_j$ for all t. Then it is straightforward to show that steady-state R&D will be given by

$$z_j^S = \max\left\{\frac{\left(\delta_j \phi\left(z_j^S\right)(\lambda - 1)\gamma Y_j - r\right)}{\delta_j \phi\left(z_j^S\right)}; 0\right\},$$

which is similar to (1.8). If there is an unanticipated change in Y_j, there continues to be no transitional dynamics (i.e., z_j immediately jumps to its new steady-state value). But it can be shown that if there is an *anticipated* increase in market size, there will be entry of new drugs in advance of the actual increase. Nevertheless, the same forces emphasized here imply that investing in R&D too far in advance would not be profitable because another firm is likely to innovate further before the actual increase in market size materializes. In terms of our empirical work, even if demographic changes are anticipated 20 or 30 years in advance, we may expect significant entry and innovation responses much later, perhaps 5 or 10 years in advance.

Second, we extended this model to incorporate entry of both generic and non-generic drugs and showed that market size has a positive effect on entry of both types of drugs, and that, under plausible circumstances, the effect of market size on generic entry should be larger than on non-generics.

1.3 Empirical Strategy

1.3.1 Empirical Specification and Estimation Issues

As $r \to 0$, equation (1.9) implies that $n_j(t)$ is proportional to $\delta_j m_j(t)$, where $m_j(t) \equiv \lambda\gamma Y_j(t)$ is the market size for drug line j at time t. We measure entry of new drugs (or innovation), $n_j(t)$, as new drug approvals by the FDA in broad drug categories as described below. This measure, denoted by N_{ct} for drug category c at time t, includes entry of generic drugs. Although generic drugs do not correspond to "innovation", their entry is driven by the same profit incentives as innovation. After presenting results using all drug approvals, our analysis focuses on the relationship between market size and entry of non-generics and new molecular entities. Non-generics include all drugs that are not identical or bioequivalent to an existing drug, while new molecular entities are drugs classified by the FDA as containing an active ingredient previously not marketed in the United States. Throughout, instead of actual market size, we use potential market size driven by demographic changes, which we denote by M_{ct}. The construction of this variable is discussed below.

Adding other potential determinants, time effects and rearranging, equation (1.9) yields a Poisson model for the conditional mean of new drugs (see Wooldridge [2002])

$$E\left[N_{ct} \mid \zeta_c, \bar{X}_c\right] = \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t), \tag{1.10}$$

where E is the expectations operator, N_{ct} is the number of new drugs in category c in time period t, M_{ct} is potential market size, X'_{ct} is a vector of controls, including a constant, ζ_c's are a full set of category fixed effects that correspond to the δ_j terms above, μ_t's are a full set of time effects capturing any common time component, and finally, $\bar{X}_c$ is the vector $((M_{c1}, .., M_{cT}) \vdots X'_{c1} \vdots \; .. \vdots X'_{cT} \vdots (\mu_1, .., \mu_T))$, with T denoting the number of time periods in our sample. This specification ensures that time effects have proportional impacts on entry of new drugs. Note also that this equation allows the coefficient of $\log M_{ct}$ to differ from 1, which could be the case if actual market size differs systematically from the potential market size, $\log M_{ct}$, or if preferences are not Cobb-Douglas (see Acemoglu (1998, 2002)).

The estimation of (1.10) would lead to biased estimates, however, since the nonlinearity in (1.10) makes it impossible to estimate the fixed effects, the ζ_c's, consistently. To deal with this problem, we follow Hausman, Hall, and Griliches (1984), and transform (1.10) to obtain a multinomial distribution for N_{ct} of the form

$$E\left[N_{ct} \mid \zeta_c, \bar{X}_c, \bar{N}_c\right] = \frac{\exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t)}{\sum_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_\tau)} \bar{N}_c, \tag{1.11}$$

where $\bar{N}_c = \sum_{\tau=1}^{T} N_{c\tau}$ is the number of drugs approved in category c over the entire sample. This transformation removes the drug category dummies, and the coefficient of interest, α, can be estimated consistently. We estimate this equation using quasi-maximum likelihood (QML). Wooldridge (1999) shows that QML has good consistency properties, even when the true model is not Poisson, for example when there is a different distribution of the error term.[10]

Below, we estimate a linear model of the form

$$\log \tilde{N}_{ct} = \alpha \cdot \log M_{ct} + d_{ct} + X'_{ct} \cdot \beta + \zeta_c + \mu_t + \varepsilon_{ct}, \tag{1.12}$$

where the left-hand side variable is defined as $\tilde{N}_{ct} = N_{ct}$ if $N_{ct} \geq 1$ and $\tilde{N}_{ct} = 1$ if $N_{ct} = 0$, and d_{ct} is a dummy that equals 1 when $N_{ct} = 0$. This procedure, first used by Pakes and Griliches (1980), is simple and flexible, but the estimates are biased, since d_{ct} is endogenous.

In addition, we estimate equations with lags and leads of $\log M_{ct}$ to determine whether there are significant delays and anticipation effects. Delayed effects are possible, since, as reported by DiMasi et al. (1991), drug approval may be as much as 15 years after the time of initial research. Anticipation effects are possible, since changes in demographics can be anticipated a long time in advance (see the discussion in Section 1.2).[11]

---

[10] Define the vector $\eta \equiv (\alpha \vdots \beta \vdots \mu)$. Then the QML estimate $\hat{\eta}$ maximizes the log likelihood function $\sum_{c=1}^{C} L_c(\hat{\eta})$ where C is the number of categories, $L_c(\hat{\eta}) \equiv \sum_{t=1}^{T} N_{ct} \log p_t(\hat{\eta})$, and $p_t(\hat{\eta}) \equiv \exp(\alpha \cdot \log M_{ct} + X'_{ct} \cdot \beta + \mu_t) / \left\{\sum_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + X'_{c\tau} \cdot \beta + \mu_\tau)\right\}$.

The (Huber-White) robust asymptotic variance-covariance matrix is calculated as $\hat{A}^{-1}\hat{B}\hat{A}^{-1}/C$, where $\hat{A} \equiv C^{-1}\sum_{c=1}^{C} \bar{N}_c \nabla_\eta p'_t(\hat{\eta}) D_c(\hat{\eta}) \nabla_\eta p_t(\hat{\eta})$, $\hat{B} \equiv C^{-1}\sum_{c=1}^{C} \nabla_\eta p'_t(\hat{\eta}) D_c(\hat{\eta}) \hat{u}_c \hat{u}'_c D_c(\hat{\eta}) \nabla_\eta p_t(\hat{\eta})$, $p(\hat{\eta}) \equiv [p_1(\hat{\eta}), \ldots, p_T(\hat{\eta})]'$, and $D_c(\hat{\eta}) \equiv [diag\{p_1(\hat{\eta}), \ldots, p_T(\hat{\eta})\}]^{-1}$. Here ∇_η denotes the gredient with respect to η and $\hat{u}_c$ is the vector of residuals calculated as $\hat{u}_c = N_c - p(\hat{\eta})\bar{N}_c$, with $N_c \equiv [N_{c1}, .., N_{cT}]$. See Wooldridge (1999) for more details.

[11] An additional issue is that the FDA approval process may be faster for more profitable drugs, and thus potentially for drugs with greater market size (see Dranove and Meltzer (1994)). Our data do not enable us to investigate this issue.

1.3.2 Potential Market Size and Identification

Throughout, we exploit the potentially exogenous component of market size driven by demographic trends, combined with differences in the age profiles of expenditure and use for different types of drugs. We obtain the age profiles from micro drug consumption data, and the changes in U.S. demographics from the Current Population Survey (CPS) data. Our (income-based) measure of potential market size is

$$M_{ct} = \sum_{a} u_{ca} \cdot i_{at}, \tag{1.13}$$

where i_{at} is the income of individuals in age group a at time t in the United States, and u_{ca} gives the age profile for drug category c. We compute u_{ca} as the average expenditure share of drugs in category c in the total income of those in age group a. This income-based measure corresponds closely to the market size in the theoretical model, which is a combination of the number of consumers and their incomes. We also check the robustness of our results with an alternative population-based measure, calculated using the U.S. population for age group a at time t for i_{at}, and the average number of drugs in category c used per person in age group a for u_{ca}. It is important that the over-time source of variation in both measures is not from changes in individual use, but purely from demographic changes captured by i_{at}—i.e., u_{ca}'s are not time-varying.[12] Consequently, changes in prices and drug quality, which may result from innovations and affect consumption patterns, will not cause over-time variation in M_{ct}. Our baseline measure uses five-year age groups and time periods corresponding to five-year intervals. We also check the robustness of our results using single year age groups and ten-year intervals.

The major threat to the validity of our empirical strategy is from potentially time-varying omitted variables (the drug category fixed effects take out any variable that is not time-varying). Omitted variables related to market size or profit opportunities may induce a bias in the implied magnitudes, but will not lead to spurious positive estimates of the effect of market size (in

[12] Because of data availability, we cannot use estimates of u_{ca} that pre-date our sample period. Therefore, our estimates of age profiles may have been affected by the availability of new drugs during the sample. This should not create a spurious relationship between potential market size, M_{ct}, and entry of new drugs, since all variation in M_{ct} is driven by aggregate demographic changes, and all of our regressions control for drug category fixed effects. In any case, the numbers in Table I suggest that age profiles do not change much over time (in fact, if preferences are Cobb-Douglas as in (1.1) and stable, the expenditure measure of u_{ca} should be constant).

other words, the presence of such variables is essentially equivalent to mismeasurement of the appropriate market size). More threatening to our identification strategy would be omitted supply-side variables that are potentially correlated with our market size measure. To show that this is not the source of our findings, we check for residual serial correlation and control for potential supply-side determinants of innovation and entry.[13]

1.4 Data and Descriptive Statistics

The demographic data come from the March CPS, 1965-2000. We compute i_{at} in equation (1.13) for five-year age groups, ranging from 0-4 to 90+. Individual income is constructed by dividing family income equally among the members of the family. For the purposes of the diagrammatic presentation, we aggregate the age groups into three broad categories, 0-30, 30-60 and 60+, corresponding to young, middle-aged and elderly users. Income and population movements of the five-year age groups within each of these broad groups are relatively similar.

Figure 1.1 shows population shares, and Figure 1.2 shows the corresponding income shares (i.e., income of the corresponding age group divided by total income in that period) for the three broad age groups. To facilitate comparison with Figure 1.3, Figure 1.2 starts in 1970. Both figures show a large amount of variation across age groups over time. In particular, it is possible to trace the baby boomers, as the fraction of individuals in the age bracket 0-30 in the 1970s, and those in the age bracket 30-60 in the 1980s and the 1990s.

The FDA classifies all prescription drugs into 20 major drug categories, which are further subdivided into 159 categories. These categories are based on a combination of therapeutic intent and chemical structure. We drop 4 of the 20 major categories from this classification: Anesthetics, Antidotes, Radiopharmaceuticals and Miscellaneous.[14] We then subdivide some

[13] Another source of endogeneity may be that innovations in certain drug categories extend the lives of the elderly, thus increasing their M_{ct}. Lichtenberg (2002, 2003) provides evidence that new drugs extend lives. This source of endogeneity is not likely to be quantitatively important, however, since the variation resulting from extended lives in response to new drugs is a small fraction of the total variation in M_{ct}. Nevertheless, we also report estimates that instrument M_{ct} with past demographics, purging it from changes in longevity.

[14] We drop the Anesthetics, Radiopharmaceuticals and Miscellaneous categories because most of the items in these categories were not developed for a distinct market. Radiopharmaceuticals are used for diagnostic purposes, and the Miscellaneous category mainly contains surgical and dental tools. The Antidote category is dropped because there are few drugs approved and there is little use of these drugs in the surveys. See the Data Appendix for further details on the contruction of our categories.

of these categories according to the conditions and diseases that the drugs are used to treat.[15] For example, within the Hematologics major category, we separate Anemia drugs from Anticoagulants because they treat different diseases. We also subdivide broader groups when the age distribution of expenditure is sufficiently heterogeneous. For example, the indications of drugs in Estrogens/Progestins and Contraceptives overlap somewhat, but the age structure of users is quite different: 20-30 year-olds use Contraceptives most, while 50-60 year-olds use Estrogens/Progestins most. In one case, we combine categories from different major classes, Antifungals and Dermatologics, because the drugs have similar indications and age distributions. The result is a classification system with 33 categories, which are listed in the Appendix Table.[16]

Our main data source for drug use is the Medical Expenditure Panel Survey (MEPS), which is a sample of U.S. households over the years 1996-1998. The survey has age and income data for each household member, and covers about 28,000 individuals each year. There is also a list of prescription drugs used by each person (if any), and the amount spent on drugs, which includes copayments and payments by insurance companies and government programs (e.g., Medicaid and worker's compensation).[17] In all, there are about 500,000 medications prescribed. We compute drug expenditure and use by five-year age groups, then divide these by the corresponding income and population numbers to construct the income-based and the population-based measures of u_{ca}.[18] The Appendix Table reports these numbers aggregated to the three broad age groups used in the figures. This table shows a large amount of variation in the age profiles of expenditure across the 33 drug categories. The elderly spend more on many categories than do younger individuals, but there are numerous exceptions. For example, Antibiotics are used most by individuals in the youngest group, while Contraceptives and Antivirals are used most

[15] Other authors, for example Lichtenberg (2003), have used a more detailed classification system based on diseases. We were unable to construct a comprehensive mapping of the prescription drugs listed in the microdata surveys to the detailed disease classes. Our classification system relies on the FDA categories, but then subdivides those according to disease and age distribution.

[16] The working paper version, Acemoglu and Linn (2003), used a system with 34 categories constructed purely based on differences in the age profiles of expenditure within the major FDA categories. Results using this alternative classification are reported in Table 1.3. Further details on the construction of the 33 categories used here and on our alternative classification system are available upon request.

[17] Respondents list the pharmacy or medical provider where they obtained the prescription drug, which are then contacted to validate this information and to gather additional information on prescription drug payments.

[18] Because income data from the CPS are more reliable, we use income estimates from the CPS to construct expenditure shares. Using the income estimates from the MEPS leads to very similar results (see the Appendix).

by 30-60 year-olds.

To investigate the stability of the age profile of users, we supplement the MEPS data with the National Ambulatory Medical Care Survey (NAMCS), which is an annual survey of doctors working in private practice and includes drug use data for the years 1980, 1981, 1985 and 1989-2000. Observations are at the doctor-patient-visit level; there are about 40,000 visits per year. Doctors are selected randomly, surveyed for a week, and patient-visits are then selected randomly from all the visits that week (further details on this survey are given in Acemoglu and Linn (2003)). We use the same classification system with the NAMCS as with the MEPS. Because the NAMCS does not contain expenditure information and its sampling scheme makes it less representative and less reliable than the MEPS, we focus on the MEPS for our main analysis and use the NAMCS mainly to check the stability of the age profiles of users.

Table 1.1 gives correlations between various measures of drug use. The first two rows of Panel A show a high degree of correlation between age profiles of use from the NAMCS surveys at various dates, both unweighted or weighted by total use of each category in the survey. These results indicate that the age profiles are similar between the 1980s and the 1990s.[19] The third and fourth rows report mean correlations by drug. These are constructed by computing the within category correlation between the measures and then averaging it across all categories. These correlations also show a substantial degree of persistence over time, especially when we look at the weighted correlation in row 4. The difference between the weighted and the unweighted correlations reflects the relatively imprecise estimates of use per person for the smaller categories.

Panel B performs the same calculation for expenditure shares from the three waves of the MEPS (weighted correlations now use total expenditure in each category as weights), and similarly shows substantial persistence in the age profiles of expenditure. Notably, there is now an even larger difference between weighted and unweighted mean correlations by drug, presumably because the MEPS, which is a more representative sample of the U.S. population than the NAMCS, has only a few observations in some of the smaller drug categories. This motivates our focus below on weighted regressions. Finally, Panel C shows high correlations

[19] Nevertheless, as we will see below, there is evidence of an increase in use per person in categories which have also experienced an increase in market size due to demographic changes.

both between the NAMCS and the MEPS measures and between expenditure shares and use per person in the MEPS.

The last major data source is a list of FDA new drug approvals. We exclude over-the-counter drugs, the so-called orphan drugs,[20] and drugs that have the same identifying characteristics (i.e., same name, company, and category, or the same FDA approval number). We focus on the time period 1970-2000. Both the quality of the approvals data and the quality of our measures of potential market size deteriorate as we go back in time for a number of reasons. First, we can only match FDA categories for drugs that are still listed by the FDA; second, before 1970 we cannot separately identify generics and non-generics; and finally, we are using age profiles from the 1990s. Our approvals dataset for 1970-2000 comprises 5,374 prescription drugs, including both generics and non-generics (see the Appendix). Since 1970 there have been 2,203 non-generic approvals and 442 new molecular entities.

Figure 1.3 shows the share of drug approvals over time to compare with changes in income shares depicted in Figure 1.2. To construct Figure 1.3, we allocate each of the 33 categories to the broad age group that has the largest expenditure in that category. The share of drug approvals is equal to the number of approvals in a given category in each five-year period divided by total approvals in that period.[21] Although this cut of the data uses only a small part of the information that the regression analysis below exploits, a positive association between changes in income shares and changes in drug approvals can be detected by comparing this figure to Figure 1.2. For example, the income share of the 30-60 group increases over the sample, as does the entry of drugs most used by this group. The shares of income and entry of drugs for those 0-30, on the other hand, show a downward trend. Finally, both the shares of income and entry of drugs for the 60+ group are relatively constant over the sample period. We explore these patterns in greater detail in the regression analysis below.

[20] These drugs treat rare conditions, affecting fewer than 200,000 people. An example is botox, first developed to treat adult dystonia, which causes involuntary muscle contractions. We drop these drugs because we have difficulty matching them consistently, and because they receive special inducements under the Orphan Drug Act.

[21] There are large fluctuations in the total number of approvals, partly because of a number of institutional changes. For example, it was discovered in 1989 that some FDA officials were taking bribes to speed up the approval process for generic drugs. As a result, in the early 1990s the approval process for generics was greatly slowed. See, for example, The Washington Post, August 16, 1989. In fact, there is a large drop in generics approvals in the early 1990s, but only a small decline for non-generics. We thank Ernst Berndt for suggestions on this issue.

1.5 Results

1.5.1 Basic Specifications

Table 1.2 provides the basic results from the estimation of equation (1.11) with Quasi-Maximum Likelihood (QML). The top panel is for all approvals. Panels B and C look at non-generics and new molecular entities (non-generics containing new molecules), and Panel D reports results for generics. Throughout the paper, the standard errors are corrected for heteroscedasticity using the Huber-White formula (see footnote 10). In this table we use the basic (income-based) measure of $\log M_{ct}$, constructed using expenditure data from the MEPS, and income from the CPS, the time periods correspond to five-year intervals, and observations are weighted by total expenditure in the corresponding drug category in the MEPS.

Column 1 of Panel A shows that the QML estimate of α for all new drugs is 6.15 with a standard error of 1.23, which is significant at the 1 percent level.

The remaining columns of Panel A investigate whether it is current market size or past or future market size that has the strongest effect on entry of new drugs. Column 2 includes current and five-year lagged market size together,[22] column 3 includes current and five-year lead market size, and finally, column 4 looks at the relationship between lead market size and entry of new drugs.

The entry of all drugs appears to respond to current or five-year lead market size. When current and lagged market sizes are included together, the coefficient on current market size has a similar magnitude to column 1, while lagged market size is negative, and neither coefficient is significant, presumably because current and previous market sizes are highly correlated. When current and lead market sizes are included together, current market size is not significant, whereas lead market size is significant at 5 percent. Moreover, column 4 shows that lead market size has a somewhat larger effect than current market size (the estimate of α is now 7.57, with standard error 1.99).

The results in Panel A combine generics and non-generics. Entry of generics and non-generics may be driven by different processes. Moreover, generics, which are identical to existing

[22] We construct the lagged market size measures for 1960s using demographic information from the CPS, so the number of observations does not decline. If we only use the post-1970 data, the results are similar, though lagged entry is somewhat stronger.

drugs, do not correspond to "innovation". Panel B shows the relationship between potential market size and entry of new non-generic drugs. The estimate of α is now 3.82, with standard error 1.15, which is also significant at 1 percent.

Perhaps more relevant for the relationship between market size and innovation is the response of new molecular entities. These are drugs classified by the FDA as containing new active ingredients, and thus correspond to more radical innovations (there are 442 new molecular entities and 2,203 new non-generics in our dataset). Panel C shows a significant relationship between market size and new molecular entities. The estimate of α is 3.54 (standard error = 1.19), and is again statistically significant at 1 percent. Interestingly, while all non-generics are most responsive to current market size, the coefficient of lead market size is larger with new molecular entities, 5.75, though the standard error is also larger, 2.37. This evidence, though not conclusive, is consistent with a limited anticipation effect in the response of innovation to market size.[23]

Finally, for completeness Panel D shows the effect of potential market size on the entry of generics. The estimate of α in the baseline specification of column 1 is 11.81 (standard error = 3.30). This effect is considerably larger than those that are obtained in the cross-sectional studies, such as Scott Morton (1999) and Reiffen and Ward (2004). This might partly reflect the complex entry dynamics of generic drugs, though we did not find evidence of such dynamics in our investigations. Since generics are not our main focus, we do not pursue potential explanations for this large effect further.

The magnitude of the effects of market size on entry of non-generics and new molecular entities in Table 1.2 is also large, in particular, larger than the proportional effects predicted by our model.[24] Note however that our key right-hand side variable measures potential market

[23]Here "limited" does not refer to the strength of the effect, but to the fact that the response to market size is five years before the change in market size, not further in advance. Two-period (ten-year) lead of market size has a large, but imprecise effect on new molecular entities (significant at 5 percent), while further leads are insignificant.

[24]Our estimates refer to the effect of market size on the flow of new drugs, which may differ from the effect on the stock of drugs. To check for this possibility, we estimate our basic models using $\sum_{\tau=1}^{t} N_{c\tau}$, i.e., the stock of drugs at time t, as the dependent variable. Consistent with our finding of limited residual serial correlation below, this procedure leads to slightly smaller estimates, 1.84 (standard error = 0.98) for non-generics, and 3.10 (standard error = 1.06) for new molecular entities. The evidence is therefore consistent with a broadly similar and somewhat smaller response of the stock of drugs to market size than the response of the flow of new drugs.

It also has to be borne in mind that these estimates are informative about the effect of market size on the *composition* of research, and the relationship between total pharmaceutical market size and aggregate research size rather than actual market size, and these two measures might differ because of changes in expenditure shares. To investigate whether the difference between actual and potential market size might affect the magnitude of our estimates, we use the NAMCS, where we can measure actual market size in terms of total use in each drug category between 1980 and 2000 (recall that the NAMCS data do not contain expenditure information). A simple regression of actual market size in each drug category on our measure of population-based potential market size and category and period dummies yields a coefficient of 4.06 (standard error = 1.60), which suggests that between 1980 and 2000 actual market size went up by 4 percent for every 1 percent increase in potential market size.[25] Assuming that this relationship also applies to the entire sample and to the income-based measure of market size, our estimates of 4-6 percent response to potential market size are consistent with a proportional relationship between entry of new drugs and actual market size.[26]

1.5.2 Robustness

Table 1.3 investigates the robustness of the effect of potential market size on the entry of non-generic drugs (Panel A) and new molecular entities (Panel B). Although the entry of new molecular entities may be responding somewhat more to the five-year lead of market size than to current market size, we report results for current market size in Panel B for compatibility with Panel A. Results for the effect of five-year leads of market size on new molecular entities

could be quite different. If we estimate (1.11) for non-generics without time effects, we obtain a coefficient of 0.22, with a standard error of 0.09 (for new molecular entities, the estimate is 0.41, with a standard error of 0.06). This is consistent with the view that the response of the composition of R&D to market size is quite different from the response of total R&D. Nevertheless, the difference between the results with and without time effects is at least partly due to the presence of other time-varying factors affecting entry of new drugs.

[25] This result implies that use per person went up in categories experiencing an increase in market size due to demographic changes, which may itself be partly due to increases in the number and quality of drugs in these categories.

[26] In any case, our estimate of a 4 percent increase in the rate of entry of new non-generics in response to 1 percent increase in market size is not implausible. There are a total of 2,203 non-generic approvals between 1970 and 2000, thus on average 10 approvals in every five-year interval in each of our 33 categories. Therefore, our estimate implies that a 2.5 percent increase in market size should lead to the entry of about 1 new drug. Total pharmaceutical sales were approximately $130 billion in 1999 (IMS (2000)), which implies an average annual expenditure of $3.9 billion per category. A 2.5 percent increase therefore corresponds to $97.5 million, or about $1.5 billion over 15 years, which is the life of a typical non-generic drug. Since entry costs for non-generics are around $800 million (in 2000 dollars, DiMasi et al (2003)), entry of one new drug in response to an increase of approximately $1.5 billion in revenue is within the range of plausible responses Naturally, this calculation is very rough and only suggestive, since it ignores the difference between average demand and the demand that a marginal entrant will capture.

are similar, and generally somewhat stronger.

Column 1 replicates the baseline results from Table 1.2 for comparison. In column 2, time periods are ten years instead of the five-year intervals. The estimate of α for non-generics is somewhat larger, 4.81 (standard error = 1.31), while the estimate for new molecular entities is similar to the baseline, 3.91 (standard error = 1.29). Both estimates are significant at 1 percent.

Column 3 looks at the effect of changes in market size driven purely by population changes (in this case, regression weights are total use in the corresponding category in the MEPS). The estimates in all three panels are larger than the baseline, and continue to be significant at 1 percent. Since the income-based measure is closer to the notion of market size suggested by theory, we continue to focus on this measure.[27] Column 4 shows similar results for non-generics and new molecular entities using the population-based measure of market size with our alternative dataset, the NAMCS.

Column 5 uses an OECD market size measure combining West European and Japanese demographic information with the U.S. information.[28] Since we only have information on population for the other countries, we perform this exercise for the population-based measure of market size. The U.S. and OECD populations by age group have a high correlation, equal to 0.81. Using the OECD market size measure leads to similar, and somewhat surprisingly, more precise results. For example, for all non-generics, the estimate of α is 3.27 (standard error = 0.86), while for new molecular entities, the estimate is 3.28 (standard error = 0.84).

Column 6 investigates the effect of weighting on the estimates. The unweighted estimate of α for all non-generics is smaller than the baseline, 1.81 (standard error = 1.61), and no longer statistically significant. For new molecular entities, the estimate is larger and still significant; 4.62 (standard error = 1.98). In both cases, the standard errors are significantly larger, reflecting the fact that expenditure shares in smaller cells are less precisely estimated.

Column 7 uses an alternative measure of market size constructed with single-age groups for i_{at}'s and u_{ca}'s in equation (1.13). This procedure uses more information about the age profiles, but since there are fewer observations in some single-age groups, the estimates of u_{ca}'s are less

[27] Instrumenting the income-based market size measure with the population-based market size measure leads to similar estimates to the baseline. For example, the estimate for non-generics is 4.07 (standard error = 1.37) and the estimate for new molecular entities is 3.84 (standard error = 1.41).

[28] These data were obtained from the United Nations website, esa.un.org/unpp/.

precise. The estimates using this alternative measure are very similar to the baseline results. For example, the estimates of α for non-generics is 3.67 with a standard error of 1.18, and the estimate for new molecular entities is 3.35 (standard error = 1.23).

Column 8 uses the alternative classification system from Acemoglu and Linn (2003), which uses only differences in the age profiles of expenditure to subdivide the major FDA categories. This classification system contains 34 categories, and because it includes a number of small FDA detailed categories that are dropped from the current system, there are now 106 more approvals (see the Appendix). The estimate for non-generics is similar to the baseline in column 1, 3.68 (standard error = 1.07), while the estimate for new molecular entities is smaller than the baseline and insignificant. However, with this classification system, there is a stronger effect of lead market size on new molecular entities (and thus somewhat stronger evidence for anticipation effects). For example, the estimate of α with new molecular entities and lead market size is 6.81 (standard error = 1.29), which is significant at the 1 percent level (not reported).

Column 9 estimates the model in (1.12) with the Pakes-Griliches transformation using OLS. The estimates are similar to those in column 1. For example, for non-generics, the estimate of α is 3.37 with a standard error of 1.75, and the estimate of α for new molecular entities is 3.54 with a standard error of 1.40. We also use the linear model to document that the relationship we observe is not driven by outliers. Figure 1.4 shows the relationship between the residuals of new molecular entities versus the residuals of market size, $\log M_{ct}$, after drug category and time period dummies are removed. Observations are labeled by their drug category codes (see the Appendix Table), and each code appears more than once, since there are multiple periods. This figure shows that there are no major outliers (the figure for non-generics is similar).[29]

Finally, column 10 checks the robustness of the results to dropping the Cardiac category, which includes the most diverse types of drugs. The exclusion of this category has little effect on the estimates.[30]

[29] Dropping the one category that appears as a slight outlier, category 14 (Anorexiants/CNS Stimulants), has no effect on the estimate, which only changes to 3.66 (standard error = 1.46).

[30] We have also experimented with dropping each of the other categories one at a time. The effect of market size on both non-generics and new molecular entities remains significant at 5 percent in all cases, except when we drop Antibiotics. Without Antibiotics, the estimates are similar to the baseline results, but no longer significant. For non-generics, the estimate is 1.72 (standard error = 2.01) and for new molecular entities, 4.14 (standard error = 2.59).

1.5.3 Potential Supply-Side Determinants of Innovation

The first part of Table 1.4 investigates the robustness of the baseline results to controlling for potential supply-side determinants of innovation, such as changes in scientific incentives or opportunities captured by the δ_j's in the theoretical model. Our main focus is the effect of market size on the entry of new non-generic drugs (Panel A) and on new molecular entities (Panel B).

First, recall that the major threat to our identification strategy is changes in the δ_j's (since permanent differences in δ_j's are already taken out by the drug category fixed effects). If the δ_j's vary over time, they are also likely to be serially correlated. Adding lags of $\log N_{ct}$ to our basic specifications is therefore a simple way to check the importance of these concerns.

Columns 1 and 2 of Table 1.4 report the results of QML estimation of a lagged dependent variable specification of the form

$$E\left[N_{ct} \mid \zeta_c, \bar{X}_c, \bar{N}_c\right] = \frac{\exp\left(\alpha \cdot \log M_{ct} + \psi \cdot \log \tilde{N}_{ct-1} + \nu \cdot d_{ct-1} + \mu_t\right)}{\sum_{\tau=1}^{T} \exp(\alpha \cdot \log M_{c\tau} + \psi \cdot \log \tilde{N}_{c\tau-1} + \nu \cdot d_{c\tau-1} + \mu_\tau)} \bar{N}_c, \tag{1.14}$$

where we use the notation in (1.11), and in particular, $\bar{N}_c = \sum_{\tau=1}^{T} N_{c\tau}$ is the number of drugs approved in category c over the entire sample. Since N_{ct-1} can be equal to 0, we follow a procedure similar to that of Pakes and Griliches (1980), used in column 9 of Table 1.3 above, and define $\tilde{N}_{ct-1} = N_{ct-1}$ if $N_{ct-1} \geq 1$ and $\tilde{N}_{ct-1} = 1$ if $N_{ct-1} = 0$, and add the dummy d_{ct-1} that equals 1 when $N_{ct-1} = 0$. The estimate of α for non-generics in this case is 3.84 (standard error = 1.07), and the coefficient on lagged entry is 0.12 (standard error = 0.10), thus small and insignificant.[31] The results for new molecular entities are similar.

The estimates in column 1 treat all right-hand side variables as strictly exogenous. Because $\log \tilde{N}_{ct-1}$ is endogenous, these estimates are inconsistent. As long as there is no additional autocorrelation in the errors, instrumenting for $\log \tilde{N}_{ct-1}$ with $\Delta \log \tilde{N}_{ct-2}$ would lead to consistent

[31] In the estimates reported in columns 1 and 2 of Table IV, we use information on approvals before 1970 to construct lags, so the sample size remains the same as in the basic specifications. Because before 1970 we cannot distinguish generics and non-generics (though we can identify new molecular entities), we use all approvals for the pre-1970 lags in Panels A.

Whether or not we use the pre-1970 data is not important for the results. Estimating the model in (1.14) without the pre-1970 data for non-generics gives an estimate of α of 3.75 (standard error = 1.93), and the lagged dependent variable is again very small and insignificant. The same is true for new molecular entities.

estimates.[32] Since the estimating equation in (1.14) is nonlinear, we perform this instrumentation strategy by adding the residuals from the first-stage regression as an additional right-hand side variable to the second stage (see Wooldridge (2002), Chapter 19.5). Column 2 reports the results of this exercise. Once again, the coefficients are very similar to the baseline estimates in all three panels, and show no evidence of significant effects of lagged entry. For example, for non-generics, the estimate of α is 3.98 (standard error = 1.16), and for new molecular entities, the estimate is 3.69 (standard error = 1.38). Furthermore, as a direct check, we tested and found no evidence for serial correlation in the residuals from the estimation of (1.11) in Table 1.2. Overall, the results show that the effect of potential market size on entry of non-generics and new molecular entities is robust to controlling for lagged entry and there is no evidence of residual serial correlation.

A plausible conjecture is that non-profit incentives to develop drugs would be related to opportunities to save lives or cure major illnesses. Our second strategy controls for differences in the health benefits of new drugs across categories. New drugs in our data set include both drugs that are demanded by the consumers but do not "save lives", such as Prozac, Paxil, Vioxx, or Viagra, and those that actually save lives such as heart medicines or cancer treatments (see Lichtenberg (2002, 2003), on the effect of pharmaceutical innovations on declines in mortality). We estimate the number of life-years lost corresponding to each drug category using the Mortality Detail Files from the National Center for Health Statistics from 1970-1998. Following Lichtenberg (2002), for each death we subtract the person's age from 65, then calculate the total number of life-years lost for all the deaths resulting from diseases related to drugs in each category.[33] Column 3 adds life-years lost to the right-hand side of our baseline regression models as a proxy for this source of non-profit incentive to undertake research. The estimate of the effect of market size on non-generics is now 3.58 (standard error = 1.70) and the estimate for new molecular entities is 3.64 (standard error = 1.79). In all cases, the variable for life-years lost is not significant.

We also investigate the implications of differences in scientific funding for various drug

[32] See, for example, Arellano and Bover (1995) and Blundell and Bond (1998). We cannot use other commonly-used moment restrictions, since equation (1.14) cannot be first-differenced.

[33] For example, if someone dies at age 32, this counts as 33 life years lost; people dying older than 65 receive no weight in this calculation. Since many of our categories contain diseases or conditions that do not lead to death, we obtain several empty cells.

categories. Using the Computer Retrieval of Information on Scientific Projects (CRISP) dataset (details are contained in Lichtenberg (2001) and Acemoglu and Linn (2003)), we construct a variable measuring the total amount of federal funding for research in each drug category, and include this variable as a control. To the extent that government funding also responds to potential market size (for example, because drug companies have a greater tendency to apply for funding in areas where they plan to do research), this variable would be correlated with our market size measure. In practice, the correlation is low, and column 4 shows that the inclusion of this variable has little effect on our estimates. The estimate of the effect of market size is 3.86 (standard error = 1.20) for non-generics and 3.56 (standard error = 1.20) for new molecular entities. The funding variable itself is positive, but small and insignificant (not reported in the table).

Next, to control for potential trends in scientific opportunities across drug categories, we add proxies for pre-existing trends. We construct an estimate for pre-existing trends as $\Delta_c = (\log N_{c,70} - \log N_{c,40})$, where $\log N_{c,70}$ is the log approvals for category c in 1970 and $\log N_{c,40}$ is the log approvals in 1940.[34] We then add a full set of interactions between Δ_c and the time dummies. This specification therefore allows drug categories that have grown differentially between 1940 and 1970 to also grow at different rates in the later periods. Column 5 reports the results of this exercise. The estimates are very similar to the baseline in all three panels. These results are perhaps not surprising, since pre-1970 approvals are considerably noisier and do not distinguish between generics and non-generics, thus are only an imperfect control for pre-existing trends.

An alternative, and substantially more demanding, strategy is to include in-sample linear time trends. To do so, we add linear time trends for each of the 16 major FDA categories. We expect technological differences to be well approximated by the 16 major drug categories, which are based on broad therapeutic intent. The estimates, reported in column 6, are quantitatively similar to the baseline, but no longer significant because the standard errors are larger, reflecting the fact that changes in market size due to demographic trends are relatively smooth, and thus highly colinear with the time trends (nevertheless, the estimate for new molecular entities

[34]Because we cannot distinguish between non-generics and generics before 1970, in Panel A, we use total approvals before 1970. Also, since there were no new molecular entities approved in the 1940 period, in Panel B we construct the pre-existing trend using 1950 and 1970 approvals of new molecular entities.

is significant if we use lead market size). Since distinguishing between linear time trends and changes in market size may be more difficult with the nonlinear model in (1.11), we also estimate the model in (1.12) with linear time trends using OLS. The results, reported in column 7, are also similar to the baseline, and now the estimate for non-generics is significant at 5 percent and the estimate for new molecular entities is significant at 10 percent.

We also investigate the potential effects of recent advances in biotechnology, such as the use of recombinant DNA or other technological changes, which may correspond to changes in the δ_j's in terms of our model. In column 8, we drop the categories of Cancer and Vascular, which, according to the FDA approval list, have witnessed the entry of the greatest number of orphan drugs (presumably by biotechnology firms). In addition, there is anecdotal evidence that biotechnology firms were first active in producing insulin (the Thyroid and Glucose category) and in the Anemia category, so in column 9 we drop these two categories.[35] In both cases, the estimates are very similar to our baseline results.

Finally, to see whether the advent of biotechnology or other technological advances of the past two decades have changed the relationship between market size and entry of new drugs, we estimate our baseline models including an interaction between a post-1985 (or post-1990) dummy and market size. Our estimates show no evidence of significant interactions.[36]

The results in this subsection therefore document that a number of controls for supply-side factors have little effect on our main finding regarding the effect of market size on entry of non-generic drugs and new molecular entities. Although these results are not conclusive on the effect of scientific or other supply-side factors in pharmaceutical research, they suggest that the effect of potential market size on entry and innovation is relatively robust.

1.5.4 Changes in Health Insurance Coverage

Our market size measure only exploits changes in potential market size driven by demographic trends. Another source of variation in market size comes from changes in coverage of drug

[35] Biotechnology firms have also been active in producing human growth factor, but since there are only a small number of individuals using these drugs in the MEPS, these drugs are not included in our approvals dataset.

[36] For example, in a specification parallel to the model for non-generics in column 1 of Table II, the estimate of α is 4.29 (s.e. = 1.66), and the interaction with the post-1985 dummy is -0.13 (s.e. = 0.18), thus small and insignificant. For new molecular entities, the estimate of α is 4.00 (s.e. = 1.46) and the post-1985 interaction is -0.10 (s.e. = 0.13).

expenditure in private or public health insurance programs. During our sample period, there were significant changes in the coverage of drug expenditures in health insurance plans. For example, the percentage of 60+ year-olds with private insurance rose from 60 percent to 75 percent between 1974 and 1996 (authors' calculations). We now investigate the effect of including information on health insurance coverage in our market size measure.

We use the National Health Interview Survey (NHIS, 1974-1996) to construct a market size measure incorporating information on health insurance coverage as follows: $\widetilde{M}_{ct} = \sum_a u_{ca} \cdot i_{at} \cdot f_{at}$, where f_{at} is the fraction of those of age a in period t with private health insurance, u_{ca} and i_{at} are as defined above. Because there is no consistent information on prescription drug coverage, we assign prescription coverage to any individual with both doctor and surgical coverage. Prescription drug coverage is highly correlated with this measure in the years we can observe it. In column 10, we use $\log \widetilde{M}_{ct}$ as our market size measure instead of $\log M_{ct}$. This leads to similar, and somewhat more precise, results. For non-generics, the estimate of α is 1.92 with standard error 0.44, while for new molecular entities, it is 2.10 (standard error = 0.51). Despite the greater precision of these results, we have more confidence in our baseline estimates, since the measure $\widetilde{M}_{ct}$ effectively assigns 0 expenditure to those without insurance and relies on information on drug coverage imputed from doctor and surgical coverage.

1.5.5 Reverse Causality

Lichtenberg (2002, 2003) shows that new drugs have increased the average age at death. This introduces the potential for reverse causality whereby the market size for successful drugs may be endogenously larger, because their users live longer. This is unlikely to be a first-order concern, since drug-induced changes in population are small relative to the demographic changes that we are exploiting. Nevertheless, we address this issue by instrumenting for current population using the corresponding population from 5 years before. For example, we use the income of 50-54 year-olds in 1975 as an instrument for the income of 55-59 year-olds in 1980. The fraction of 50-54 year-olds is highly correlated with the fraction of 55-59 year-olds 5 years later, but is unaffected by new drugs that are developed in the intervening 5 years. As described above, the instrumental-variables procedure is performed by adding the residuals from the first-stage as an additional right-hand side variable (see Wooldridge (2002)).

These instrumental-variables estimates, reported in column 11 of Table 1.4, are similar to the baseline results and show no evidence of reverse causality. For example, for non-generics, the estimate of α is 2.93 (standard error = 1.45), and for new molecular entities, the estimate is 3.08 (standard error = 1.32).

1.5.6 Patents

The results so far show a large and robust effect of potential market size on entry of non-generic drugs and new molecular entities, and suggest a strong link between market size and innovation. In this subsection, we briefly investigate the relationship between market size and another measure of pharmaceutical innovation, patents.[37]

We obtained data on pharmaceutical patents from Thomson Derwent Inc., and with the help of a specialist at this company, we mapped these patents into our classification system.[38] Using these data, we found no significant relationship between market size and patents, which might be due to a variety of reasons.[39] First, this result may simply reflect the imperfect match between the patent data and the FDA categories, especially bearing in mind the potential use of certain chemical structures in multiple drug lines. Second, the significant costs and uncertainty involved in the development of new molecules and patentable products may be creating substantial attenuation (e.g., a drug intended for the 1990s may be patented in the 1980s or 1990s, depending on delays in the research process). Third, pharmaceutical companies may respond to profit incentives more during the later stages of the research process than during the earlier stages. Finally, because U.S. patents include those taken by foreign companies, they may be more responsive to OECD demand than to U.S. demand. To investigate the last possibility, we estimated the relationship between changes in OECD market size derived

[37]Firms typically apply for a patent prior to the clinical trial stage of drug development, or about 5-10 years before the drug is approved, and therefore lose a significant fraction of the life of the patent before it can begin marketing the drug. Part of the 1984 Hatch-Waxman Act allowed pharmaceutical companies to apply to the FDA for an extension of the life of their patents, if they could show that they lost marketing time while waiting for approval. The maximum extension is 5 years, and depends, among other things, on the length of the FDA approval process. Overall, companies have a maximum of 14 years of patent protection after FDA approval.

We were unable to obtain data for a sufficient number of categories for another possible proxy for pharmaceutical innovation, clinical trials.

[38]We could not use the data from the Hall-Jaffe-Trachtenberg patent dataset (see Jaffe and Trachtenberg, 2002) because we were unable to map their classification based on chemical structure to our drug categories.

[39]Finkelstein (2004) also finds weaker results for vaccine patents than for later stages of development.

from European, Japanese and U.S. demographic changes. In this case, we find a significant relationship between five or ten-years leads of OECD market size and patents. With the five-year of lead market size, the estimate of α is 3.49 (standard error = 1.02) and with the ten-year lead, the estimate is 5.02 (standard error = 0.63).[40] Although this result suggests that OECD demand may be more important for patents, we are currently unable to make more progress in distinguishing between these various explanations, and the weaker results for patents remain a puzzle.

1.6 Concluding Remarks

This paper investigates the response of entry of new drugs and pharmaceutical innovation to changes in potential market size driven by demographic changes. Our results indicate that a 1 percent increase in the potential market size for a drug category leads to approximately a 4 percent growth in the entry of new non-generic drugs and new molecular entities.

The effect of market size on entry of new drugs and new molecular entities, if further proven to be robust, has important implications both for research on the pharmaceutical industry, and for the endogenous growth and directed technical change literatures. It provides evidence that, as posited by these models, R&D and technological change are directed towards more profitable areas. These findings also imply that there may be little pharmaceutical research towards drugs with small markets, especially towards those intended for groups with limited ability to pay, which is a key premise of recent work by Kremer (2002). Based on this premise, Kremer suggests that there needs to be selective government incentives for developing drugs against malaria and other third-world diseases.

We view this research as part of a broader investigation of the effects of profit incentives on innovation. Evidence from a single industry may be nonrepresentative, for example because pharmaceuticals may be more research oriented than other industries. Future research investigating the response of innovation and entry of new products to market size both in specific industries and at the economy-wide level is necessary to substantiate the results presented here.

[40] Part of the difference between the U.S. and OECD results is driven by the fact that we are using income-based measures for the U.S. and the population-based measures for the OECD. Using the population-based measure for the U.S. lead market size with patents yields a coefficient of 2.34 (standard error = 4.63).

1.7 Appendix

1.7.1 Medical Expenditure Panel Survey (MEPS) and Construction of the Drug Classification System

The MEPS is an annual survey of randomly sampled households; we use the 1996, 1997 and 1998 surveys. We obtain each person's age, the name and national drug code of the prescription drug(s) used, and total expenditure (there are multiple records for people who use more than one prescription drug). Over the 3 years, we have about 500,000 drugs used and about 85,000 individuals. Expenditure information includes out-of-pocket expenses, as well as amounts paid by insurance companies and government payments (e.g., Medicaid and worker's compensation). These data are collected from the pharmacies and medical providers listed by the respondents.

We begin with the 159 therapeutic categories, obtained from the FDA's National Drug Code (NDC) Directory. The names of these categories can be found in the second column of the Appendix Table. The NDC Directory contains a file with the therapeutic category for most FDA approved drugs currently on the market. We assign each drug in the MEPS to one of the 159 categories by matching it by national drug code with a drug in the NDC file. We cannot match about 10 percent of the drugs mentioned in the MEPS; these are usually not commonly used drugs, and make up less than 5 percent of the total drugs used.

Drug expenditure shares and use per person are calculated by computing drug expenditure and use by five-year age group, and then dividing by the income and population of the age group. We use the population numbers from the MEPS (so use per person is the weighted average of use per person of respondents in the MEPS), but income estimates from the CPS. We prefer the CPS income estimates because the MEPS income data are likely to contain greater measurement error; in the MEPS, the sample is smaller, wage and salary incomes for almost half of the sample are imputed either based on broad income ranges or other information, non-wage incomes were generally imputed, and the imputation methods changed between the 1996-97 and 1998 surveys. Nevertheless, the results are almost identical if we construct expenditure shares for individuals in the survey (i.e., without using CPS information in the same way as we do for use per person). For example, the estimate of the effect of market size on the entry of non-generics is 4.08 (standard error = 1.31) and the estimate for new molecular entities is

3.59 (standard error = 1.35). We also checked the robustness of our results using an alternative market size measure constructed with single-age groups, and the results are reported in Table 1.3. We prefer the measure using five-year age groups, since there are only a few observations in some single-age groups in the MEPS.

The FDA has assigned the 159 categories to one of 20 major therapeutic categories. As noted in the text, we drop four major categories: Anesthetics, Antidotes, Radiopharmaceuticals, and Miscellaneous.[41] Within each major category, we first subdivide categories whose drugs have different indications (we determine drugs' indications by searching by name on the National Institute of Health website, www.nlm.nih.gov/medlineplus/druginformation.html). For categories that have not been subdivided based on indications, we then divide them if there is sufficient heterogeneity in the age profile of users for subcategories. The Appendix Table shows that we create subcategories when there is considerable age variation within broad categories. For example, within the Hormones major category, Estrogens/Progestins are used predominately by 30-60 year-olds, while Contraceptives are used fairly evenly by 0-30 year-olds and 30-60 year-olds. This classification system differs somewhat from the working paper version, in which we divided major classes based entirely on age structure. The previous system includes several FDA categories that are dropped from the current one: the CNS Miscellaneous, Hyperlipidemia and Calcium Metabolism categories, which contain drugs used for heterogeneous conditions, and several categories that would have been subdivided based on drug indications, but had fewer than 1,000 observations in the MEPS. The details of the previous classification system are in Acemoglu and Linn (2003), and Table 1.3, column 8, reports results using this older classification.

1.7.2 Drug Approvals from the FDA

The list of FDA drug approvals were obtained by Lichtenberg and Virahbak (2002) under the Freedom of Information Act. We thank Frank Lichtenberg for generously sharing these data with us. Over-the-counter drugs and orphan drugs (of which only a few can be matched) are excluded. Biologics, which go through a separate approval process, are not in this dataset.

[41] We also drop several minor categories when there are not sufficient observations to estimate a reliable age profiles. We use about 1,000 observations as our cutoff rule. We obtain this number from observing that only categories with more than 1,000 observations have fairly smooth age profiles.

We match drugs in the approval list to FDA categories by drug name and FDA approval number. 14,432 of 16,772 prescription drugs (86 percent) approved since 1970 are matched, while before 1970, the match rate is about 51 percent. This motivates our focus on drug approvals between 1970 and 2000. Drugs that have the same approval number as a previously approved drug and drugs for which the corresponding FDA category is dropped because of insufficient observations in the MEPS are excluded. Finally, we drop drugs with the same name, MEPS category and company as a previously approved drug.

1.8 References

Acemoglu, Daron. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality," *Quarterly Journal of Economics.* CXIII (1998), 1055-1090.

Acemoglu, Daron. "Directed Technical Change," *Review of Economic Studies.* LXIX (2002), 781-810.

Acemoglu, Daron. "Labor- and Capital-Augmenting Technical Change," *Journal of European Economic Associations*, I (2003), 1-37.

Acemoglu, Daron and Joshua Linn. "Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry," NBER *Working Paper No.* 10038, 2003.

Aghion, Philippe and Peter Howitt. "A Model of Growth Through Creative Destruction," *Econometrica*, LX (1992), 323-351.

Arellano, Manuel and Olympia Bover. "Another Look at the Instrumental-Variable Estimation of Error-Components Models," *Journal of Econometrics*, LXVIII (1995), 29-52.

Arrow, Kenneth J. "The Economic Implications of Learning-by-Doing," *Review of Economic Studies*, XXIX (1962), 155-173.

Blundell, Richard and Stephen Bond. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, LXXXVII (1998), 115-144.

Cerda, Rodrigo. "Drugs, Market Size and Population," University of Chicago Ph.D. thesis, 2003.

Cockburn, Iain and Rebecca Henderson. "Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research," *Journal of Health Economics*, XX (2001), 1033-1057.

Danzon, Patricia M., Sean Nicholson and Nuno Sousa Pereira. "Productivity in Pharmaceutical-Biotechnology R&D: The Role of Experience in Alliances," *NBER Working Paper No.* 9615, 2003.

DellaVigna, Stefano and Joshua Pollet. "Attention, Demographics and the Stock Market," University of California Berkeley mimeo, 2004.

DiMasi, Joseph A., Ronald W. Hansen and Louis Lasagna. "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics*, X (1991), 107-142.

DiMasi, Joseph A., Ronald W. Hansen and Henry G. Grabowski. "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, XXII (2003), 151-185.

Drandakis, E. and Edmund Phelps. "A Model of Induced Invention, Growth and Distribution," *Economic Journal*, LXXVI (1965), 823-840.

Dranove, David and David Meltzer. "Do Important Drugs Reach the Market Sooner?" *RAND Journal of Economics*, XXV (1994), 402-423.

Finkelstein, Amy. "Static and Dynamic Effects of Health Policy: Evidence from the Vaccine Industry," *Quarterly Journal of Economics*, CXIX (2004), 527-564.

Galambos, Louis and Jeffrey Sturchio. "Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation," *Business History Review*, LXXII (1998), 250-278.

Gambardella, Alfonso. *Science and Innovation: The U.S. Pharmaceutical Industry During the 1980s.* (Cambridge, UK: Cambridge University Press, 2000).

Griliches, Zvi. "Hybrid Corn: An Exploration in the Economics of Technological Change," *Econometrica*, XXV (1957), 501-522.

Grossman, Gene and Elhanan Helpman. *Innovation and Growth in the Global Economy.* (Cambridge, MA: MIT Press, 1991).

Hausman, Jerry, Bronwyn H. Hall and Zvi Griliches. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, LII (1984), 909-938.

Hayami, Yujiro and Vernon Ruttan. "Factor Prices and Technical Change in Agricultural Development: the U.S. and Japan, 1880-1960," *Journal of Political Economy*, LXXVII (1970). 1115-1141.

Henderson, Rebecca and Iain Cockburn. "Scale, Scope and Spillovers: The Determinants of Research Productivity in Drug Discovery," *Rand Journal of Economics,* XXVII (1996), 32-59.

IMS. "World Review" at www.ims-global.com/insight/report/market_growth/report0600.htm. 2000

Jaffe, Adam and Manuel Trachtenberg and Rebecca Henderson. *Patents, Citations, and Innovations.* (Cambridge, MA: MIT Press, 2002).

Kennedy, Charles "Induced Bias in Innovation and the Theory of Distribution," *Economic Journal,* LXXIV (1964), 541-547.

Kling, Jim. "From Hypertension to Angina to Viagra," *Modern Drug Discovery,* I (1998), 31-8.

Kremer, Michael. "Pharmaceuticals and the Developing World," *Journal of Economic Perspectives.* XVI (2002), 67-90.

Lichtenberg, Frank R. "The Allocation of Publicly Funded Biomedical Research," in *Medical Care Output and Productivity, Studies in Income and Wealth*, LXIII, Ernst Berndt and David Cutler, eds. (Chicago, IL: University of Chicago Press, 2001).

Lichtenberg, Frank R. "Sources of U.S. Longevity Increases, 1960-1997," *NBER Working Paper No.* 8755, 2002.

Lichtenberg, Frank R. "The Impact of New Drug Launches on Longevity: Evidence from Longitudinal, Disease-level Data from 52 Countries, 1982-2001," Columbia mimeo, 2003.

Lichtenberg, Frank R. and Suchin Virabhak "Pharmaceutical Embodied Technical Progress, Longevity and Quality of Life: Drugs as Equipment for Your Health," *NBER Working Paper No.* 9351, 2002.

Lichtenberg, Frank R. and Joel Waldfogel. "Does Misery Love Company? Evidence From Pharmaceutical Markets Before and After the Orphan Drug Act," *NBER Working Paper No.* 9750, 2003.

Ling, Davina C., Ernie Berndt, and Richard G. Frank. "General Purpose Technologies, Capital-Skill Complementarity, and the Diffusion of New Psychotropic Medications among Medicaid Populations," MIT mimeo, 2003.

Malerba, Franco and Luigi Orsenigo. "Innovation and Market Structure in the Dynamics of the Pharmacetutical Industry: Towards a History Friendly Model," *Industrial and Corporate*

Change, XI (2002), 667-703.

Newell, Richard, Adam Jaffee and Robert Stavins. "The Induced Innovation Hypothesis and Energy-Saving Technological Change," *Quarterly Journal of Economics.* CXIV (1999), 907-940.

Pakes, Ariel and Zvi Griliches. "Patents and R&D at the Firm Level: A First Look," *Economic Letters*, V (1980), 377-381.

Pakes, Ariel and Mark Schankerman. "An Explanation into the Determinants of Research Intensity," in *R&D, Patents and Productivity,* Zvi Griliches ed. (Chicago, IL: University of Chicago Press, 1984).

Popp, David. "Induced Innovation and Energy Prices," *American Economic Review*, XCII (2002), 160-180.

Reiffen, David and Michael R.Ward. 2004 "Generic Drug Industry Dynamics," *Review of Economics and Statistics* (2004), .

Romer, Paul M. "Endogenous Technological Change," *Journal of Political Economy*, XCVIII (1990), S71-S102.

Samuelson, Paul. "A Theory of Induced Innovations Along Kennedy-Weisacker Lines," *Review of Economics and Statistics*, XLVII (1965), 444-464.

Schmookler, Jacob. *Invention and Economic Growth.* (Cambridge, MA: Harvard University Press, 1966).

Scott Morton, Fiona. "Entry Decisions in the Generic Drug Industry," *The Rand Journal*, XXX (1999), 421-440.

Wooldridge, Jeffrey M. "Distribution-free Estimation of Some Nonlinear Panel Data Models," *Journal of Econometrics*, XC (1999), 77-97.

Wooldridge, Jeffrey M. *Econometrica Analysis of Cross Section and Panel Data.* (Cambridge, MA: MIT Press, 2002).

TABLE 1.1

Correlations Between Different Drug Use Measures

	Panel A: NAMCS use per person		
	1980/1990	1990/2000	1980/2000
Correlation	0.897	0.861	0.861
Weighted correlation	0.906	0.843	0.856
Mean correlation by drug	0.709	0.651	0.626
Weighted mean correlation by drug	0.820	0.825	0.790
	Panel B: MEPS expenditure share		
	1996/1997	1997/1998	1996/1998
Correlation	0.961	0.965	0.929
Weighted correlation	0.962	0.973	0.937
Mean correlation by drug	0.698	0.686	0.575
Weighted mean correlation by drug	0.865	0.881	0.796
	Panel C: NAMCS/MEPS use and MEPS use/expenditure		
	NAMCS/MEPS use	MEPS use/expenditure	
Correlation	0.869	0.954	
Weighted correlation	0.891	0.956	
Mean correlation by drug	0.804	0.902	
Weighted mean correlation by drug	0.935	0.940	

The numbers refer to the correlation of use per person or average expenditure share between the indicated dates and datasets. Observations are for five-year age groups by drug category (there are 33 x 19 = 627 observations in each case). In weighted correlations, observations are weighted by total use or expenditure from the MEPS or NAMCS. Mean correlation by drug computes correlations separately by drug category, then calculates the average.

TABLE 1.2

Effect of Changes in Market Size on New Drug Approvals

	(1)	(2)	(3)	(4)
	Panel A: QML for Poisson model, dep var is count of drug approvals			
Market size	6.15 (1.23)	6.84 (4.87)	-2.22 (4.12)	
Lag market size		-0.61 (3.85)		
Lead market size			10.16 (4.28)	7.57 (1.99)
	Panel B: QML for Poisson model, dep var is count of non-generic drug approvals			
Market size	3.82 (1.15)	6.72 (7.63)	2.91 (5.31)	
Lag market size		-2.49 (5.97)		
Lead market size			-1.77 (6.94)	1.73 (2.02)
	Panel C: QML for Poisson model, dep var is count of new molecular entities			
Market size	3.54 (1.19)	5.79 (6.66)	-1.38 (5.16)	
Lag market size		-1.99 (5.28)		
Lead market size			7.35 (5.11)	5.75 (2.37)
	Panel D: QML for Poisson model, dep var is count of generic drug approvals			
Market size	11.81 (3.30)	8.55 (6.85)	1.28 (7.17)	
Lag market size		3.12 (5.94)		
Lead market size			13.24 (8.66)	14.65 (3.71)
Number of observations	198	198	165	165

Huber-White robust standard errors are reported in parentheses. The dependent variable in Panel A is count of drug approvals, in Panel B the dependent variable is count of non-generic drug approvals, in Panel C the dependent variable is new molecular entities, and in Panel D, it is generic drug approvals, all calculated from the FDA dataset of New Drug Approvals (see Appendix). Market size is log potential market size calculated from the MEPS and the CPS, using 5-year age groups (see text). Lag market size refers to one-period lag of market size, and Lead market size refers to one-period lead of market size. All regressions include drug and time dummies, and use the income-based measure of market size. Time intervals are 5 years. Estimates are weighted by total expenditure for the category in the MEPS. The Poisson model is estimated by Quasi Maximum Likelihood (QML), with the Hausman, Hall and Griliches (1984) transformation. See equation (1.11) in the text.

TABLE 1.3

Robustness Checks

	Baseline QML	10 year intervals	Population-based market size	NAMCS market size	OECD market size	Unweighted regressions	Market size uses single-age groups	Market size uses previous classification	Linear model	Drop Cardiac
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: dependent variable is count of non-generic drug approvals										
Market size	3.82 (1.15)	4.81 (1.31)	5.35 (1.63)	4.53 (1.12)	3.27 (0.86)	1.81 (1.61)	3.67 (1.18)	3.68 (1.07)	3.37 (1.75)	3.90 (1.38)
Approvals	2203	2203	2203	2203	2203	2203	2203	2309	2203	2078
Panel B: dependent variable is count of new molecular entities										
Market size	3.54 (1.19)	3.91 (1.29)	5.13 (1.22)	4.16 (1.01)	3.28 (0.84)	4.62 (1.98)	3.35 (1.23)	2.73 (1.74)	3.54 (1.40)	3.17 (1.46)
Approvals	442	442	442	442	442	442	442	492	442	397
Number of observations	198	99	198	198	198	198	198	204	198	192

Huber-White robust standard errors in parentheses. The dependent variable is count of non-generic drug approvals in Panel A and count of new molecular entities in Panel B, computed from the FDA dataset of New Drug Approvals. All columns except 3-5 use the income-based measure of market size; columns 3-5 use the population-based measure. All regressions include drug and time dummies, and, except for column 6, are weighted. Market size in column 4 is computed using the NAMCS data for drug use, and in column 5 market size is computed using total OECD population, as explained in text. In column 7 market size is computed using single-age groups (see text). In column 8 market size is constructed as in Table II, except that the classification from Acemoglu and Linn (2003) is used (see text). The Poisson model is estimated using Quasi Maximum Likelihood, with the Hausman, Hall and Griliches, 1984, transformation in columns 1-8 and 10. In column 9 the linear model in equation (1.12) is estimated. If a cell is empty, the log approvals variable is set equal to zero, and a dummy variable is added, equal to 1 when the cell is empty (see text).

TABLE 1.4

Potential Supply-Side and Demand-Side Determinants of Innovation

	Lag dep var	Lag dep var, IV	Life years lost	Public funding	Pre-existing trends	Major cat trends	Major cat trends, linear model	Drop Cancer, Vascular	Drop Thyroid, Anemia	Health insurance mkt size	IV with previous mkt size
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: dependent variable is count of non-generic drug approvals											
Market size	3.84 (1.07)	3.98 (1.16)	3.58 (1.70)	3.86 (1.20)	3.24 (0.91)	2.93 (5.11)	5.46 (2.43)	3.59 (1.26)	3.72 (1.11)	1.92 (0.44)	2.93 (1.45)
Lagged dependent variable	0.12 (0.10)	-0.43 (0.33)									
Panel B: dependent variable is count of new molecular entities											
Market size	3.57 (1.39)	3.69 (1.38)	3.64 (1.79)	3.56 (1.20)	3.84 (1.17)	7.64 (5.83)	3.79 (2.20)	4.59 (1.40)	3.27 (1.23)	2.10 (0.51)	3.08 (1.32)
Lagged dependent variable	-0.19 (0.11)	-1.34 (0.82)									
Number of observations	198	198	198	198	198	198	198	186	186	198	198

Huber-White robust standard errors in parentheses. The dependent variable is count of non-generic drug approvals in Panel A and count of new molecular entities in Panel B, computed from the FDA dataset of New Drug Approvals. Market Size is constructed as in Table 1.2 for columns 1-9 and 11. All regressions include drug and time dummies, use the income-based measure of market size and are weighted by expenditure. In column 2 the lagged dependent variable is instrumented with the twice lagged first difference of the dependent variable (see text). Life years lost is years prior to age 65 for each death in the U.S., calculated from the Mortality Detail Files (see text). Column 3 includes a count of total life years lost due to diseases in the corresponding category and time interval. Column 4 includes the amount of funding from NIH research grants in each category and time interval, calculated from the CRISP database. 1940/1970 trend is the log difference of drug approvals for category c between 1940 and 1970. In column 5 the 1940/1970 trend is interacted with period dummies (see text). Major drug category trends are linear time trends interacted with dummies for the 16 major drug categories (see text). In column 10, market size includes information on health-care coverage (see text). In column 11 current market size is instrumented with the market size of the same cohort 5 years earlier.

Appendix: Summary of Disease Classification and Drug Expenditure by Age Group

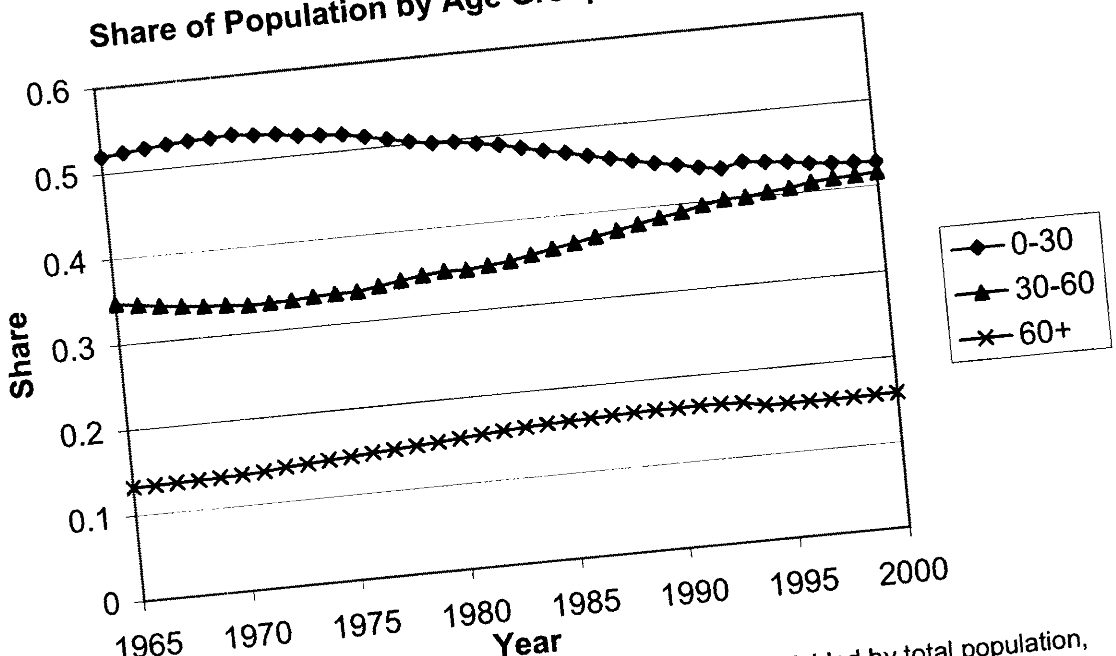
Class	Description	Expenditure share x 1000 {Share of expenditure by age group in total expenditure in brackets} 0-30	30-60	60+	Age group with largest expenditure
1	Antibiotics	0.95 {0.41}	0.62 {0.40}	0.90 {0.19}	0-30
2	Antivirals	0.03 {0.05}	0.36 {0.91}	0.05 {0.04}	30-60
3	Antiparasitics	0.01 {0.10}	0.05 {0.60}	0.08 {0.29}	30-60
4	Antifungals	0.26 {0.32}	0.23 {0.44}	0.38 {0.24}	30-60
5	Anemia	0.00 {0.07}	0.00 {0.47}	0.01 {0.47}	60+
6	Anticoagulants	0.00 {0.01}	0.06 {0.19}	0.77 {0.80}	60+
7	Glaucoma	0.00 {0.01}	0.03 {0.14}	0.58 {0.85}	60+
8	Acid/Peptic Disorders	0.17 {0.06}	0.89 {0.46}	2.87 {0.49}	60+
9	Antidiarrheals, Laxatives	0.00 {0.07}	0.01 {0.45}	0.03 {0.49}	60+
10	Cardiac	0.03 {0.01}	0.72 {0.32}	4.68 {0.67}	60+
11	Vascular	0.12 {0.02}	1.24 {0.34}	7.00 {0.64}	60+
12	Sedatives/Hypnotics, Antianxiety	0.05 {0.06}	0.25 {0.54}	0.58 {0.40}	30-60
13	Antipsychotics/Antimanics, Antidepressants	0.46 {0.13}	1.64 {0.70}	1.28 {0.18}	30-60
14	Anorexiants/CNS Stimulants	0.08 {0.52}	0.05 {0.45}	0.01 {0.03}	0-30
15	Vitamins/Minerals	0.00 {0.07}	0.01 {0.36}	0.05 {0.58}	60+
16	Electrolyte Replenishment/Regulation, Water Balance	0.01 {0.03}	0.05 {0.26}	0.46 {0.71}	60+
17	Adrenal Corticosteroids	0.05 {0.26}	0.05 {0.40}	0.14 {0.34}	30-60

Appendix (cont.)

Class	Description	Expenditure share x 1000 {Share of expenditure by age group in total expenditure in brackets} 0-30	30-60	60+	Age group with largest expenditure
18	Androgens/Anabolic Steroids	0.00 {0.03}	0.01 {0.21}	0.07 {0.77}	60+
19	Estrogens/Progestins	0.31 {0.17}	0.71 {0.58}	0.97 {0.26}	30-60
20	Contraceptives	0.11 {0.47}	0.08 {0.52}	0.00 {0.01}	30-60
21	Blood Glucose Regulators, Thyroid/Antithyroid	0.08 {0.03}	0.75 {0.43}	2.90 {0.54}	60+
22	Topical Steroids	0.01 {0.21}	0.02 {0.36}	0.06 {0.43}	60+
23	Topical Anti-Infectives	0.01 {0.32}	0.01 {0.41}	0.02 {0.27}	30-60
24	Extrapyramidal Movement Disorders	0.00 {0.01}	0.03 {0.25}	0.29 {0.74}	60+
25	Skeletal Muscle Hyperactivity, Anticonvulsants	0.21 {0.19}	0.46 {0.64}	0.39 {0.18}	30-60
26	Oncolytics	0.20 {0.19}	0.29 {0.42}	0.84 {0.39}	30-60
27	Ocular Anti-Infective/Anti-Inflammatory	0.05 {0.29}	0.04 {0.32}	0.14 {0.39}	60+
28	Topical Otics	0.01 {0.41}	0.01 {0.30}	0.02 {0.28}	0-30
29	Vertigo/Motion Sickness	0.02 {0.19}	0.02 {0.38}	0.08 {0.42}	60+
30	Pain Relief	0.27 {0.09}	1.08 {0.55}	2.09 {0.35}	30-60
31	Antiasthmatics/ Broncodilators	0.38 {0.23}	0.39 {0.37}	1.28 {0.40}	60+
32	Nasal Decongestants, Antitussives, Cold Remedies	0.05 {0.23}	0.07 {0.57}	0.08 {0.20}	30-60
33	Antihistamines, Inhalation/Nasal	0.36 {0.25}	0.51 {0.53}	0.64 {0.22}	30-60

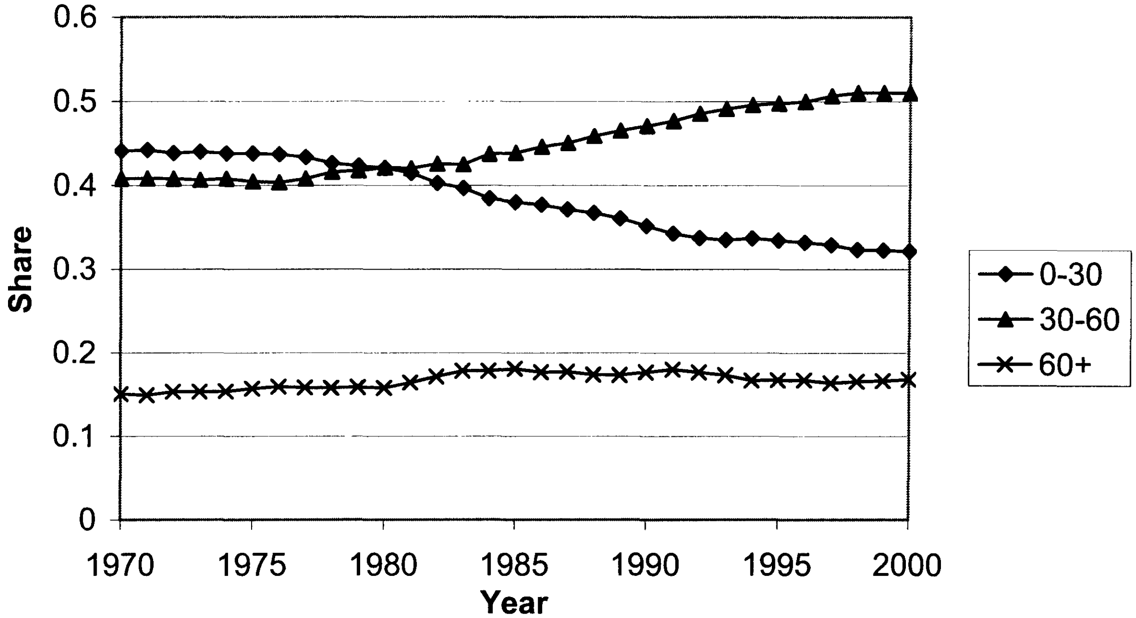
All data is from the MEPS, 1996-1998. Construction of the 33 categories is described in the Appendix. Each category includes the indicated FDA sub-categories. Expenditure share is the total expenditure on drugs in the category divided by the total income of users in that age group. Share of expenditure by age group is the fraction of total expenditure in the category accounted for by the age group. In this table shares of expenditure by category are calculated for 30-year broad age groups. Age group with largest expenditure is the broad age group with the greatest expenditure on the corresponding category.

Figure 1.1
Share of Population by Age Group from CPS, 1965-2000



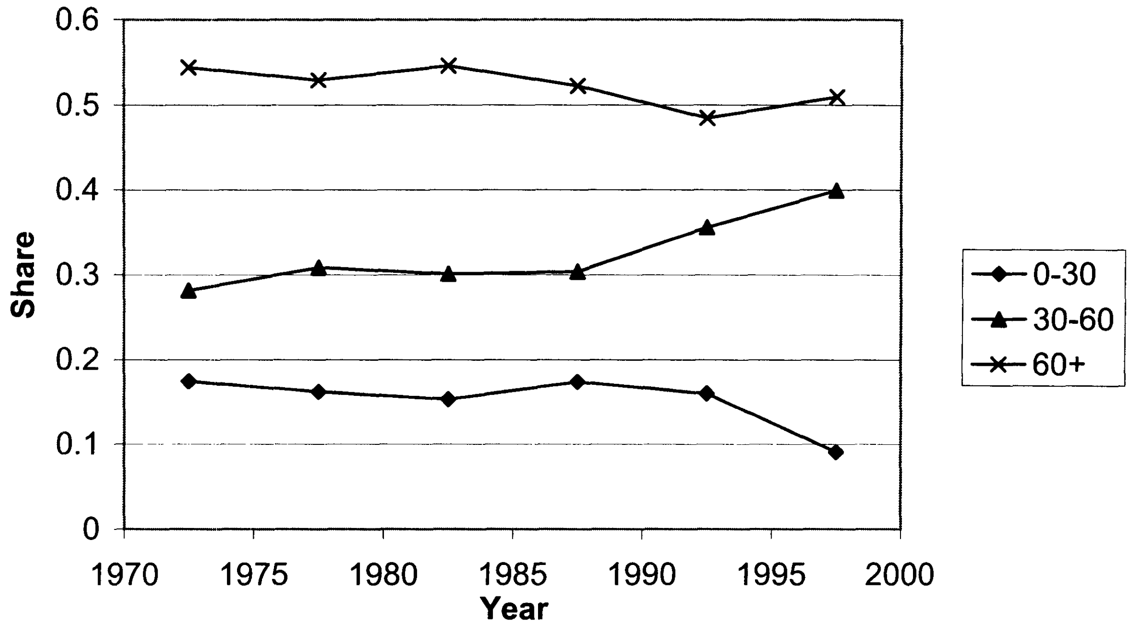
Share of population is the population of the corresponding age group divided by total population, computed from the March CPS.

Figure 1.2
Share of Income by Age Group from CPS, 1970-2000



Share of income is income of the corresponding age group divided by total income, computed from the March CPS. Individual income is obtained by dividing total family income equally among family members.

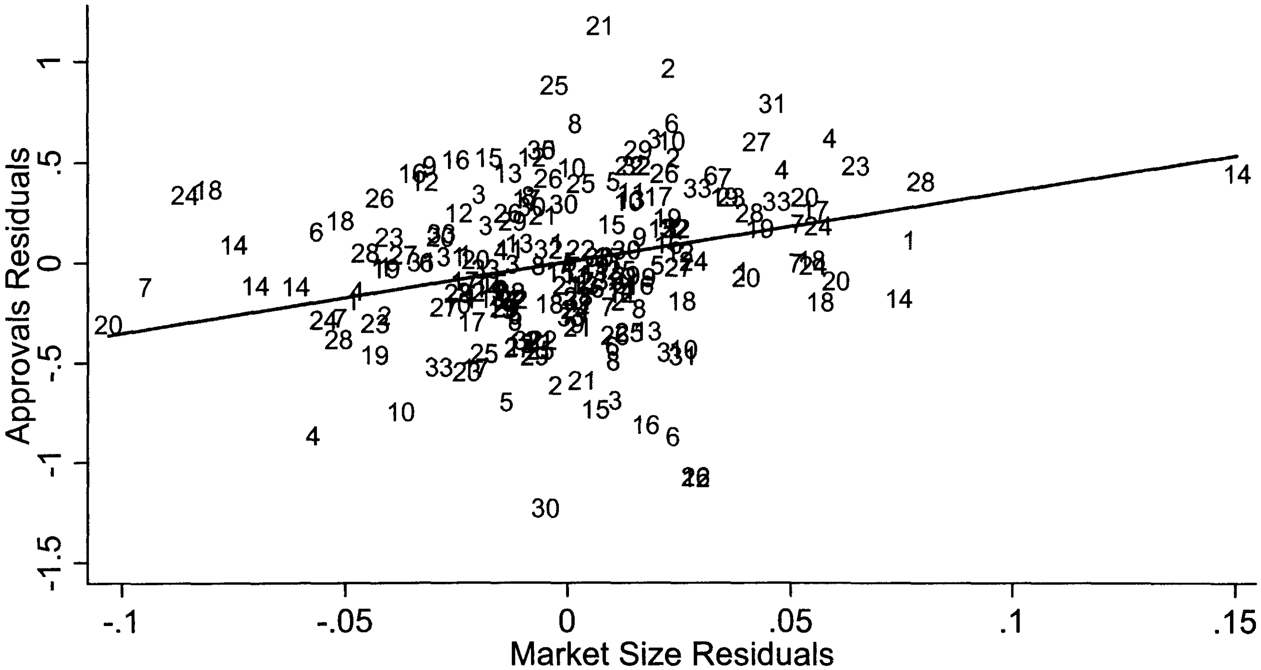
Figure 1.3
Share of FDA Approvals by Age Group, 1970-2000



Share of FDA approvals is computed as approvals of drugs in the corresponding broad age group divided by total approvals in that period, calculated from the FDA data set of New Drug Approvals. Each of the 33 drug categories is assigned to one of the three broad age groups according to which broad age group has the largest expenditure (see Appendix Table).

Figure 1.4

Approvals Residuals vs Market Size Residuals for New Molecular Entities



Approvals residuals and market size residuals are residuals from OLS regressions of log approvals and log income-based market size on category and time dummies, weighted by expenditure with five year intervals. Fitted values are predicted approvals residuals obtained from the OLS regression in Table 1.3 Panel B, column 9.

Chapter 2

Energy Prices and the Adoption of Energy Saving Technology

Summary 2 *This chapter investigates the effect of price-induced technology adoption on energy demand in U.S. manufacturing. I use plant data from the Census of Manufactures, 1967-1997, and identify technology adoption by comparing the energy efficiency of entrants and incumbents. I find a statistically significant effect of technological change, though the magnitude is small relative to changes in energy use due to factor substitution. The results suggest that technological change can reduce the long run effect of energy prices on growth, but by significantly less than previous research has suggested.*

2.1 Introduction

This paper investigates the link between factor prices, technology and factor demands. I estimate the effect of price-induced technology adoption on energy demand in the U.S. manufacturing sector, and find that an increase in the price of energy causes technology adoption which significantly reduces demand.

The relationship between technology and energy demand is important for understanding the effect of energy prices on long run growth. Amid recent concerns that rising oil prices may considerably harm the U.S. economy, Federal Reserve Bank Chairman Alan Greenspan said, "Unless oil prices fall back, some of the more oil-intensive parts of our capital stock would

... be displaced, as was the case following the price increases of the late 1970s."[1] That is, technological change would reduce both energy demand and the negative effect of energy prices on the economy.

Atkeson and Kehoe (1999) make the same argument more formally. They present a model where technology adjusts to a price increase, reducing energy demand in the long run. Technology is "putty-clay," and a plant chooses the capital to energy ratio of its machines, and cannot adjust it later. Energy demand is inelastic in the short run and a price shock causes a large decrease in output. In the long run, plants choose a higher capital to energy ratio, reducing energy demand and allowing output to recover.

In accordance with this model's predictions, total energy demand fell soon after the price increases in the 1970s, and output recovered. There is little direct evidence, however, that technological change explains these patterns. There are two primary difficulties: first, the oil shocks were aggregate, making it impossible to infer causality from a purely time series relationship between the price of energy and energy demand.

Second, aggregate data does not distinguish substitution along the demand curve for energy from a shift of the curve due to technological change. Figure 2.1 shows this ambiguity, plotting the price of energy and energy demand, $D(A, \sigma)$, where A is the level of technology and σ is the elasticity of demand to the price of energy. The economy begins at the point (E_0, P_0) and a positive shock to the price of energy causes it to move to (E_1, P_1). If the demand curve were $D^{\sigma}(A_0, \sigma_1)$, plants would substitute away from energy and move along the curve. Alternatively, substitution possibilities may be limited for the original technology, and the initial demand curve would be $D(A_0, \sigma_0)$. After adopting technology that requires less energy at any given price, the demand curve shifts to $D^A(A_1, \sigma_1)$. The latter case corresponds to the Atkeson-Kehoe model.

This paper addresses these issues in two ways, and estimates the effect of price-induced technology adoption on energy demand. First, I use plant-level data from the Census of Manufactures, exploiting geographic and time series variation in energy prices and cross-industry variation in the effect of a given price shock.

Second, I compare the energy efficiency of entering and existing plants, assuming that existing plants cannot adjust their technology. In response to a price increase, existing plants

[1] Remarks to the National Italian American Foundation, Washington, D.C., October 15, 2004.

move along their demand curve, so if entering plants use less energy than incumbents, this difference corresponds to an inward shift of their demand curve.[2]

I find that a one percent increase in the price of energy leads to technology adoption that causes a 0.1 percent decrease in energy demand for entrants. In the sample, the real price of energy doubles between 1972 and 1982, so that, relative to the difference in 1972, entrants in 1982 are 10 percent more efficient than incumbents.

The results imply that technology adoption is of lesser importance than theoretical and anecdotal evidence suggest. Substitution and changes in industry composition explain a larger fraction of the reduction in energy use in U.S. manufacturing.[3] The calculation in section 2.6 suggests that technology adoption reduces a price shock's long run effect on output by a relatively small amount.

I now discuss the identification strategy in more detail. Energy prices are potentially endogenous; for example technological change could reduce demand and lower prices. To address this concern I construct a fixed-weight index of the total price of energy. The index uses the sector-wide average prices of different fuels and electricity, and should not include the effects of technological change, substitution or industry-specific shocks to output demand.[4]

The baseline estimates rest upon three additional assumptions. First, the total factor productivity (TFP) differential between entering and existing plants is uncorrelated with the price of energy. Second, the elasticity of substitution is constant and equal for entrants and incumbents. Third, existing plants do not adopt new technology. While these assumptions are strong, they allow me to estimate a simple equation and compare the response of incumbents and entrants to energy prices. In the remainder of the empirical analysis, I document that the first and third assumptions seem to hold in practice, and that relaxing the second assumption does not affect the results.

[2] As I discuss below, to the extent that incumbents adopt technology, my results underestimate the total effect of technology adoption. Note that the adoption of newly invented machines and ones that previously existed both cause the demand curve to shift.

[3] Industry composition has not been particularly important since about 1980, but explained much of the decline in total energy demand before 1980 (Wing and Eckaus, 2004).

[4] Another issue related to the price of energy is the fact that in most of the analysis I use the current price of energy as the independent variable. That is, I assume that the current price captures the permanent component of a price shock; this would be the case if the price follows a random walk. Below I relax this assumption, and compute a forecasted price of energy, which yields similar results. Note that if the price of energy is mean-reverting, the results underestimate the actual effect of technology adoption.

More specifically, a plant's TFP affects its energy demand, so if an increase in the price of energy causes entry of plants with higher than average TFP (as in the model below) the estimate would be biased away from zero. That is, when the price is high, entrants would be more efficient because they require less of all inputs, but otherwise use the same technology. Several empirical approaches suggest this is not a major concern.

Second, regarding the substitution elasticity of entrants and incumbents, Diamond, McFadden and Rodriguez (1978) show that an identifying assumption is necessary to distinguish technological change from a change in the substitution elasticity. For example, consider Figure 2.1, and assume the economy begins on the demand curve $D(A_0, \sigma_0)$. After a price shock entrants can either be on the demand curve $D^{\sigma}(A_0, \sigma_1)$, which has a greater substitution elasticity than the original technology, or on the curve $D^{A}(A_1, \sigma_0)$, which is an inward shift of the original curve. The results below suggest that the estimate reflects the latter case; the short run price elasticity is similar for entrants and incumbents.

Finally, I assume that existing plants do not adopt technology. To the extent that incumbents adopt technology, the results underestimate the total effect of technology adoption. As documented below, the average incumbent does not invest in new machines or retire old capital in response to the price of energy, which is consistent with this assumption, and suggests that the bias may not be large.

This paper estimates the strength of the relationship between the price of energy, technology and energy demand. Theoretical work on price-induced technological change began with Hicks (1932), who argued that technology is directed towards more expensive factors. Work on induced innovation (e.g. Binswanger, 1974) and directed technological change (e.g. Acemoglu, 2002) has refined the original analysis, and predict that relative profit incentives increase the demand for certain types of technology. In this case, a rise in the price of energy increases the benefit of adopting energy-saving technology. These models typically focus on innovation and abstract from the adoption decision, but a finding of limited adoption would imply a limited amount of innovation.

A number of recent papers consider the price of energy to be an exogenous source of variation in the profitability of innovation and adoption.[5] Typically, this work focuses on specific

[5] Early empirical tests of these models were concentrated in the agricultural sector (e.g. Griliches, 1957 and

technologies in a few industries, and it is difficult to extrapolate the results to the sector- or economy-wide level. Closest to this study, Popp (2001 and 2002) looks at induced innovation and technology adoption in several manufacturing industries. He estimates the elasticity of energy-saving patents to the price of energy, and the response of energy efficiency to patents, and finds that one-third of the observed decline in energy use was due to induced innovation. In contrast, this paper reports a much smaller effect, primarily because I focus on technology adoption across the entire manufacturing sector.[6]

Doms and Dunne (1995) and Pizer *et al* (2002) find that high energy prices cause the adoption of specific types of energy-saving technologies in several manufacturing industries. They focus on the adoption of a variety of automated technologies, e.g. climate control, waste-heat recycling and adjustable speed motors. Both studies implement a different empirical strategy from this paper, exploiting cross sectional variation in energy prices in the late 1980s and early 1990s (respectively). It is difficult to compare their estimates to those reported here, but they similarly conclude that energy prices have a significant effect on technology adoption.

Several other studies complement this work, examining the effect of energy prices on technology adoption, in other areas of the economy. Jaffe and Stavins (1995) find a positive response of the adoption of thermal insulation technology to energy prices, though the magnitude is small compared to the effect of other variables. Rose and Joskow (1990) conclude that electricity generators adopt fuel saving technology in response to an increase in the price of fuel. Newell, Jaffe and Stavins (1999) report that air conditioners and water heaters became more energy efficient after increases in the price of energy, relative to periods of price stability.

A few studies have investigated the correlation of plant age with technology adoption, finding mixed results.[7] Dunne (1994) reports that younger plants are not more likely to have certain

Hayami and Ruttan, 1970), largely for reasons of data availibility.

[6]Anderson and Newell (2004) investigate the Department of Energy's (DOE) Industrial Assessment Centers program, in which the DOE performs audits of manufacturing plants and suggests projects that might reduce energy requirements. Anderson and Newell find that high energy prices increase the probability of undertaking a project, though other variables, such as expected implementation costs, have larger effects. The program covers a limited number of small and medium-sized plants who request the audit, so it is difficult to extend the results to the entire sector; nevertheless, the finding of a small but precise effect of energy prices agrees with the results presented in this paper.

[7]Much of the literature on technology adoption characterizes the factors that affect adoption, emphasizing a variety of explanations, such as plants size and expected profitability. Recent work has studied network effects (e.g. Saloner and Shepard, 1995) and the need for complementary inputs such as skilled labor (e.g. Caselli and Coleman, 2001). This paper takes energy prices to be exogenous to these factors, and the emphasis of this paper

advanced technologies in four 2-digit SIC industries. Luque (2000) models technology adoption as an investment decision, and estimates the effect of irreversibility and uncertainty. Luque finds that after controlling for these factors, small young plants are more likely to adopt technology than small old plants, though there is no correlation between age and adoption for large plants. However, these papers do not explicitly differentiate between entrants and incumbents (e.g. Luque separates plants into 15-year age groups). Future work is needed to reconcile their findings with the results of this paper, which shows that new plants use different technology from older plants.

This paper is organized as follows. Section 2.2 develops a model of technology adoption and discusses the empirical strategy, emphasizing the assumptions underlying the estimating equation. Section 2.3 discusses the measurement of the price of energy, and section 2.4 describes the data. The main results are presented in section 2.5, which also discusses the empirical support for the main assumptions. In section 2.6, I use the results to calculate the effect of technology adoption on the long run response of output to a price shock. Section 2.7 concludes.

2.2 Model and Identification

In this section, I present a simple model of technology adoption that highlights the simplifying assumptions in the empirical strategy. I derive the estimating equation, and discuss the methods for assessing the validity of the main assumptions.

2.2.1 Motivating Theory

I model technology adoption in response to an unexpected and permanent increase in the price of energy. I compare the short run response of energy demand to the shock, holding technology constant, to the long run response, when technology changes.

The Baseline Model

Consider a single industry in which plants take input prices as given and choose their energy technology and capital stock when they enter. The industry begins in the steady state, with

is on the entrant/incumbent distinction, which has received little attention in the literature.

all input prices and demand for the industry product constant. I first characterize profit maximization for existing plants, then the decisions of new plants that have entered but not yet produced, and finally the entry decision.

Each plant $i \in I$ produces output, Y_{it}, and has a production function, $f(A_i, K_i, A_i^E E_{it}, L_{it})$, where A_i is total factor productivity; K_i is the capital stock chosen by the plant; A_i^E is a state variable, representing the energy technology; E_{it} is the energy consumed at time t; and L_{it} is employment. I refer to E_{it}/Y_{it} as energy efficiency. A_i is drawn from the distribution function $F(A_i)$, with support $[0, \infty)$, and plant i learns A_i before entering. I assume a constant elasticity of substitution (CES) production function, which yields a simple solution, and is necessary for obtaining the result that the substitution elasticity does not depend on the price of energy:[8]

$$Y_{it} = A_i(\alpha_K K_i^\rho + \alpha_L L_{it}^\rho + \alpha_E(A_i^E E_{it})^\rho)^{1/\rho},$$

where $-\infty < \rho < 1$, and the elasticity of substitution is $\frac{1}{1-\rho}$.

The maximization problem of a potential entrant can be written as:

$$\max_{\{E_{it},L_{it}\}_{t=0}^{\infty},A_i^E,K_i} E_0\{\sum_{t=0}^{\infty} \frac{1}{(1+r+\phi)^t}(p_t Y_{it} - p_t^E E_{it} - p_t^L L_{it}) - g(A_i^E, K_i) - p^K K_i\} \tag{2.1}$$

$$s.t. Y_{it} = A_i(\alpha_K K_i^\rho + \alpha_L L_{it}^\rho + \alpha_E(A_i^E E_{it})^\rho)^{1/\rho},$$

where $E_0\{\cdot\}$ is the expectation operator; p_t, p_t^E, p_t^L and p^K are the prices of output, energy, labor and capital; r is the discount rate; and $g(A_i^E, K_i)$ is the cost of selecting energy technology A_i^E, where $g(A_i^E, K_i)$ is increasing in both arguments. Profit maximization includes three decisions. First, a potential entrant calculates its expected profits, and decides to enter if profits are non-negative. Second, the new plant selects its energy technology and capital stock as functions of expected future prices, and it pays a fixed cost of $g(A_i^E, K_i) + p_K K_i$. Finally, in each period, the plant purchases labor and energy after observing current factor prices; operating costs are

[8] While the CES functional form is essential for the results below, it can be somewhat generalized, without affecting the main results. For example, if capital and energy are combined according to a CES technology to form an intermediate input, which is used to produce output with labor (similar to Atkeson and Kehoe), the resulting empirical specification would be identical to that used below, with the addition of the initial capital stock as a control variable.

$p_t^E E_{it} + p_t^L L_{it}$.

Energy technology augments energy use, but because it is chosen at the same time as capital and both are fixed, it is a feature of the capital stock.[9] In other words, this is a vintage capital model, where plants purchase capital and choose the energy characteristic of the capital stock at the time of entry.

I now discuss the three components of profit maximization previously noted, in reverse chronological order. Since the production function is constant returns to scale, I fix each plant's capital stock to $K_i = \overline{K}$. I analyze the steady state, where all prices are constant, and the subscript 0 denotes the initial steady state. Having chosen A_i^E, a plant solves the following problem each period:

$$\max_{E_{i0}, L_{i0}} \{pY - p_0^E E_{i0} - p_0^L L_{i0}\}$$

$$s.t. Y = A_i(\alpha_K \overline{K}^\rho + \alpha_L L_i^\rho + \alpha_E (A_i^E E_{it})^\rho)^{1/\rho}.$$

The first order condition for E_{i0} yields the equation:

$$\ln(E_{i0}/Y) = \frac{\rho}{1-\rho} \ln A_{i0}^E - \frac{1}{1-\rho} \ln \widetilde{p}_0^E + \frac{1}{1-\rho} \ln(\alpha_E) + \frac{\rho}{1-\rho} \ln A_i, \tag{2.2}$$

where $\widetilde{p}_0^E = p_0^E/p$, the real price of energy. Equation (2.2) shows several results. First, holding the price of energy constant, a higher A_i^E can increase or reduce energy demand. When $\rho < 0$ inputs are gross complements, and a plant with better energy technology uses less energy. Second, the coefficient on the price of energy is the energy own-price elasticity, and since $\rho < 1$, it is negative; energy demand decreases when the price of energy increases. Third, productivity, A_i, affects the level of energy demand, but not the elasticities to the price of energy or technology.

Next, I examine the maximization problem of a new plant, after it has decided to enter. Each period proceeds as follows. A fixed fraction, ϕ, of plants are exogenously destroyed.[10]

[9] The assumption that the plant's capital stock and energy technology are fixed captures the main features of a model with adjustment costs, where entering plants have greater flexibility and respond more elastically to a price shock.

[10] The main results are qualitatively the same if I allow for endogenous exit and a fixed fraction of plants entering each period (the simple model cannot accomodate both endogenous entry and exit).

New plants enter and select their energy technology. All plants, new and existing, purchase factors, produce and receive revenue. Since the industry is in steady state, the choices of all inputs and energy technology are independent of time, and I drop the time subscripts.

Consider the choice of energy technology. To obtain an explicit solution, I assume the following cost function for $g(A_i^E, \overline{K})$:[11]

$$g(A_i^E, \overline{K}) = \frac{\overline{K}(A_i^E)^\gamma}{\gamma(r+\phi)},$$

where $\gamma > 1$. Note that the cost is directly proportional to the size of the capital stock, $\overline{K}$.

Solving the first order condition for energy technology gives:

$$\ln A_{i0}^E = \beta_1 \ln \widetilde{p}_0^E + \beta_2, \tag{2.3}$$

where $\beta_1 = \rho/[\gamma(\rho-1)+\rho]$. β_1 is the elasticity of technology to the expected price; I discuss its interpretation below. The second term is a constant, which depends on the output price, other input prices, and the plant's total factor productivity.

To analyze the effect of technology adoption on energy demand, I combine equations (2.2) and (2.3) to obtain:

$$\ln(E_{i0}/Y) = \sigma^T \ln \widetilde{p}_0^E + \sigma^S \ln \widetilde{p}_0^E + \beta_i^C, \tag{2.4}$$

where $\sigma^T = \frac{\rho}{1-\rho} \cdot \beta_1$, and σ^T captures the effect of price-induced technology adoption (i.e. the shift of the demand curve for energy). σ^S is the own price elasticity of substitution for energy and is negative; this captures movement along the demand curve. Finally, β_i^C is a plant specific constant which depends on productivity. Equation (2.4) is a central result because it captures the two channels by which the price of energy can affect energy demand: factor substitution

The assumption that ϕ is independent of A_i is for simplicity. The main results are unchanged if, for example, ϕ is a decreasing function of A_i.

[11] I assume plants have the same adoption costs. In a more general case, adoption costs could be $\widehat{g}(A_i^E, c_i, K_i) = \frac{K_i(A_i^E)^{\gamma+c_i}}{\gamma+c_i}$, where a higher c_i denotes more costly adoption. In this case the elasticity of energy technology to the price of energy (B_2) depends on c_i, where a lower c_i confers a greater (in magnitude) elasticity.

The function $g(\cdot)$ is chosen for analytical simplicity, and the exact form is not important for the main results. For example, if the cost were proportional to the capital stock raised to some exponent, the resulting estimating equation would include the capital stock as an additional control. As discussed below, the results are unaffected by this, or similar modifications to the baseline equation.

(σ^S) and technology adoption (σ^T).

The entry decision completes the characterization of a plant's profit maximization. Expected profits are increasing in productivity, A_i, so there is generally a cutoff productivity, $\underline{A}_i(\widetilde{p}_0^E)$, below which a potential entrant does not enter. This boundary is also a function of other prices, but I emphasize the dependence on the price of energy because the other prices remain constant in the following analysis. The number of entrants is $\int_{\underline{A}_i(\widetilde{p}_{E0})}^{\infty} dF(A_i)$, and in the steady state, this is equal to the number of exiting plants, ϕn_0, where n_0 is the number of plants operating. Note that there is a negative relationship between the average productivity of entrants and the number of entrants.

Next, I analyze a shock to the steady state and compare the energy efficiency of entering and existing plants. The industry begins in the steady state with the price of energy equal to p_0^E. At the end of $t = 0$ (after all decisions have been made and markets have cleared), there is an unexpected, one-time and permanent increase of p^E, to p_1^E.

Suppose demand for the industry's output is completely elastic at the price p.[12] The energy demand of plants that survive from the initial steady state is:

$$\ln(E_{i1}^I/Y) = \sigma^T \ln \widetilde{p}_0^E + \sigma^S \ln \widetilde{p}_1^E + \beta_i^C, \tag{2.5}$$

where the subscript 1 indicates that the price shock has occurred and the superscript I indicates plants that survive from the initial steady state. The constant β_i^C is unchanged because the output price and other input prices do not change, and σ^S and σ^T are the same as before because the elasticities do not depend on factor prices (a result of the CES assumption). The energy demand of new plants is given by:

$$\ln(E_{i1}^N/Y) = \sigma^T \ln \widetilde{p}_1^E + \sigma^S \ln \widetilde{p}_1^E + \beta_i^C. \tag{2.6}$$

where the superscript N indicates new plants.

I now use equations (2.5) and (2.6) to compare the elasticities of entering and existing plants at time $t = 1$:

[12]This assumption allows me to obtain a simple solution and does not affect the main results.

$$\frac{\partial \ln(E_{i1}^N/Y)}{\partial \ln \widetilde{p}_1^E} - \frac{\partial \ln(E_{i1}^I/Y)}{\partial \ln \widetilde{p}_1^E} = \sigma^T. \tag{2.7}$$

Thus, the difference in the two elasticities is equal to the effect of technology adoption on energy demand.[13]

This result illustrates the LeChatelier principle, that demand is more elastic in the long run when technology can adjust. σ^S is the effect of the price of energy on energy demand holding technology fixed, and σ^T is the effect of the price of energy on demand when technology changes. Since $\mid \sigma^T + \sigma^S \mid > \mid \sigma^S \mid$, long run energy demand is more elastic.

Profits are decreasing in the price of energy, so the cutoff productivity for entry, $\underline{A}_i(\widetilde{p}_1^E)$, increases. There are fewer entrants each period than in the initial steady state, and a constant fraction is destroyed each period. The number of operating plants (and hence industry output) declines monotonically to the new steady state.[14] I define e as the number of entrants each period after the shock. The fraction of plants with the new technology at time t is given by:

$$\frac{\sum_{s=0}^{t}(1-\phi)^t e}{(1-\phi)^t n_0 + \sum_{s=0}^{t}(1-\phi)^t e},$$

which converges to one.

In summary, after the price shock the energy demand for new plants is given by equation (2.6). Plants that survive from the initial steady state have the demand shown in equation (2.5). The difference in the energy-price elasticities of the two types of plants is equal to σ^T, which is the effect of price induced technology adoption on energy demand. In addition, there are fewer entrants each period than before the price shock. Since the distribution function $dF(A_i)$ does not change after the price shock, the reduction in entry is accompanied by an increase in the average productivity of entrants.

[13] The conclusion is the same if other prices change due to the price shock, for example if $p = p(p^E)$. As equation (2.3) shows, the change in the output price affects the desired technology, which I refer to as the indirect effect of the price of energy on technology. From equation (2.2), the output price effects energy demand, but this effect is the same for entrants and incumbents. Thus, comparing entrants and incumbents, the difference in elasticities is the effect of technology adoption, which now includes the indirect effect.

[14] An increase in the price of energy may also cause the exit of less efficient plants (i.e. with lower total factor productivity or worse energy technology). This would bias the estimate towards zero because incumbents would be more efficient on average when the price is high. However, after controlling for covariates, exit does not appear to be correlated with the price of energy, nor does the energy efficiency of exiting plants.

Variable elasticity of substitution

I relax the assumption that after the price shock, entering plants have the same elasticity of substitution as existing plants. For simplicity, I assume that the price shock coincides with a shock to ρ, changing it to ρ_1. This setup is sufficient to show that the difference in total energy-price elasticities in equation (2.7) depends on the effect of energy technology, as before, but also on the different substitution elasticities.[15]

The energy-price elasticity of plants remaining from before the shock is σ^S. For entering plants, the parameters from equation (2.4) differ from the baseline case because ρ changes. The long run elasticity is $\widetilde{\sigma}^S + \widetilde{\sigma}^T$, with $\widetilde{\sigma}^S$ and $\widetilde{\sigma}^T$ defined analogously as before, replacing ρ with ρ_1. Thus, the comparison of the elasticities at $t = 1$ is:

$$\frac{\partial \ln(E^N_{i1}/Y)}{\partial \ln \widetilde{p}^E_1} - \frac{\partial \ln(E^I_{i1}/Y)}{\partial \ln \widetilde{p}^E_1} = \widetilde{\sigma}^T + (\widetilde{\sigma}^S - \sigma^S). \tag{2.8}$$

There are two terms: the effect of technology adoption and the difference between the substitution elasticities of entrants and incumbents.

2.2.2 Empirical Specification

In this subsection, I derive the estimating equation from the model in subsection 2.2.1. I then discuss the assumptions necessary for this comparison to yield an accurate measure of the effect of technology adoption on energy demand.

I Return to the case where ρ is constant. The log energy efficiency of an entering plant at time t is:

$$\ln(E^N_{it}/Y) = \sigma^T \ln \widetilde{p}^E_t + \sigma^S \ln \widetilde{p}^E_t + \beta^C_i, \tag{2.9}$$

The current price of energy affects the energy efficiency of an entrant through factor substitution and the choice of energy technology.

In comparison, the log energy efficiency of an incumbent at time t is:

[15] This case yields the same conclusion as a more realistic model, in which plants choose technology from a menu of ρ and A^E.

$$\ln(E_{it}^I/Y) = \sigma^T \ln \widetilde{p}_0^E + \sigma^S \ln \widetilde{p}_t^E + \beta_i^C, \tag{2.10}$$

where $\widetilde{p}_0^E$ is the real price of energy at the time the plant entered. The price of energy affects the energy efficiency of an existing plant via substitution; by assumption, it cannot adopt technology. I define the constant N_{it}, a dummy variable equal to one if the plant is an entrant, which allows me to combine equations (2.9) and (2.10):

$$\ln(E_{it}/Y) = \sigma^T N_{it} \ln \widetilde{p}_t^E + \sigma^S \ln \widetilde{p}_t^E + \beta_i^C + (1 - N_{it})\sigma^T \ln \widetilde{p}_0^E. \tag{2.11}$$

Thus, three factors determine energy demand: technology adoption by entrants, substitution, and a plant-specific constant, which depends on productivity and whether or not the plant is an entrant. Note that by the CES assumption, the substitution elasticities are equal in equations (2.9) and (2.10), allowing me to obtain equation (2.11).

I add an error term and a matrix of controls, and assume that the total factor productivity component of β_i^C is uncorrelated with the other variables, to arrive at the estimating equation:

$$\ln(E_{it}/Y_{it}) = \delta_1 N_{it} \ln \widetilde{p}_t^E + \delta_2 \ln \widetilde{p}_t^E + \delta_3 N_{it} + \delta_4 C_i + X_{it}\eta + \varepsilon_{it}, \tag{2.12}$$

where $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ are parameters to be estimated, X_{it} is a matrix of controls with coefficient vector η, and ε_{it} is a random disturbance term. C_i is a constant and a set of cohort dummies, equal to one if the plant entered in the corresponding year and zero in the year of entry. This accounts for the dependence of the intercept on the price of energy at the time of entry. In all specifications, the matrix X_{it} includes industry-region-year interactions, to remove the effects of industry-region shocks.

The regression compares the average log energy efficiency of entering and existing plants. The parameter δ_1 measures the effect of the price of energy on the difference of log energy efficiency between entrants and incumbents. Under the simplifying assumptions discussed below, this captures the effect of price-induced technology adoption on energy demand. δ_2 measures the own price elasticity of substitution for entrants and incumbents. In the model, δ_2 is negative, which would also be true for a general concave production function. The coefficients δ_3

and δ_4 allow entrants and each cohort of incumbents to have different average efficiencies.

To address the possible endogeneity of the price of energy, I compute a fixed-weight price index, described in section 2.3. Because the weights do not vary over time, and the prices of individual fuels and electricity are sector-wide averages, industry shocks, technological change and substitution should not affect the price index.[16]

Note that I use the current price of energy. This allows for the simultaneous measurement of the substitution elasticity and the effect of technology on energy demand. This is the appropriate measure if the price of energy follows a random walk process. In the sample, the average plant operates for about 10 years, and other work (e.g., Pindyck, 1999) suggests that a random walk is a reasonable approximation over such a time horizon. Furthermore, mean reversion would imply that the results underestimate the effect of technological change. Below I relax the random walk assumption and use the a vector autoregression (VAR) to forecast energy prices, which yields similar results.

I now discuss the three main assumptions. First, I assume that the price of energy is uncorrelated with the difference in average TFP of entrants and incumbents. The estimating equation, (2.12), does not include the plant's unobserved productivity, which would cause biased estimates if that variable were correlated with the independent variables. Several approaches described below suggest that the price of energy is not strongly correlated with the productivity of entrants and incumbents (after controlling for industry by region shocks).

The second assumption is that entering and existing plants have an equal and constant substitution elasticity. Diamond, McFadden and Rodriguez argue that an identifying assumption is necessary to distinguish technological change from a change in the substitution elasticity, because of two related issues. First, as equation (2.8) shows, if entrants are more elastic, the estimate of δ_1 would be biased away from zero. This would imply that entrants can substitute more easily and are not permanently more efficient; δ_1 would measure the combined effect of a rotation and shift of the energy demand curve (see Figure 2.1).

An additional concern is that the substitution elasticity may depend on the price of energy

[16] As noted above, other factor prices and the output price may respond to an energy price shock. The identification strategy includes these indirect effects of the shock on technology adoption, acting through changes in other prices. Entrants respond to the changes in other prices by selecting a different demand curve. Having chosen their demand curve, entrants and incumbents substitute by the same amount because of the CES assumption.

(as in the case of a translog production function). Suppose entrants and incumbents have the same technology, and an increase in the price of energy causes both to move along the demand curve, changing their substitution elasticity. The estimate of δ_2 would reflect the average substitution elasticity of all plants, before and after the shock, while the estimate of δ_1 would capture the difference between the substitution elasticity of entrants and δ_2, and would be statistically significant. It appears, however, that the estimate of δ_1 reflects a shift of the demand curve. The results below support the assumption that the substitution elasticity of entrants and incumbents is the same, and that the substitution elasticity does not vary with the price of energy.

The final assumption is that existing plants do not adopt technology, i.e., there is no term $\sigma^{IT} \ln p_{Et}$ in equation (2.10). Technology adoption by existing plants would bias δ_1 downwards, so this assumption concerns the estimate's accuracy, and not a possible spurious correlation. It is straightforward to augment the model and allow existing plants to adopt technology after the price shock. Under reasonable assumptions, the effect of technology adoption on energy demand is greater for entrants than incumbents, i.e., $\sigma^T > \sigma^{IT}$.[17] The coefficient on the price of energy, δ_2, is equal to $\sigma^{IT} + \sigma^S$. Because the total elasticity for entrants is $\sigma^T + \sigma^S$, the interaction coefficient, δ_1, is equal to $\sigma^T - \sigma^{IT}$. Thus, δ_1 is less than the total effect of technology adoption on energy demand, $\sigma^T + \sigma^{IT}$. This is a second illustration of the LeChatelier principle, and it is an important result because it implies that δ_1 is a lower bound of the total effect of technology adoption on energy demand.

To assess the magnitude of the downward bias, I investigate whether existing plants adopt technology. For these plants, I find no response of investment and capital retirements to the price of energy. This suggests that existing plants do not adopt technology, but it is possible that they respond on unobserved margins.

[17]For example, suppose the marginal cost of adoption is higher for incumbents, and they recover the cost $g(A_i^E)$ before choosing their new technology. Then A_i^E is smaller for incumbents, and the effect of technology on energy demand is smaller.

Alternatively, suppose the marginal cost is the same for both types of plants, but plants differ in productivity. Incumbents pay a fixed cost to adjust their technology, so only a subset of incumbents change. Conditonal on changing, they select the same technology as entrants, but the average elasticity of demand to technology for entrants is at least as large as that of incumbents.

2.3 Construction of the Price of Energy

Before discussing the data and the main results, I describe the construction and variation of the price of energy variable. I compute the nominal price of energy as the weighted sum of fuel and electricity prices:

$$p_{ist}^{E} = \sum_{f} w_{isf} \cdot \pi_{sft}, \tag{2.13}$$

where p_{ist}^{E} is the price of energy in industry i, state s and year t, and w_{isf} is the share of energy (in BTUs) for industry i in state s for energy source f (electricity, natural gas, petroleum and coal). In 1975 the average cost shares across states and industries were 0.47 for electricity, 0.29 for natural gas, 0.17 for petroleum and 0.07 for coal. π_{fst} is the price in current dollars per million BTUs by state, source and year. I compute the real price of energy, $\widetilde{p}_{ist}^{E}$, by dividing p_{ist}^{E} by the output price for industry i in region r and year t (see below).[18]

I now discuss the variation of the price of energy in equation (2.13). It is important to note that the weights do not change over time. Thus, industry-specific substitution between energy sources and technological change, both of which are endogenous responses to energy prices, do not affect p_{ist}^{E}.

The regressions below include industry-region-year interactions and state fixed effects, so I discuss the remaining variation. That is, if the prices of energy sources are uniformly high, either in a given year and region, or in one state over the entire period, this has no effect on p_{ist}^{E}. On the other hand, a state price shock, to one energy source or to all of them, affects relative energy prices in a given industry and region (and the size of the effect varies according to the weights).

An aggregate or regional shock to an individual energy price can affect relative energy prices within a given industry-region-year cell. For example, in the Nitrogen Fertilizer industry (SIC 2873), the natural gas cost share is higher in Texas than in other states in the same region. In response to a regional shock to natural gas prices, the price of energy for Fertilizer plants in

[18] I also construct the price of electricity and fuels separately. A plant's real electricity price is the price of electricity in the corresponding state and year, divided by the output price. The nominal fuel price is computed similarly to equation (2.13), except that each weight is the cost share of the fuel in total fuel costs for the industry and state.

Texas would rise compared to Fertilizer plants in the rest of the region.

Finally, the price of energy depends on time-invariant state differences in individual energy prices due to variation in the cost shares within states. As discussed below, the ratio of the price of a particular source to other sources may be higher in some states than in others. For example, the price of coal is relatively high in Maine. Since cement plants (SIC 3241) have a higher coal cost share than the average industry, the price of energy for cement plants in Maine is higher than for other plants in the same state. Even if coal prices did not vary over time, cement plants in Maine would face a high price of energy relative to other cement plants.

I next discuss the historical patterns of the different energy prices. Figure 2.2 shows the average real prices of electricity, natural gas, petroleum products and electricity from 1967-1997. Natural gas and oil prices moved together closely, and electricity prices followed a similar pattern, but varied less over time. Coal prices were considerably different, peaking in the early 1970s and declining afterwards.

Electricity prices were regulated over most of this period at the state level, and did not respond quickly to demand. They depended, *inter alia*, on fuel costs (most often coal), capital equipment, other operating costs and transmission costs, and the composition of electricity generators (e.g. coal-fired versus hydropower). Regulators set prices to guarantee utilities a certain rate of return. There was considerable geographic, and some inter-temporal variation in the availability of hydroelectric and nuclear power, which I assume was exogenous to plant activity.

Natural gas prices varied dramatically over time and space, due to regulation, transportation costs, and oil prices. In the 1970s, federal regulation reduced the domestic supply of natural gas and contributed to a severe shortage (MacAvoy and Pyndick, 1975). Supply was completely halted to many customers, so that observed prices may not reflect the true costs. Beginning in the late 1970s, prices were gradually deregulated, and by the end of the sample, prices mainly reflected differences in transportation costs. Natural gas is often a close substitute for oil, which partly explains the high correlation of these prices across states and time.

Import prices of oil, transportation costs and regulation were the main sources of variation for petroleum product prices. For much of the period, OPEC held nominal crude oil prices stable, except during the two major shocks. Regulatory bodies such as the Texas Railroad

Commission further maintained price stability by controlling the domestic supply (Hamilton, 1984). As noted, the OPEC shocks affect the price of energy for plants in states with a high cost share of petroleum, relative to other states in the same industry and region.

Coal prices varied differently from other sources, due to the location of mines in certain regions of the country and environmental regulation. Most industries use little coal directly, though coal prices had a large effect on electricity prices. Transportation costs are particularly high for coal, relative to extraction costs. The variation in coal prices comes disproportionately from geographic variation, compared to other energy sources. Furthermore, sulfur emissions regulation in the 1970s raised the demand for low sulfur coal supplied from the west. The regulation affected both coal and electricity prices in a much different manner from the OPEC shocks and natural gas shortage.

2.4 Data Sources and Variable Construction

The main data sources are the Census of Manufactures (CM), 1963, and every five years from 1967-1997, and the Department of Energy (DOE) State Energy Price Report, 1960-2000. I supplement the CM with other data from the Longitudinal Research Database (LRD), the Annual Survey of Manufactures (ASM), 1973-1988, and the ASM fuel surveys for 1975-1981. The main variables are energy efficiency, the real price of energy and the entrant dummy variable.

Energy efficiency is in units of million BTUs per 1972 dollar of output. The CM contains output and energy expenditure in nominal dollars, so I primarily describe the price deflators. I construct a detailed price index, which varies both geographically and by SIC product code. The LRD contains product files, which for each Census year contains revenues and quantities for a large number of 7-digit products. Missing or imputed quantity data and concerns about data quality prevent me from using plant level product prices. Instead, I aggregate the plant-level 7-digit product prices. I begin with 7-digit by state prices, and compute chain-weighted prices for different aggregation levels, up to 4-digit product code. For each product a plant produces, I use the least aggregated price for which there were at least 50 plants in the product files.[19]

[19] In computing the prices I also drop prices below the 5th and above the 95th percentile, to limit the effect of

In comparison, most other studies use the 4-digit industry prices from the NBER Manufacturing Productivity Database.[20] I construct the more detailed product prices for three reasons. First, there is considerable geographic variation of output prices in the data, both within and across years. Second, because I use state-level energy price data, the detailed prices prevent all geographic variation in the real price of energy from being attributed to the DOE prices and industry weights. Third, I use the product prices to compute a plant-level output price, as the weighted average of the prices of the plant's products.[21] In most Census years the correlation between these prices and the NBER prices is high, typically over 0.80. As the results show below, using the NBER prices instead of the CM- and DOE-based prices yields similar estimated magnitudes, though the precision is considerably lower.

The nominal price of energy is the weighted sum of the state prices of the different energy sources (see equation (2.13)). The weights are the average cost shares computed from the 1975 ASM fuel survey, and they vary by industry and state.[22] The state prices are the average delivered prices for the industrial sector in a state and year, in dollars per million BTUs. I deflate the nominal energy price by the industry by region product price. The latter is the weighted sum of product prices across an industry and region.

The entrant dummy variable is equal to one when a plant appears in the CM for the first time. I use the plant identification number, which links observations of each plant over time, to construct the variable.[23] As a result, I measure entry for some plants several years after they enter (and miss plants that enter and exit between Census years). The data is not available to measure entry more precisely for the entire sample.

The dependent variable is log energy efficiency, the log of energy use divided by output,

---

outliers. The main results are insensitive to using the median price and to changing either the cutoff percentile or the number of plants needed to use a price.

[20] An exception is Foster, Haltiwanger and Syverson, who find that geographic variation in output prices greatly affects estimated productivity.

[21] Many plants within a given 4-digit industry produce products of other industries, and the NBER prices are weighted averages of the output prices of plants in the industry.

[22] The fixed weights imply that the measured price of energy over-estimates the true price index. This would bias the results if the measurement error were correlated with the entrant dummy. This does not appear to be a significant concern, however. The price of electricity and electricity efficiency do not depend on the weights, and the results are similar for electricity and total energy (see Table 7).

[23] There may be some measurement error for this variable, as plants can appear in the CM which do not operate. Using employment to impute entry (see Davis, Haltiwanger and Schuh, 1996) yields almost identical results, because most plants with zero employment also report zero output, and are dropped from the regressions.

in million BTUs per 1972 dollar of output. The numerator is the plant's nominal energy expenditure, divided by the nominal price of energy. The denominator is the nominal value of shipments, divided by the plant's output price.

The sample includes all plant-year observations in the CM, from 1967-1997. (The 1963 Census is used to identify entrants in 1967.) I drop "administrative record" plants, as is customary in working with the CM. These are typically small plants for whom most variables except payroll have been imputed. I also omit observations with non-positive output or energy expenditure. The dataset includes 1,250,203 observations for 549,094 plants.

Table 2.1 provides key descriptive statistics for the CM data. For each CM from 1967-1997, I separate plants according to whether they entered that year. The table reports the mean and standard deviation of log energy efficiency, unweighted and weighted by plant output, and the nominal cost share of energy (energy expenditure divided by total shipments).[24] The different measures show a common pattern: the difference in energy efficiency between entrants and incumbents increases when the price of energy is high.[25]

Figure 2.3 shows the same pattern graphically, for 1972-1987.[26] The series for incumbents is the output-weighted mean log energy efficiency of plants that operated in the previous Census (plants for which the entrant dummy in equation (2.12) is zero). The 1972 incumbents point includes plants that entered in 1967 and plants still operating from the 1963 census. The other lines show the mean energy efficiency by cohort.

The figure shows that entering plants are relatively more efficient than incumbents when the price of energy is high. The 1972 cohort, which entered before the price shock, is 0.19 below the incumbents in 1972. The average log energy efficiency of the 1977 cohort, most of which entered before or during the first shock, is about 0.35 below the incumbents, but is similar to the 1972 cohort. In contrast, the 1982 cohort is further below the incumbents and previous entry cohorts. The 1987 cohort, which entered as energy prices fell, is closer to the incumbents than is the 1982 cohort. Also note that the 1982 cohort remains more efficient than the other groups

[24] There is a considerable amount of variation across plants in the same entry cohort and year, some of which appears to be due to reporting errors. In particular, the standard deviation of the cost share is often close to, or greater than one. This is driven by the presence of a few outliers; in no year do more than 30 plants report a cost share larger than 1. As shown in Table 2, omitting these outlying plants from the regressions does not affect the results.

[25] The median energy efficiency (not reported) shows a similar pattern.

[26] I omit the other years to present a clearer picture.

in 1987. As discussed above, one concern is that entrants may have a different substitution elasticity than incumbents. The fact that the 1982 cohort does not increase energy demand by more than the other groups suggests that this is not the case.

2.5 Results

2.5.1 Basic Specification

This section discusses the estimated coefficients from equation (2.12), using plant data from 1967-1997. The model is estimated by Weighted Least Squares, using plant output as weights to account for the reporting errors of small plants.[27] Standard errors are robust to heteroskedasticity using the Huber-White formula. Column 1 of Table 2.2 shows the baseline specification, which includes industry-region-year and state dummies to control for demand shocks and fixed differences across states. In all regressions, I include a full set of entry cohort dummies. The estimate of the interaction of the log price of energy and the entrant dummy is -0.096, with standard error 0.032 (significant at 1 percent), meaning that the relative efficiency of entrants improves when the price of energy rises.[28] The estimate of δ_2, the elasticity of substitution with respect to the price of energy, is -0.188 with standard error 0.049. The price of energy is demeaned in all regressions, so entrants are about 8 percent more efficient on average.

The magnitudes imply that technology explains a significant, though relatively small amount of the observed changes in energy demand. For example, between 1972 and 1982 the real price of energy (weighted by plant output) rose 106 percent while energy efficiency improved by 31 percent. The estimate of the interaction coefficient thus implies that about 9 percent of the observed change in efficiency was due to technology adoption.[29] In comparison, Popp (2001) performs an analogous calculation for his set of industries, and finds that one-third of the observed decline in energy use was due to innovation. He estimates an elasticity -0.37, which is

[27] Dhrymes (1992) finds that there is less measurement error for large plants. Presumably this is because the Census Bureau is most interested in publishing aggregate statistics. As a result, it checks the responses from large plants more carefully for missing values or nonsensical answers. In the regression analysis, the measurement error mainly affects the dependent variable, so it should not bias the estimates.

[28] By comparison, the unweighted estimate is -0.059 (standard error 0.025) for the interaction term. The R-squared is 0.44, much lower than that reported in column 1, which supports using the weighted estimates.

[29] Specifications with other prices reported below gives similar results; technology adoption by entrants typically explains between 5-15 percent of the decline in energy use.

much larger than the estimate here. However, he studies several energy intensive industries, and the mean is driven by a few outlying industries. This underscores the need to measure the effect of technological change across a broad range of industries. I will return to the interpretation of the estimate in section 2.6.

Recall that some cross sectional variation of the price of energy is due to differences in industry weights within a given state. In column 2, I include a full set of industry-state interactions, to restrict the price variation to the time series sources described above. The estimate of δ_1, -0.077, is similar to the baseline (significant at the 10 percent level). This suggests that transportation costs and other time-invariant factors are not driving the results.

It is possible that the specification in column 1 does not adequately account for industry heterogeneity. In column 3, I add to the baseline specification the interactions of the log price of energy with a set of industry dummies. This allows each industry to have a different average log efficiency in each region and year, and a different substitution elasticity over the entire period. The estimate of δ_1 is quite similar to the baseline.

Next, I relax the assumption that after removing industry-region-year means, plants in the same entry cohort have the same average log energy efficiency. For column 4, I compute each industry's cost share of energy in 1963, and construct a set of dummy variables denoting the 10 deciles of the cost share distribution. I interact these dummies with the cohort dummies, and obtain similar estimates to the baseline. In column 5, I interact the cohort dummies with a set of 2-digit industry dummies, which has little effect on the estimates.

Finally, since the model is estimated by Weighted Least Squares, it is possible that outlying observations affect the results. Reporting errors are not uncommon in the CM, and are a particularly likely source of measurement error. To address this concern I calculate the 5th and 95th percentile of the energy cost share (energy expenditure divided by value of shipments) for each industry-region-year-entrant cell. Column 6 omits outlying observations, and the results remain robust.[30]

The magnitude of the energy price-entrant interaction, δ_1, is relatively stable and statistically significant after controlling for industry, geographic, and entry cohort heterogeneity, and

[30]The specification in column 6 omits outliers due to reporting errors, since the energy cost share is computed directly from values in the CM. Alternatively, the results may be affected by outlying values of energy efficiency; dropping these observations in a similar manner yields the same conclusion.

does not appear to be affected by outliers. The remaining specification checks assess the validity of the simplifying assumptions.

2.5.2 Sample Selection and Plant Heterogeneity

The assumption that the substitution elasticity is equal for entrants and incumbents is central for the identification of the baseline equation. I postpone that discussion to the next section, because it is first necessary to establish that the sample used in Table 2.2 is representative.

Since the productivity distribution depends on entry and exit, there is a problem of sample selection. Existing approaches, such as a Heckman correction or methods based on the propensity score (e.g., Angrist, 1995) require a sample where the selection decision is observed and the exogenous variables are uncorrelated with the unobserved variables. The CM does not meet these conditions, because I do not observe potential entrants who do not enter.

Consequently, I consider a regional market for entry, analogous to Dumais, Ellison and Glaeser (2002). That is, I assume that the distribution from which potential entrants are drawn, $F(A_i)$, is constant in a given industry, region and year. In the model in section 2.2.1, the number of entrants at time t is $\int_{\underline{A}_i(p_t^E)}^{\infty} dF(A_i)$. The price of energy is positively correlated with the average productivity of entrants if the cutoff productivity, $\underline{A}_i(p_t^E)$, is increasing in the price of energy. In other words, a high price of energy would cause the entry of fewer, more productive plants. Entrants would be more efficient when the price is high, introducing the possibility of a spurious correlation.

I use the log of a count of entrants for the plant's state, industry and year of entry (the same level of variation as the price of energy) to proxy for unobserved productivity. A negative correlation of this variable with the price of energy would suggest that a high price of energy causes the entry of more productive plants.

In addition, given a stable distribution of potential entrants, plants from an entry cohort with greater than average productivity should be more likely to survive to the next Census. A positive correlation of the price of energy with survival would indicate that the price of energy is correlated with the productivity of entrants.

This subsection first shows that the price of energy is not correlated with entry or survival of entrants. I then add several controls to the baseline regression which proxy for plant productivity, and find that the results are unchanged. I conclude that my estimates do not appear to be biased by sample selection.

Entry and Survival of Entrants

Table 2.3 shows the correlations between the log price of energy and entry, survival of entrants and exit of incumbents. In all regressions, the dependent variable is aggregated to the industry-state-year level. I include state fixed effects and a full set of industry-region-year interactions. I weight observations by the total output of plants in the corresponding cell. In column 1, the dependent variable is the log of the count of entrants, and the coefficient on the log price of energy is -0.109, with standard error 0.128. The estimate is insignificant, so entry does not appear to be strongly correlated with the price of energy (after controlling for industry-region shocks), though the magnitude is large.[31] This estimate would imply a large bias of the estimate δ_1 if the energy efficiency of marginal entrants were highly correlated with the price of energy (i.e., if a high price causes the entry of fewer, more productive plants). However, as Table 2.4 suggests (see below), the marginal entrant is not significantly less efficient than the average entrant, and the relatively large magnitude in Table 2.3 does not appear to be a significant concern in practice.

For a given industry and region, there may be certain characteristics that make it more profitable to enter in some states than in others, such as proximity to consumers or inputs. If these characteristics were correlated with the price of energy, then entry would be correlated with the price, even if the price does not affect the distribution of entrants. Consequently, in column 2, I use the ratio of plant entry to the total number of plants operating in the state in the base year (1963), which captures these unobserved characteristics.[32] The coefficient is -0.012, and is not significant. Since the entry rate is about 0.27, the implied elasticity is smaller than in column 1.

In column 3, the dependent variable is the log of survival, and in column 4, it is survival divided by the 1967 number of entrants. I do not include industry-state cells for 1997 because

[31] The industry-region-year interactions are particularly important. If I include only industry and year controls, for example, the price of energy is strongly and negatively correlated with entry, as one would expect given other work on the effects of energy prices on manufacturing activity (e.g., Davis and Haltiwanger, 2001).

[32] The sample size is larger in column 2 because there are many industry-state-year cells with zero entrants, which are dropped in the log specification in column 1.

I do not observe whether those entrants survive. In both cases the price of energy has an insignificant effect on survival.

For completeness, I show the results for exit by incumbents between the current and next Census. Although there is no general relationship between this measure and plant productivity, large or significant coefficients would cause concern. The estimates in columns 5 and 6 are insignificant, and the magnitudes are similar to columns 1 and 2.[33] I conclude that after controlling for state fixed effects and industry-region-year interactions, the price of energy is not strongly correlated with entry and exit.

Controls for Unobserved Heterogeneity

Table 2.4 reports the results of several tests of bias due to unobserved heterogeneity. Although the estimates in Table 2.3 are insignificant, the magnitude in column 1 is large, and it is important to assess whether the decrease in entry is associated with an increase in the energy efficiency of entrants. Columns 1 through 4 use the dependent variables from columns 1 through 4 in Table 2.3 as controls. I use the entry or survival measure in the year the plant entered, so the variable is constant during the life of the plant. In each case the estimates of δ_1 and the substitution elasticity are unchanged from the baseline estimates.[34] The coefficient on log entry in column 1 is small and insignificant, which suggests that this variable is not strongly correlated with the average efficiency of entrants, as would be the case if sample selection were driving the results. The survival variables are negative and significant, in agreement with the hypothesis that plants with greater TFP use less energy and are more likely to survive, but this does not affect the main estimates.

Recent methods of controlling for unobserved heterogeneity rely on predicted relationships between unobserved and observed variables. For example, Olley and Pakes (1996) argue that log investment can proxy for unobserved productivity, but missing values (i.e., zero reported

[33] If energy-intensive plants are more likely to exit when the price of energy is high, this could bias the estimate of the effect of technology adoption. This does not appear to be a significant concern; although plants that use more energy are more likely to exit on average, there is no correlation between the price of energy and the probability that an energy-intensive plant exits.

[34] The baseline estimates for these samples are almost identical to the estimates in columns 1-4. For example, the estimate of the interaction term using baseline specification for the sample in column 4 is -0.086 with standard error 0.030.

investment) greatly reduce the sample size. Levinsohn and Petrin (2001) use intermediate materials, but the data does not support the separability assumption that this method requires. Consequently, I do not report the results from implementing these methods, though the baseline results are unaffected by including the log of investment or materials.

Other research in the manufacturing sector (e.g. Audretsch and Mahmood, 1995) has observed that plants within the same year and industry enter with different capital stocks and employment, possibly due to productivity differences. In columns 5 and 6, I control for the log capital stock and log employment the year the plant enters. I only observe capital stock from 1972-1992, so I begin with that year and assign 1972 incumbents their 1972 value.[35] Both of the productivity proxies are positive and precisely estimated, but the other coefficients are unaffected.

2.5.3 Substitution Elasticities of Entrants and Incumbents

This subsection considers the CES assumption, that entrants and incumbents have the same substitution elasticity, and that the substitution elasticity is constant for a given technology. The results provide general support for the assumption.

Subsection 2.2.1 shows that if entrants' technology allows for easier substitution, the interaction term would overestimate the shift of the demand curve. In column 1 of Table 2.5, I estimate a separate substitution elasticity for incumbent plants operating in each year. I then test whether plants that entered in a given year have different substitution elasticities from the corresponding elasticity of incumbents. I estimate the following equation:

$$\ln(E_{it}/Y_{it}) = \phi_0 \ln p_t^E + \sum_j (\phi_j O_j \ln p_t^E + \varphi_j T_j \ln p_t^E) + \sum_j (\iota_j O_j + \kappa_j T_j) + X_{it}\eta + \varepsilon_{it}, \quad (2.14)$$

where j indexes Census years. O_j is a set of dummy variables, equal to one if the plant operates in the corresponding year (i.e., it is constant over the life of the plant). T_j is a set of dummy variables equal to one if the plant enters that year, and is also constant over the life of the plant.

[35] The estimate of the interaction terms using the corresponding samples with the baseline specification are –0.090 (standard error 0.031) and -0.094 (standard error 0.033).

The term $O_j \ln p_t^E$ is the interaction of O_j with the price of energy, and similarly for $T_j \ln p_t^E$. This specification estimates a separate substitution elasticity for each year, maintaining the assumption that existing plants do not adopt technology. The energy price-entrant interaction allows the substitution elasticity of an entry cohort to differ from that of the corresponding population of incumbents. As column 1 shows, these interactions are generally small and insignificant; only the 1992 interaction is significant at 10 percent. Furthermore, an F test fails to reject the hypothesis that the interactions are jointly equal to zero. These results suggest that entering plants have a similar substitution elasticity to incumbents.[36]

A second concern is that the substitution elasticity may not be constant for a given technology. If this were the case, an increase in the price of energy would cause the elasticity of existing plants to change as they move along their demand curve. Entrants would have a different elasticity from the average of incumbents, even if they use the same technology. Columns 2-4 return to the baseline specification, and allow the substitution elasticity of incumbents to vary. Column 2 includes price of energy by year interactions, which allows for a different average substitution elasticity of incumbents each year. The specifications in columns 3 and 4 estimate a separate substitution elasticity by region and year, and by 2 digit industry and year. In each case the estimate of δ_1 is quite similar to the baseline. Note that the reported coefficient on the price of energy depends on the omitted category, which explains the discrepancy with the baseline estimate in Table 2.2.

2.5.4 Robustness

Alternative Measures of the Price of Energy

In Table 2.6, I use several alternative measures of the price of energy and find broadly similar results. As discussed earlier, much of the variation in the price of energy is driven by aggregate shocks to the price of oil and the natural gas shortage. The geographic variation reflects regulatory and transportation differences, for which the exogeneity argument may be weaker. In column 1, I use the log aggregate price of energy, which is the log of a weighted average of the industry by state energy prices. The estimate of the interaction is -0.163 with standard error

[36] The conclusions are not driven by cross-plant heterogeneity. The results are similar for a specification that takes first differences of the variables.

0.063, which is larger than the estimate in column 1 of Table 2.2. The fact that the estimate is negative and significant suggests that the main results do not arise from cross sectional price variation.[37]

Recall that in the model I consider a one-time and permanent increase in the price of energy. In that case, or if the price of energy follows a random walk, the current price is the appropriate measure of the expected future price. Note that if the actual price is mean-reverting, the results underestimate the effect of technology adoption on energy demand. In column 2, I relax the random walk assumption. I use forecasted prices generated from a VAR of the prices of residual, distillate, natural gas and electricity, with two lags, using the cross section of state prices from 1960-2000.[38] I construct the forecasted price for each energy source as the discounted sum of forecasted prices for the following 10 years, weight these prices and take the log, to obtain the log forecasted aggregate energy price. The estimate in column 2 is significant at the 1 percent level, and is similar to the corresponding estimate with the aggregate price of energy in column 1. The forecasted price varies by about three-fifths as much as the actual price; for example, between 1972 and 1982 the actual price increases by 106 percent, while the forecasted price increases by 63 percent. Thus, the implied reduction in energy demand due to technology adoption is smaller for the forecasted price.

In columns 3 and 4, I account for the fact that some entrants operated before the corresponding Census year (i.e., 1977 entrants may have first operated between 1973-1977), and that technology adoption may respond gradually to the price of energy. In column 3, I use the three year lag, and in column 4 the log of the average price over the previous five years, for the plant's state and industry. In both cases the estimate of δ_1 is quite similar to the baseline.

Finally, in column 5 I use the energy and shipments prices from the Manufacturing Productivity Database, to check the quality of the constructed prices. The prices vary by industry and year, so the industry-region-year interactions absorb the main effect of the price of energy. The coefficient on the interaction is smaller than the estimate in Table 2.2, and the standard error is considerably larger, validating the use of the DOE- and LRD-based prices.

[37]The standard errors are considerably larger if I use the aggregate price and cluster the standard errors by year: for example, the standard error on the interaction term is 0.096, and the estimate is significant at the 10 percent level.

[38]It is also possible to estimate a separate VAR for each region. In many cases a unit root process cannot be rejected, and I do not report the corresponding results.

Additional Robustness Checks

I discuss the results from a variety of alternative specifications. Several papers report an asymmetric response to the price of energy. For example, Davis and Haltiwanger (2001) find that employment responds more to a price increase than to a decrease. Learning by doing may explain an asymmetric effect of technology adoption.[39] To investigate this possibility, I compute a dummy variable equal to one if the price of energy is at least as large as it was in the corresponding state and industry, five years earlier. The interaction of this variable with the other independent variables allows for both an asymmetric elasticity of substitution and an asymmetric effect of technology adoption on energy demand. The results suggest that technology adoption may respond more to a price increase, though the imprecision of the estimates prevents a strong conclusion. The estimate of the average effect of technology adoption is -0.054, with standard error 0.030, considerably smaller than the baseline and insignificant. The magnitude of the additional effect of a price increase is large, -0.058, but it is insignificant, with standard error 0.054.

Table 2.7 reports several additional results. Columns 1 and 2 show that clustering the standard errors by year or plant does not affect the precision of the estimates.

The response of technology may differ for electricity and fuels, since the prices varied differently over the sample (see Figure 2). In columns 3 and 4, I separate energy use into electricity and fuel consumption divided by output (both in million BTUs per 1972 dollar of output). For comparison, the estimates of δ_1 for log energy efficiency, with the samples corresponding to columns 3 and 4, are -0.090 (standard error 0.038) and -0.091 (standard error 0.037). Thus the fuel efficiency estimate is smaller (and significant at 10 percent), and the electricity estimate is slightly larger than the baseline, though they are fairly similar to one another.

The remainder of Table 2.7 tests for various forms of heterogeneity of the response of technology to the price of energy. Other work with the Census of Manufactures (e.g., Dunne, Roberts and Samuelson, 1989) suggests that plants belonging to multi-plant firms behave differently from single-plant firms. For example, entering plants that are part of pre-existing firms

[39] Consider an initial period of low energy prices, when the cost of adopting efficient technology is relatively high. If the price increases and plants adopt better technology, this reduces the cost of future adoption. If the price falls to the initial level, plants may use better technology than before the initial shock because it is less costly to adopt it.

may have a better knowledge of energy-saving technology. To test this possibility, I construct a dummy variable equal to one if the entering plant is a new firm (or part of a new firm). I interact the dummy with the log price of energy and include both variables in the baseline specification in column 5. The coefficient on the energy price-new firm interaction (unreported) is 0.055 (standard error 0.056) while the estimate of δ_1 is -0.123 (standard error 0.036). This suggests that plants in new firms may differ somewhat, though the difference is not precisely estimated.

Columns 6-8 check for heterogeneity across industries. In column 6, I omit industries in the top quartile of the distribution of energy cost share. The estimates are robust to restricting the sample in this manner.[40]

Beginning in the early 1980s, many plants produced electricity and sold the surplus. For these plants, changes in the price of energy affect output as well as input prices. I cannot observe electricity sales, so in column 7 I drop the four 2-digit industries for which electricity sales comprise the largest share of output, as identified in the Edison Electric Institute publication, Capacity and Generation of Non-Utility Sources of Energy (1991). These industries are Paper, Chemicals, Petroleum and Primary Metals. The estimates are similar to the baseline.

In column 8, I exploit the fact that coal prices varied differently from other energy prices. This provides an additional check for the existence of an omitted variable correlated with the price of energy. As discussed in section 2.3, the sources of variation of coal prices were different from other energy sources (see Figure 2.2). Consider an omitted variable correlated with natural gas and oil prices, which affects the energy efficiency of entrants. It is unlikely that this variable is strongly correlated with coal prices, so it would have a smaller effect on plants in coal-intensive industries. In column 8, I allow these industries to have a different response to the price of energy. I interact the main independent variables with a dummy variable, equal to one if the plant's industry is in the 90th percentile of the cost share of coal in total energy expenditure. The estimate of δ_1 is only slightly larger than the baseline, and none of the coal industry interactions are significantly different from zero, which suggests that these industries respond similarly to the price of energy.

[40] During the 1970's, the Federal government subsidized R&D aimed at reducing energy demand. This funding was aimed exclusively at energy intensive industries, so the specification in column 6 is a joint test of heterogeneity across industries and the effects of the subsidy.

During the natural gas shortage, supply to many areas was eliminated, and the observed price of natural gas may not always reflect the true cost. Since I cannot observe the availability of natural gas to an individual plant, I use the ASM to identify counties from 1975-1981 with unusually low natural gas use (controlling for industry composition). In column 9, I omit observations of plants in counties belonging to the bottom 15th percentile of the distribution, and the estimates are unchanged from the baseline.

It is possible that federal environmental regulation (the Clean Air Act, Clean Water Act, etc.), which accompanied some of the price shocks, affected the demand for energy. The provisions of the Clean Air Act may have affected both entrants and incumbents. For example, existing plants could avoid installing cleaner technology if they did not undertake major investment, and entering plants in non-attainment counties were subject to additional regulation, relative to entering plants in other counties. Also note that the most energy-intensive industries were also among the most regulated.

As a proxy for (unobserved) regulatory stringency, I use county level attainment status for several pollutants with respect to the Clean Air Act. Greenstone (2002) uses the same data and describes it in detail. I focus on Carbon Monoxide (CO), which Greenstone finds has the largest effect on manufacturing activity. I construct a dummy variable equal to one when the county is designated non-attainment, and interact it with a dummy equal to one if the plant's industry emits CO, to obtain the CO non-attainment dummy. Column 10 includes this variable plus its interaction with the price of energy. Since the attainment data is not available before 1972 and for some counties after 1972, I rerun the baseline specification with this sample. The estimates are nearly identical to those in column 10. In addition, the CO main effect and interaction variables are small and insignificant (unreported).

I perform two additional robustness checks, not reported in Table 2.7. Plants may substitute towards intermediate materials when energy prices rise, for example, if they find it less expensive to import some energy intensive goods. Since the regressions use gross output, this substitution would appear to increase energy efficiency, even though plants do not produce more from a given amount of energy. This does not appear to be a major concern, however; estimating the baseline specification with value added (i.e., gross output minus materials and energy, deflated by the appropriate value added deflator), yields identical results.

Second, the fact that I cannot observe the precise year of entry introduces measurement error. It is possible to determine the entry year for plants that enter after 1975. If I redefine the entrant variable, so that it equals one only if the plant entered in the Census year, the results are very close to estimating the baseline specification (where entry is equal to one if the plant did not appear in the previous Census), using the same sample. Consequently, this source of measurement error does not appear to bias the estimates.

2.5.5 Technology Adoption by Existing Plants

I investigate the response of incumbents to the price of energy. I use the CM and ASM from 1972-1988, to construct a balanced panel of plants, so that the results are not affected by entry and exit. Since the sample may be selected on underlying parameters (e.g., total factor productivity), the estimates may differ from those of the entire population of plants.

I find that the baseline estimate of the substitution elasticity reflects a plant's ability to change energy use, and not cross-plant heterogeneity. Investment and capital retirements do not respond to the price of energy, and I find no evidence of technology adoption by existing plants.

Substitution Elasticity

There are two concerns related to the substitution elasticity. First, the CES assumption implies that existing and entering plants have the same elasticity. An implication is that the elasticity of plants in the balanced panel should be similar to the overall estimate. Second, the baseline estimate may reflect variation in energy use across plants, rather than a plant's ability to substitute.

The sample includes 11,565 plants appearing in every CM and ASM from 1972-1988. In column 1 of Table 2.8, I use the baseline specification (industry-region-year dummies and state fixed effects) and regress log energy efficiency on the log price of energy. The estimate is -0.192 with standard error 0.087, which is quite similar to the estimate in Table 2.

In column 2, the variables are in first differences and the regression includes industry-region-year dummies; the estimated elasticity is -0.174, with standard error 0.053. This is similar to column 1, and suggests that the estimated substitution elasticity measures the average plant's ability to adjust energy consumption, and does not reflect cross-plant variation.

Investment and Retirements

It is likely that incumbents increase investment and capital retirements when they adopt technology. The remainder of Table 2.8 investigates whether the price of energy causes investment or retirements.

I compute each plant's capital stock using the book value of machines in 1972 and the perpetual inventory method. I use 2-digit depreciation rates from the Bureau of Economic Analysis and deflate investment using the Manufacturing Productivity Database investment deflators. Log investment is the first difference of the log capital stock.

Columns 3-5 report results for investment. I include the price of energy and its interaction with the 4-digit industry energy cost share from 1963. The interaction captures the effect of the price of energy on energy-related investment, since investment by plants in energy intensive industries is more likely to include energy saving technology. All variables are in first differences, to remove plant fixed effects, and all regressions include industry-region-year interactions. I use the current price of energy in column 3, the 3-year lag price of energy in column 4 and the aggregate forecasted price of energy in column 5. I do not find a positive response of investment to any of the price measures. However, this does not establish that existing plants do not adopt technology, because I only measure total investment, and I do not observe investment in specific types of capital.

If a plant adopts technology, it likely retires capital with the old technology. Retirement data may be less noisy than the investment measure, and I report results using data on capital retirements for the years 1977-1988 (due to data availability). As columns 6-8 show, log retirements do not respond to any of the measures of the price of energy. There is no evidence that the average existing plant adopts technology, but without more detailed capital data, it is impossible to make further progress on this question.

2.6 Implications for Energy Prices and Growth

The Introduction noted the connection between the price of energy and economic activity, and recent concerns over the prospects of rising energy prices. In this section I use the empirical estimates to calculate the effect of technology adoption on output in the long run. The results suggest a limited effect of technological change.

I consider a 10 percent permanent increase in the price of energy, which represents the effect of declining natural resource supplies, or a carbon tax. I compare the effect on steady state output for two models: one without technological change, and one with technological change.

I use the baseline model in section 2.2.1, except that I fix the number of plants in the industry. A constant number of plants enter and exit the industry in each period, so industry output is determined by output per plant.

I define the elasticity, ε_Y^{NT}, as the elasticity of output to the price of energy in the model without technology adoption. An increase in the price of energy causes energy consumption to decrease, lowering output.

Analogously, ε_Y^T is the elasticity of output for the model with technology adoption. In response to the shock, entering plants use better energy technology, which causes the demand curve for energy to shift towards the origin. This increases the level of output for a given amount of energy use, and allows output to recover in the long run.

To compare the two models, I calculate the ratio $\varepsilon_Y^T/\varepsilon_Y^{NT}$, which measures the amount that long run output falls in the model with technological change, relative to model without technological change. This ratio is equal to $1-\varepsilon_{AE}$, where ε_{AE}, is the elasticity of energy technology to the price of energy. This elasticity is unobserved, but can be calculated from the empirical estimates, using the fact that the estimate of σ^T, the effect of technology adoption on energy demand, is equal to $\frac{\rho}{1-\rho}\varepsilon_{AE}$ (see equation (2.4)). Thus, $\varepsilon_{AE} \simeq 0.12$, and the long run decrease of output in the model with technological change is about 88 percent of the decrease of the model without technological change. This suggests a much smaller effect of technology adoption than the analogous results of Atkeson and Kehoe. However, this is a rough calculation which relies on the functional form assumptions of the model and other simplifications. Future work may use the empirical estimates presented in this paper to calibrate a general equilibrium model.

2.7 Conclusions

This paper measures the effect of price-induced technology adoption on energy demand. The results have important implications for understanding the long run relationship between the price of energy and growth. I find that a 1 percent increase in the price of energy leads to a 0.1 percent increase in the relative efficiency of entering plants versus incumbents. I find general support for the simplifying assumptions, and the estimate is robust to a variety of specification checks. The calculation in section 2.6 suggests that technology adoption has a small effect on the sensitivity of steady state output to the price of energy.

The results are important for two other recent discussions. First, several papers (e.g., Goulder and Schneider, 1999) have incorporated price-induced technological change into climate change models. Popp (2004) calibrates a long run growth model and calculates the net benefit of implementing a carbon tax to reduce greenhouse gas emissions. The general conclusion from these studies is that accounting for endogenous technological change increases the benefit of implementing the policy, but not by a very large amount. However, to accurately assess the role of price-induced technological change, these models should be calibrated to appropriate empirical estimates; future work may incorporate the results reported here.

Second, recent energy price increases have renewed interest in the effects of energy prices on the value of installed capital. Alpanda and Peralta-Alva (2004) argue that the 1973/4 oil shock caused the obsolescence of energy intensive capital, which explains the coinciding decline in the stock market. In contrast, Wei (2003) argues that the effect of the price of energy on the value of capital should be roughly proportional to its cost share, and that the oil shock did not cause the decline in the stock market, and subsequent investment in energy-saving capital. The results from this paper cannot be directly integrated into these models, so further work is needed to assess whether the oil shocks led to the replacement of a large fraction of the capital stock, and whether this mechanism can explain the apparently large effects of energy prices on the economy.

2.8 References

Acemoglu, Daron, 2002. Directed Technological Change. *Review of Economic Studies*, 69, 781-809.

Adelman, M. A., 1995. *The Genie out of the Bottle: World Oil Since 1970.* MIT Press, Cambridge, MA.

Alpanda, Sami and Adran Peralta-Alva, 2004. "Oil Crisis, Energy-Saving Technological Change and the Stock Market Crash of 1973-74." mimeo.

Anderson, Soren T. and Richard G. Newell, 2004. Information Programs for Technology Adoption: the Case of Energy-Efficiency Audits. *Resource and Energy Economics*, 26, 27-50.

Angrist, Joshua, 1995. Conditioning on the Probability of Selection to Control for Selection Bias. *NBER Technical Working Paper* #181.

Atkeson, Andrew and Patrick J. Kehoe, 1999. Models of Energy Use: Putty-Putty versus Putty-Clay. *American Economic Review*, 89, 1028-43.

Audretsch, David and T. Mahmood, 1995. New Firm Survival: New Results Using a Hazard Function. *Review of Economics and Statistics*, 77, 97-103.

Binswanger, Hans, 1974. A Microeconomic Approach to Induced Innovation. *Economic Journal*, 84, 940-958.

Capone, Charles and Kenneth Elzinga. Technology and Energy Use Before, During and After OPEC: The U.S. Portland Cement Industry. *The Energy Journal*, 93.

Caselli, F. and W. Coleman, 2001. Cross-country Technology Diffusion: The Case of Computers. *American Economic Review*, 91, 328-35.

Davis, Steven and John Haltiwanger, 2001. Sector- Job Creation and Destruction Responses to Oil Price Changes. *Journal of Monetary Economics*, 48, 465-512.

Davis, Steven, John Haltiwanger and Scott Schuh, 1996. *Job Creation and Job Destruction.* MIT Press, Cambridge, MA.

Dhrymes, Phoebus, 1992.

Diamond, McFadden and Rodriguez, 1978. Measurement of the Elasticity of Substitution and Bias of Technical Change, in *Production Economics: a Dual Approach to Theory and Applications*, vol 2, 125-148. Fuss, Melvyn and Daniel McFadden, eds., Amsterdam: North-Holland.

Doms, Mark E. and Timothy Dunne, 1995. Energy Intensity, Electricity Consumption, and Advanced Manufacturing Technology Usage. *Technological Forecasting and Social Change*, 49, 297-310.

Dumais, Guy, Glenn Ellison and Edward Glaeser, 2002. Geographic Concentration as a Dynamic Process. *Review of Economics and Statistics*, 84, 193-204.

Dunne, Timothy, 1994. Plant Age and Technology Use in U.S. Manufacturing Industries. *RAND Journal of Economics*, 25, 488-99.

Dunne, Timothy, Mark Roberts and Robert Samuelson, 1989. The Growth and Failure of U.S. Manufacturing Plants. *Quarterly Journal of Economics*, 104, 671-98.

Foster, Lucia, John Haltiwanger and Chad Syverson, 2004. Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability? mimeo

Goulder, Lawrence H. and Stephen H. Schneider, 1999. Induced Technological Change and the Attractiveness of CO_2 Abatement Policies. *Resource and Energy Economics*, 21, 211-53.

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, 2002.

Griliches, Zvi, 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica*, 25, 501-22.

Hamilton, James, 1984. Oil and the Macroeconomy Since World War II. *Journal of Political Economy*, 91, 228-48.

Hayami, Yujiro and Vernon Ruttan, 1970. Factor Prices and Technical Change in Agricultural Development: The U.S. and Japan, 1880-1960. *Journal of Political Economy*, 77, 1115-1141.

Hicks, John, 1932. *The Theory of Wages*. London: McMillan.

Jaffe, Adam and Robert Stavins, 1995. Dynamic Incentives of Environmental Regulations: the Effects of Alternative Policy Instruments on Technology Diffusion. Journal of Environmental Economics and Management, 29: S43-63.

Levinsohn, James and Amil Petrin, 2001. Estimating Production Functions Using Inputs to Control for Unobservables. mimeo

Luque, Adela, 2000. An Option Value Approach to Technology in U.S. Manufacturing:

Evidence from Plant Level Data. *Center for Economic Studies working paper #00-12.*

MacAvoy, Paul and Robert Pindyck, 1975. *The Economics of the Natural Gas Shortage, 1960-1980.* American Elsevier Publishing Company, New York.

Mansfield, E., 1968. *Industrial Research and Technological Innovation.* Norton, New York.

Newell, Richard, Adam Jaffe and Robert Stavins, 1999. The Induced Innovation Hypothesis and Energy-Saving Technological Change. *Quarterly Journal of Economics*, 114, 907-940.

Olley, Steven and Ariel Pakes. The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 64, 1263-98.

Pindyck, Robert S., 1999. The Long Run Evolution of Energy Prices. *The Energy Journal*, 20, 1-27.

Pizer, William A., Winston Harrington, Raymond J. Kopp, Richard D. Morgenstern and Jhih-Shyang Shih, 2002. Technology Adoption and Aggregate Energy Efficiency. *Resources for the Future Discussion Paper*, 02-52.

Popp, David, 2001. The Effect of New Technology on Energy Consumption. *Resource and Energy Economics*, 23, 215-39.

Popp, David, 2002. Induced Innovation and Energy Prices. *American Economic Review*, 92, 160-80.

Popp, David, 2004. ENTICE: Endogenous Technological Change in the DICE Model of Global Warming. *Journal of Environmental Economics and Management*, 48, 742-68.

Rose, Nancy and Paul Joskow, 1990. The Diffusion of New Technology: Evidence from the Electric Utility Industry. *RAND Journal of Economics*, 21, 354-373.

Rotemberg, Julio and Michael Woodford, 1996. Imperfect Competition and the Effects of Energy Price Increases on Economic Activity. *Journal of Money, Credit and Banking*, 28, 549-77.

Saloner, G. and A. Shepard, 1995. Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines. *RAND Journal of Economics*, 26, 479-501.

Wei, Chao, 2003. Energy, the Stock Market, and the Putty-Clay Investment Model. *American Economic Review*, 93, 311-24.

Wing, Ian Sue and Richard Eckaus, 2004. Explaining Long Run Changes in the Energy

Intensity of the U.S. Economy. mimeo.

Table 2.1

Summary Statistics from the Census of Manufactures, 1967-1997

	1967	1972	1977	1982	1987	1992	1997
				Panel A: Entrants			
Log Energy Efficiency	-2.111	-2.193	-2.539	-3.133	-2.824	-2.641	-2.577
	(1.063)	(0.934)	(0.994)	(1.154)	(1.149)	(1.225)	(1.266)
Weighted Log Energy Efficiency	-2.584	-2.579	-2.794	-3.362	-3.079	-3.121	-3.145
	(1.178)	(1.108)	(1.208)	(1.480)	(1.516)	(1.572)	(1.748)
Energy Cost Share	0.027	0.019	0.022	0.024	0.025	0.027	0.027
	(0.639)	(0.158)	(0.126)	(0.133)	(0.459)	(0.572)	(0.968)
				Panel B: Incumbents			
Log Energy Efficiency	-2.236	-2.221	-2.507	-2.892	-2.682	-2.595	-2.580
	(1.026)	(0.946)	(1.004)	(1.117)	(1.085)	(1.161)	(1.226)
Weighted Log Energy Efficiency	-2.465	-2.382	-2.454	-2.724	-2.830	-2.858	-3.038
	(1.097)	(1.072)	(1.215)	(1.264)	(1.297)	(1.451)	(1.125)
Energy Cost Share	0.018	0.017	0.022	0.039	0.023	0.023	0.019
	(0.032)	(0.035)	(0.043)	(4.624)	(0.664)	(0.520)	(0.471)
				Panel C: Aggregate Statistics			
Entry Rate	0.23	0.30	0.28	0.27	0.25	0.28	0.26
Number of Plants	146,779	160,339	171,356	188,802	181,720	197,904	203,203

Variables are constructed from the Census of Manufactures (CM), 1967-1997. Panel A includes plants in the indicated Census year that did not appear in the previous Census. Panel B includes plants that operated in the previous Census. Energy use for each plant is computed as reported energy expenditure divided by the price of energy for the plant's state and industry, in million BTUs (see text). Output is the reported total value of shipments divided by the plant's output price deflator (see text). Log energy efficiency is the log of the ratio of energy use to output across plants. Each cell reports the mean, with the standard deviation in parentheses. Weighted log energy efficiency uses plant output as the weight. Energy cost share is the ratio of energy expenditure to value of shipments. Entry rate is the number of entrants in the subsample, divided by the total number of plants.

Table 2.2

Response of Energy Efficiency to the Price of Energy, 1967-1997

	Baseline: industry x region x year, cohort and state	Baseline and industry x state	Baseline and industry x price of energy	Baseline and cohort x industry energy intensity	Baseline and cohort x industry	Omit outliers
	Dependent Variable: Log Energy Efficiency					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Price of Energy x Entrant	-0.096	-0.077	-0.095	-0.091	-0.109	-0.098
	(0.032)	(0.044)	(0.032)	(0.030)	(0.031)	(0.039)
Log Price of Energy	-0.188	-0.198	-0.248	-0.181	-0.189	-0.229
	(0.049)	(0.023)	(0.055)	(0.047)	(0.049)	(0.050)
Entrant	-0.078	-0.023	-0.082	-0.126	-0.147	-0.033
	(0.012)	(0.012)	(0.013)	(0.037)	(0.111)	(0.012)
R^2	0.74	0.74	0.74	0.74	0.74	0.79
Number of Observations	1,250,203	1,250,203	1,250,203	1,250,203	1,250,203	1,121,608

Huber-White standard errors in parentheses. The dependent variable is log energy efficiency, constructed as in Table 1. Log price of energy is the log of the weighted sum of state by year energy source prices from the DOE, using weights constructed from the LRD, divided by the industry by region output price (see equation (2.13) and text). Entrant is a dummy equal to one if the plant appears for the first time in the CM. Log price of energy x entrant is the interaction of the two variables. The sample is constructed as described in the text. All regressions are estimated by Weighted Least Squares, using plant output as weights. The cohort of a plant refers to the year the plant first appears in the Census. All regressions include a full set of cohort dummy variables, equal to one if the plant belongs to the corresponding cohort, and equal to zero the year the plant enters. All columns include a full set of industry-region-year and state dummies, and column 2 also contains industry-state interactions. Column 3 includes interactions of the price of energy with the industry's 3-digit industry. Industry energy intensity is a set of dummy variables for the deciles of the industry's energy cost share in 1963. Column 4 includes the full set of interactions of industry energy intensity with the entry cohort dummies. Column 5 includes the interactions of the plant's entry cohort with its 2-digit industry. Energy cost share is constructed as in Table 2.1, and the 5th and 95th percentiles of the cost share is computed for each industry-region-year-entrant cell. Column 6 omits outlying observations.

Table 2.3

Effect of the Price of Energy on Entry, Survival of Entrants and Exit

	(1)	(2)	(3)	(4)	(5)	(6)
	Log entry	Entry fraction	Log survival of entrants	Survival probability of entrants	Log exit of incumbents	Exit probability of incumbents
Log Price of Energy	-0.109 (0.128)	-0.012 (0.017)	-0.017 (0.141)	-0.066 (0.065)	-0.104 (0.169)	0.014 (0.020)
R^2	0.91	0.63	0.89	0.59	0.88	0.61
Number of Observations	59,643	92,069	42,293	50,865	39,864	72,115

Huber-White standard errors in parentheses. Observations are at the industry-state-year level. The dependent variable is the log of a count of entering plants in column 1; the count of entering plants divided by the number of operating plants in the corresponding cell in 1963 in column 2; the log of the number of entrants who survive to the next Census in column 3; the number of survivors divided by the number of 1967 entrants in column 4; the log of the number of incumbents that exit before the next Census in column 5; and the probability that an incumbent exits before the next Census in column 6. All regressions are estimated by Weighted Least Squares, with total output in the corresponding cell as weights. All regressions include industry-region-year and state dummies. Columns 1 and 2 include observations from 1967-1997, and columns 3-6 include 1967-1992.

Table 2.4

Controls for Plant Productivity

	Log entry	Entry fraction	Log survival of entrants	Survival probability of entrants	Log initial capital stock	Log initial employment
	Dependent Variable: Log Energy Efficiency					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Price of Energy x Entrant	-0.087 (0.033)	-0.095 (0.032)	-0.087 (0.030)	-0.087 (0.029)	-0.104 (0.031)	-0.096 (0.033)
Log Price of Energy	-0.173 (0.054)	-0.190 (0.050)	-0.200 (0.051)	-0.191 (0.051)	-0.162 (0.054)	-0.191 (0.049)
Entrant	-0.085 (0.013)	-0.075 (0.013)	-0.086 (0.013)	-0.086 (0.012)	-0.053 (0.013)	-0.043 (0.013)
Plant Productivity Variable	-0.004 (0.005)	-0.034 (0.027)	-0.017 (0.005)	-0.055 (0.021)	0.046 (0.003)	0.041 (0.003)
R^2	0.73	0.74	0.69	0.70	0.74	0.74
Number of Observations	1,184,514	1,242,622	1,082,928	1,184,568	1,019,478	1,244,384

Huber-White standard errors in parentheses. The dependent variable, log price of energy and entrant variables are computed as in Table 2. The sample is constructed as in Table 2.2, except that column 5 does not include observations from 1967. Regressions are estimated by Weighted Least Squares, using plant output as weights. All regressions include the same control variables as in column 1 of Table 2. Plant Productivity Variable refers to the variable added to the regression that is indicated in the column heading. In columns 1-4 plant productivity is the dependent variable from columns 1-4 of Table 2.3, for the plant's industry, in the year and state it entered. Column 5 includes the log book value of the capital stock the year the plant entered (or the 1972 value for plants that entered before 1972) and column 6 uses the log of plant employment in the first observed year of operation.

Table 2.5

Substitution Elasticities of Entrants and Incumbents

	Energy price elasticity by entry cohort	Baseline specification, with log price of energy x year interactions	Log price of energy x region x year interactions	Log price of energy x industry x year interactions
	Dependent Variable: Log Energy Efficiency			
	(1)	(2)	(3)	(4)
Log Price of Energy x Entrant		-0.095 (0.032)	-0.097 (0.032)	-0.094 (0.032)
Log Price of Energy	-0.293 (0.076)	-0.308 (0.118)	-0.292 (0.119)	-0.299 (0.118)
Entrant		-0.078 (0.012)	-0.073 (0.032)	-0.071 (0.032)
Log Price of Energy x 1967 New Plant	-0.039 (0.034)			
Log Price of Energy x 1972 New Plant	-0.005 (0.047)			
Log Price of Energy x 1977 New Plant	0.004 (0.067)			
Log Price of Energy x 1982 New Plant	0.008 (0.071)			
Log Price of Energy x 1987 New Plant	-0.073 (0.057)			
Log Price of Energy x 1992 New Plant	0.120 (0.074)			
Log Price of Energy x 1997 New Plant	-0.043 (0.083)			
R^2	0.74	0.74	0.74	0.74
Number of Observations	1,250,203	1,250,203	1,250,203	1,250,203

Huber-White standard errors in parentheses. The dependent variable, log price of energy and entrant dummy are computed as in Table 2.2. The sample is the same as column 1 of Table 2.2. Regressions are estimated by Weighted Least Squares, using plant output as weights, and include industry-region-year and state dummies. Columns 2-4 include cohort dummies. Operate is a set of dummy variables, equal to one if the plant operated in the corresponding year. New plant is a set of dummy variables, equal to one if the plant entered that year. Column 1 includes the operate and new plant variables, plus the interaction of all variables with the log price of energy. Column 2 includes a full set of log price of energy by year interactions. Column 3 includes log price of energy by region by year interactions, and column 4 includes log price of energy by 2 digit industry by year interactions.

Table 2.6

Alternative Measures of the Price of Energy

	Log aggregate price	Log aggregate forecasted price	3-year lag	5-year average	NBER prices
			Dependent Variable: Log Energy Efficiency		
	(1)	(2)	(3)	(4)	(5)
Log Price of Energy x Entrant	-0.163 (0.063)		-0.089 (0.032)	-0.097 (0.032)	-0.056 (0.037)
Log Forecasted Price x Entrant		-0.201 (0.055)			
Log Price of Energy			-0.157 (0.046)	-0.327 (0.059)	
Entrant	-0.082 (0.013)	-0.082 (0.013)	-0.078 (0.012)	-0.078 (0.012)	-0.061 (0.012)
R^2	0.74	0.74	0.74	0.74	0.72
Number of Observations	1,250,203	1,250,069	1,248,286	1,250,103	1,046,679

Huber-White standard errors in parentheses. The dependent variable, log price of energy and entrant dummy are computed as in Table 2.2. The sample is constructed as in Table 2.2, except column 5 which does not include 1997. Regressions are weighted by plant output, estimated by Weighted Least Squares, and include the same control variables as in column 1 of Table 2.2. Column 1 uses the log aggregate price of energy, the log of the weighted mean of the individual plant prices. The log aggregate forecasted price is calculated from a VAR with two lags, estimated using the DOE state prices for natural gas, electricity, petroleum distillate and residual. The forecasted price of each energy source is the sum of the current price and the linear prediction of the price over the next ten years, discounted by ten percent. The log aggregate forecasted price of energy is the log of the weighted sum of the energy source forecasts, using aggregate weights (see text). Column 2 includes the interaction of the log forecasted price and the entrant dummy. Column 3 uses the 3-year lag of the log price of energy for the plant's industry and state. Column 4 uses the mean log price of energy for the plant's industry and state over the previous 5 years. Column 5 uses the industry by year log prices of energy and shipments from the NBER Manufacturing Productivity Database.

Table 2.7

Additional Robustness Checks

	Cluster standard errors by year	Cluster standard errors by plant	Dep var is log fuel intensity	Dep var is log electricity intensity	Include dummy if plant is part of a new firm	Omit top quartile of energy intensive industries	Omit electricity producing industries	Separate effect for coal intensive industries	Omit low natural gas counties	Control for CO non-attainment and entrant interaction
	Dependent Variable: Log Efficiency									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Price of Energy x Entrant	-0.096 (0.018)	-0.096 (0.028)	-0.075 (0.043)	-0.107 (0.036)	-0.123 (0.036)	-0.107 (0.039)	-0.092 (0.036)	-0.093 (0.034)	-0.096 (0.034)	-0.084 (0.036)
Log Price of Energy	-0.188 (0.053)	-0.188 (0.052)	-0.472 (0.061)	-0.224 (0.075)	-0.187 (0.049)	-0.165 (0.058)	-0.217 (0.052)	-0.176 (0.049)	-0.182 (0.053)	-0.166 (0.057)
Entrant	-0.078 (0.020)	-0.078 (0.014)	-0.073 (0.016)	-0.026 (0.015)	-0.102 (0.018)	-0.072 (0.015)	-0.086 (0.014)	-0.084 (0.012)	-0.082 (0.013)	-0.087 (0.012)
R^2	0.74	0.74	0.74	0.66	0.74	0.68	0.71	0.74	0.74	0.84
Number of Observations	1,250,203	1,250,203	943,034	1,239,533	1,250,203	957,305	1,122,357	1,205,203	1,119,907	1,007,795

Huber-White standard errors in parentheses. The dependent variable in columns 1, 2 and 5-10 is log energy efficiency, constructed as in Table 2.2. The dependent variable is log fuel efficiency in column 3 and log electricity efficiency in column 4, both in million BTUs per 1972 dollar of output, constructed using reported fuel and electricity expenditure and the real price of fuels and electricity (see text). All specifications include the same controls as in column 1 of Table 2.2, and observations are weighted by plant output. In column 1 standard errors are clustered by year, and by plant in column 2. In column 5 a dummy equal to one if the plant does not belong to a pre-existing firm is added, plus its interaction with the price of energy. Column 6 omits plants belonging to industries in the top quartile of energy cost share, computed from the 1963 Census. Column 7 excludes plants in the four 2-digit industries with the largest share of electricity sales in total output: Paper, Chemicals, Petroleum and Primary Metals (see text). Coal cost share is the ratio of coal expenditure to total energy costs, by industry. Column 8 includes the interactions with the main independent variables of a dummy equal to one if the industry is in the 90th percentile of coal cost share. Column 9 excludes counties with natural gas consumption below the 15th percentile (see text). CO non-attainment is the interaction of a dummy variable equal to one if the plant's county is in non-attainment status with respect to the Clean Air Act for Carbon Monoxide, with a dummy equal to one if the plant's industry emits CO. Column 10 includes the interaction of CO non-attainment with the entrant dummy, plus the main effect.

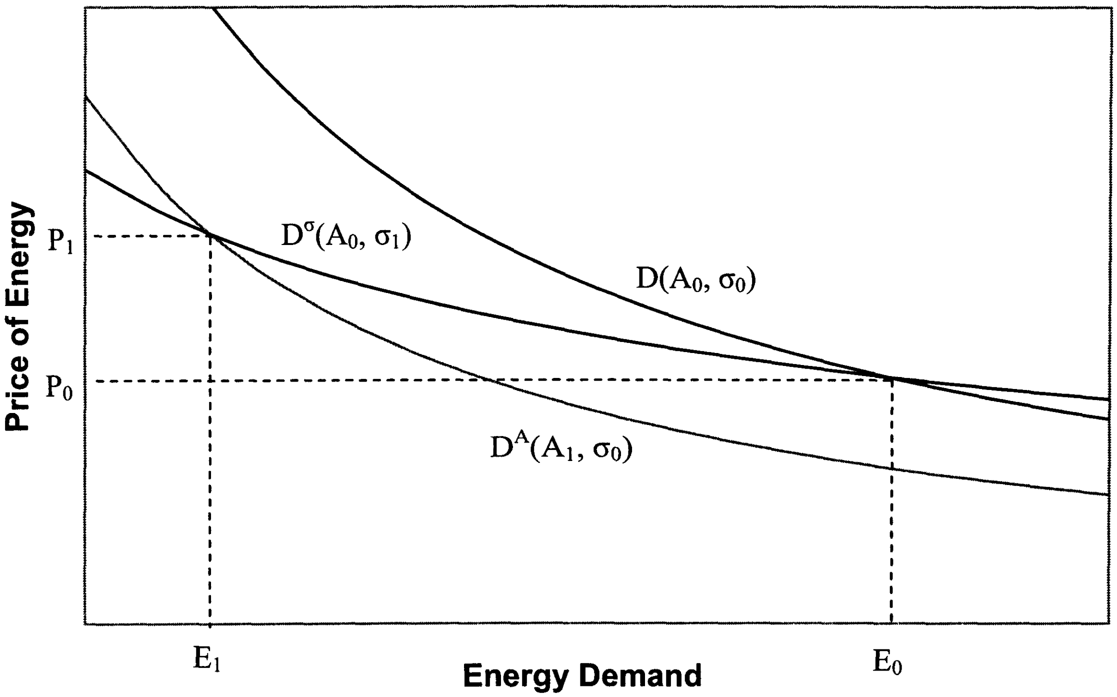
Table 2.8

Energy Efficiency, Investment and Capital Retirements of Existing Plants, Using a Balanced Panel, 1972-1988

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dep var is log energy efficiency, baseline spec	Dep var is log energy efficiency, first difference	Log investment	Log investment with 3-year lag price of energy	Log investment with forecasted price of energy	Log retirements	Log retirements with 3-year lag price of energy	Log retirements with forecasted price of energy
Log Energy Price	-0.192 (0.087)	-0.174 (0.053)	0.012 (0.042)	-0.082 (0.058)		0.411 (0.284)	0.289 (0.301)	
Log Energy Price x Energy Intensity			-0.675 (0.464)	-0.592 (0.501)		-2.314 (2.833)	-2.247 (3.025)	
Log Forecasted Price x Energy Intensity					-0.335 (0.242)			-1.162 (1.648)
R^2	0.80	0.39	0.26	0.26	0.26	0.44	0.44	0.48
Number of Observations	196,605	185,040	161,805	161,805	161,805	146,094	146,094	146,094

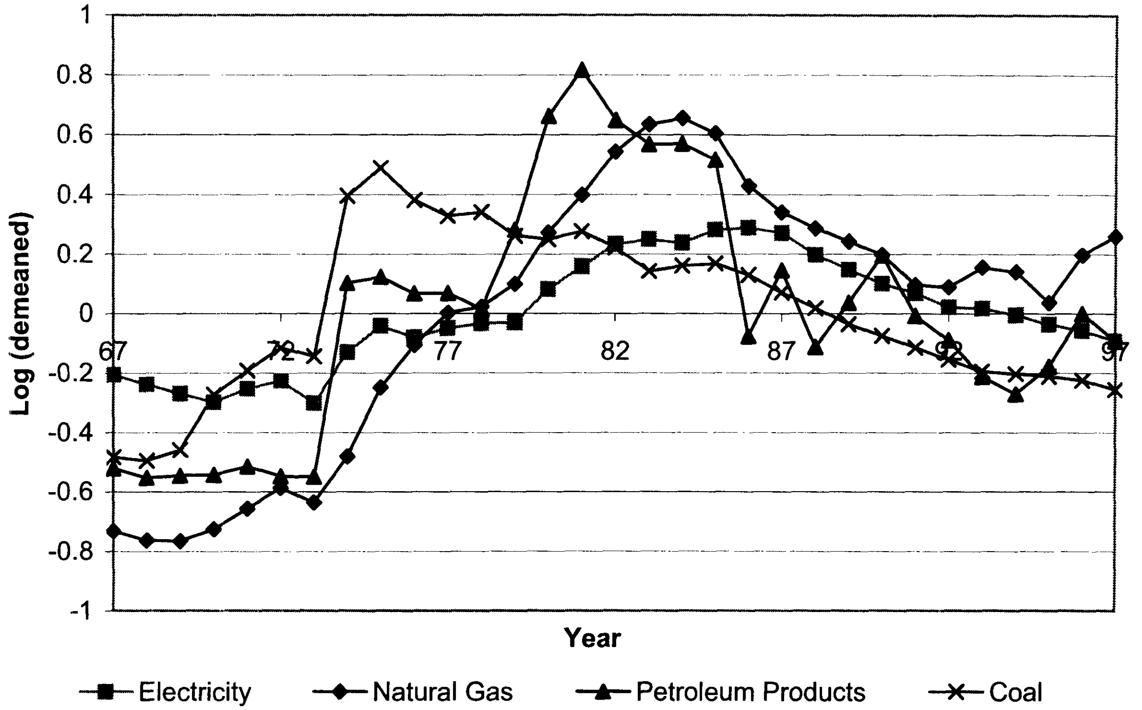
Huber-White standard errors in parentheses. All variables in columns 2-8 are in first differences. All regressions include industry-region-year dummies and are estimated by Weighted Least Squares, with plant output as weights. Column 1 also includes state dummies. The sample includes a balanced panel of plants appearing in all CM and ASM years from 1972-1988. The dependent variables are indicated in the column headings. The dependent variable in columns 1 and 2 is log energy efficiency, constructed as in Table 2.1. Plant capital stocks are constructed using the perpetual inventory method, beginning with the 1972 book value of capital, subtracting retirements and deflating investment, using the BEA depreciation rates and NBER investment deflators (see text). Log investment is defined as the difference in the log capital stock between the current and previous years. The dependent variable in columns 3-5 is log investment, and in columns 6-8 it is log retirements. Columns 3 and 6 use the log current price of energy, columns 4 and 7 use the 3-year lag, and columns 5 and 8 use the log aggregate forecasted price of energy, using a VAR with two lags (see Table 2.6 and the text). Energy intensity is the ratio of total energy costs to output in 1963 for the plant's industry, calculated from the 1963 Census of Manufactures.

Figure 2.1: Substitution Elasticity vs Technological Change



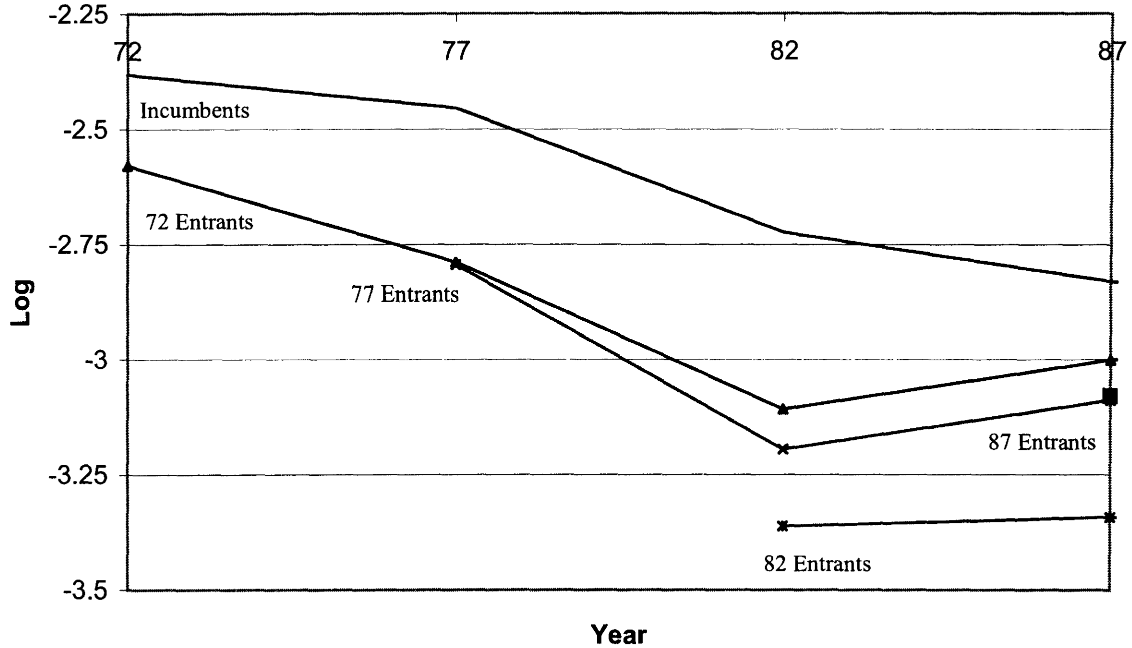
Notes: The curves represent the demand for energy, as functions of the energy technology, A, and the substitution elasticity, σ (see text). The price of energy is P_0 before a price shock, after which it is P_1. The initial energy consumption is E_0 and final consumption is E_1. The curve $D^{\sigma}(A_0, \sigma_1)$ has the same technology and a greater substitution elasticity than $D(A_0, \sigma_0)$. The curve $D^{A}(A_1, \sigma_0)$ has the same substitution elasticity and a larger energy technology than $D(A_0, \sigma_0)$.

Figure 2.2: Real Energy Prices, 1967-1997



Notes: The real price of each energy source is the nominal price, from the State Energy Price Report, 1967-1997, divided by the output price, from the Census of Manufactures, 1967-1997. The numerator is the price in current dollars per million BTUs. The denominator is the output deflator, in current dollars per 1972 dollars. Real prices are computed separately for each plant in the sample, then weighted by plant output (see text). All series have been demeaned.

Figure 2.3: Mean Energy Efficiency of Entrants and Incumbents, 1972-1987



Notes: Incumbents include plants which operated in the previous Census. Each entrant cohort includes plants that entered in the indicated year, still operating in the corresponding year. Each data point is the weighted mean log energy efficiency, in million BTUs per 1972 dollar of output, computed as in Table 2.1, with plant output as weights (see text).

Chapter 3

Technology Adoption and the Effects of Climate Change Policy on Manufacturing

Summary 3 *This chapter studies the response of the manufacturing sector to a carbon tax. I estimate long run price elasticities for fuels and electricity, exploiting the ability of entering plants to choose their technology in response to expected prices. A tax of $10 per metric ton of carbon would reduce emissions by 2 percent and raise operating costs by 8 percent in the short run. Emissions would be 5 percent lower in the long run, and costs would be 5 percent higher. The tax would make plants more vulnerable to subsequent natural gas and distillate oil price shocks, and less sensitive to coal, residual and electricity shocks. Exit would increase by 0.2 percentage points.*

3.1 Introduction

Researchers have devoted considerable effort to estimating the benefits and costs of climate change policies. Recent forecasting models have become increasingly sophisticated. For example, the National Commission on Energy Policy (2004) uses the Department of Energy's National Energy Modeling System to study the effects of tradeable permits system. The academic literature, such as Goulder and Schneider (1999) and Popp (2004), has relaxed many of the assumptions of earlier climate change models, such as Nordhaus (1994).

Despite recent improvements in these models, there is great uncertainty, both on the benefit and cost side. A carbon tax (or permit system) would permanently affect the costs of using fuels and electricity, so the models should include the response of technology to relative costs.[1] The more easily plants can adopt technology that allows them to switch to the less carbon-intensive energy sources, the less costly and more beneficial a policy would be.

Most climate change studies ignore endogenous technological change. They often use estimates of short run elasticities, and assume a constant rate of exogenous technological improvement. This paper focuses on a tax which affects relative energy prices, and documents the importance of accounting for price-induced technology adoption. I compare the change in energy costs and emissions in the short run and the long run, when technology can adjust. As I argue below, the main conclusions are valid for an emissions tax, despite the fact that I cannot directly estimate how much plants would reduce their emissions per unit of energy used.

These calculations rely on estimated parameters of short and long run cost functions. I isolate the effect of technology adoption by assuming that a plant selects its technology when it enters, and cannot change it afterwards.

Figure 3.1 illustrates the empirical strategy. The curve $D^{SR}(A^{NT})$ shows the short run demand curve for coal, as a function of the price of coal. The curve represents the substitution capabilities of a plant that enters with no carbon tax, holding fixed technology, A^{NT}, and other prices. At the initial coal price, P_0, the plant uses C_0 units of coal. The tax increases the price of coal to P_1. By assumption, the plant cannot change its technology. It remains on the initial demand curve, and moves to the point C_1^{NT}.

In contrast, a plant that enters after the tax can select a different technology, A^T. Its short run demand curve is $D^{SR}(A^T)$, and it consumes C_1^T units of coal. Thus, the long run demand curve, D^{LR}, includes the points (P_0, C_0) and (P_1, C_1). There are two effects of the tax on entrants: they select a different short run demand curve from the incumbent, and then move along that curve to the point C_1^T. I refer to the combined effect as the long run response to the tax, because it includes the effect of technology adoption. On the other hand, the incumbent

[1]Either a tax on carbon emissions or a tax on energy would create incentives for technological change. As I argue below, the analysis is similar for either type of policy, and I focus on an energy tax because of data availability.

can only move along its demand curve, and this corresponds to the short run effect.[2]

Given the assumption on technology choice, empirically distinguishing the short run from the long run cost function is straightforward. I estimate the parameters of a translog cost function. This functional form does not impose restrictions on the own- and cross-price elasticities, and it is a second order approximation to an arbitrary, twice differentiable, constant returns to scale cost function.

As is customary with the translog, I recover elasticities from a regression of cost shares on log prices. For each plant, an observation consists of current prices, and the expected prices when the plant entered. Technology can only respond to the latter by assumption. Holding current prices constant, the effect of expected prices on cost shares is equal to the effect of price-induced technology adoption on demand. In contrast, the short run response to energy prices is estimated from the effect of current prices on cost shares, holding initial expected prices fixed.

The results show that technology adoption would greatly reduce the costs, and increase the benefits, of imposing a \$10 carbon tax.[3] Short run costs would increase by about 8 percent, while long run costs would increase by 5 percent. The policy would reduce emissions by about 2 percent in the short run, and by 5 percent in the long run.

I investigate two additional effects of the tax, which are typically ignored. I find that plants adopt technology that requires less coal, residual and electricity, and more natural gas and distillate. These changes would cause plants' total energy costs to be somewhat more sensitive to natural gas and distillate shocks, and less sensitive to other shocks.

Second, the increase in energy costs would increase exit by 0.2 percentage points. This would represent an additional cost of the policy, to the extent that installed capital cannot be resold.

This paper focuses on the response of technology to relative factor prices. The effect of

[2]If the price of another input changes, the incumbents' demand curve would shift by an amount equal to the short run price elasticity. The entrant would have the flexibility to select a different demand curve. Note that the long run and short run cross price elasticities can have the opposite sign.

[3]As discussed below, these calculations assume that the tax is levied on energy producers, and affects the relative prices paid by manufacturing plants. An alternative type of carbon policy would impose a tax on emissions, so that manufacturing plants would pay the tax directly. As I argue below, the changes in emissions and costs for such a policy would be proportional to the first case, so that the conclusions about the importance of technology adoption, and the costs and benefits of the tax are qualitatively similar.

market incentives, such as input prices, on technological change has been studied extensively in the literature. In models of induced innovation and directed technical change, e.g., Binswanger (1974) and Acemoglu (2002), technology responds to expected profits. For a carbon tax, an increase in the price of one input raises the benefit of adopting technology that reduces demand for that input; this spurs both adoption and innovation of that technology. Note that this paper looks at the combined effect of the adoption of previously existing technology and the adoption of price-induced innovations. This paper is a test of one of the necessary conditions of the theoretical literature, that adoption responds to relative prices.

Several recent studies have investigated the effect of energy prices on technology. Linn (2005) estimates the effect of price-induced technology adoption on total energy demand in U.S. manufacturing. This paper uses a similar identification strategy, but focuses on substitution among energy sources, with direct implications for a carbon tax. Several other papers, such as Popp (2001, 2002) and Pizer *et al* (2002), look at the effect of technology on total energy demand, but there has been little recent work looking at individual energy inputs.

Closely related to this paper is Doms (1993). Using plant data from the 1985 Manufacturing Energy Consumption Survey, Doms documents the relationship between plant size, fuel prices and fuel use, finding that many plants use only one fuel, and some have the ability to switch costlessly between fuels (most often, between natural gas and oil). These patterns mean that it is important to use a flexible functional form for the cost function, one that allows arbitrary cross price elasticities, and can incorporate the fact that many plants use only a subset of the available fuels.

A number of papers, including Jorgenson and Wilcoxen (1992), Goulder and Mathai (2000), and Popp (2004) have investigated the effect of a carbon tax on carbon emissions and economic growth. These authors simulate long run growth models under various policy scenarios, estimating the costs and benefits of the alternatives. They do not allow for the endogenous response of technology to relative fuel and electricity prices.

The paper is organized as follows. The next section presents a model illustrating the effect of relative energy prices on technology adoption, and highlights the short and long run effects of a permanent price increase. Section 3.3 discusses the strategy for estimating the cost functions. Section 3.4 describes the data, Section 3.5 discusses the results, and Section 3.6 concludes.

3.2 Theoretical Motivation

3.2.1 Baseline Model

I present a model of technology adoption. Plants choose their technology in response to expected energy prices when they enter. In the initial steady state, entering and previously existing plants use the same technology. I then consider a tax on energy, which affects fuel and electricity prices in proportion to their emissions rates, and causes entrants to select a different technology. I calculate the reduction in carbon emissions and increases in energy costs due to the tax, for entering and existing plants.

Plants operate in a single industry with no uncertainty, and take prices as given. I assume that energy is separable from other inputs, such as labor and intermediate materials, so that a plant chooses its technology and fuel demands to minimize the cost of obtaining a given quantity of energy. This separability assumption is strong, but I use it to focus on energy demand and technology, and relax it in the empirical work.

More specifically, each period there is a set of potential entrants which determine whether it would be profitable to enter, pay entry costs if they decide to enter, then begin operating. I assume that a fraction, ϕ, of plants that operated in the previous period exogenously fail. A potential entrant solves the following cost minimization problem (I omit plant subscripts):

$$\min_{\{\mathbf{X}_t\}_{t=0}^{\infty};\mathbf{A}} \sum_{t=0}^{\infty} \frac{(\mathbf{X}_t \cdot \mathbf{P}_t)}{(1+r+\phi)^t} + G(\mathbf{A}) \tag{3.1}$$

$$s.t. E = f(\mathbf{X}_t, \mathbf{A}),$$

where $\mathbf{X}_t$ is the vector of energy inputs at time t: coal, natural gas, distillate fuel oil, residual and electricity. The plant purchases these inputs each period it operates. $\mathbf{P}_t$ is the vector of energy prices, so that the cost at time t is the inner product of $\mathbf{X}_t$ and $\mathbf{P}_t$. Future costs are discounted by the interest rate, r, plus the failure rate, ϕ. The plant minimizes costs subject to the constraint that it obtain E BTUs of energy, according to the function $f(\mathbf{X}_t, \mathbf{A})$. $\mathbf{A}$ is a vector of factor augmenting technology parameters, which the plant selects at the time of entry. For example, X_c units of coal provide $A_c X_c$ units of "coal services." The plant pays an entry cost of $G(\mathbf{A})$, where G is increasing in each argument. Thus, the plant can reduce its demand of input i by selecting a higher A_i, but must pay a higher entry cost. I will return to the interpretation of $\mathbf{A}$ below.

After entering, the plant takes $\mathbf{A}$ as given, and minimizes costs each period, subject to the constraint above. I define the cost function, $c(\mathbf{P}_t, \mathbf{A}, E)$, which is the dual of the production function, $f(\mathbf{X}_t, \mathbf{A})$. I approximate the cost function with a translog function, and rewrite the minimization problem as:

$$\min_{A} \sum_{t=0}^{\infty} \frac{c_t(\mathbf{P}_t, \mathbf{A}, E)}{(1+r+\phi)^t} + G(\mathbf{A}) \tag{3.2}$$

$$s.t. \ln c_t = \ln E + \sum_j \alpha_j \ln(P_{jt}/A_j) + 1/2 \sum_j \sum_k \beta_{jk} \ln(P_{jt}/A_j) \ln(P_{kt}/A_k).$$

I impose the homogeneity and symmetry restrictions, that $\sum_j \alpha_j = 1$, $\sum_j \beta_{jk} = 0$, and $\beta_{jk} = \beta_{kj}$. Under these assumptions, the translog is the second order approximation to an arbitrary, twice differentiable, constant returns to scale cost function. The α's are the first order coefficients in the approximation, and correspond to the average cost shares; a higher α_c means the plant uses more coal. The β's are the second order coefficients, and as discussed below, they are related to the substitution elasticities. A higher β_{cc} means that the plant reduces its coal demand by less in response to a coal price increase. I normalize prices such that costs are directly proportional to E.

As is customary in working with the translog function, I focus on the input share equations. For simplicity, I assume that energy prices are constant at $\mathbf{P}_0$. The cost share of input j, s_{j0}, is obtained by differentiating the cost function with respect to P_j. Assuming that the plant is a price-taker, $\partial c_t/\partial P_{jt}$ equals the demand for input j. Thus, the jth cost share is:

$$s_{j0}^I = \alpha_j - \sum_k \beta_{jk} \ln(A_k^I) + \sum_k \beta_{jk} \ln(P_{k0}), \tag{3.3}$$

where the superscript I denotes that the plant is an incumbent. Note that the intercept depends on the technology, which the plant selects at the time of entry. The role of the β's is clarified by the short run own- and cross-price elasticities, ε_{jj}^I and ε_{jk}^I:

$$\varepsilon_{jj}^I = \beta_{jj}/s_{j0} + s_{j0} - 1 \tag{3.4}$$

$$\varepsilon_{jk}^I = \beta_{jk}/s_{j0} + s_{k0} \tag{3.5}$$

A smaller β_{jj} means that the own-price elasticity is larger in magnitude. The condition $\beta_{jk} > 0$ is a sufficient condition that inputs j and k are substitutes.

In comparison to equation (3.3), the cost share for an entrant is:

$$s_{j0}^E = \alpha_j - \sum_k \beta_{jk} \ln(A_k^E) + \sum_k \beta_{jk} \ln(P_{k0}), \tag{3.6}$$

where the superscript E denotes an entrant. Equations (3.3)-(3.6) show that if entering and existing plants have the same technologies, as is true in the steady state with constant prices, they have the same cost shares and substitution elasticities.

I now consider the choice of technology parameters. I assume the following functional form for $G(\mathbf{A})$:

$$G(\mathbf{A}) = \frac{1+r+\phi}{r+\phi} \sum_k \frac{1}{\rho_k} A_k^{\rho_k}, \tag{3.7}$$

where k is an index over coal, natural gas, distillate, residual and electricity. I assume that $\rho_k > 0$, so the marginal cost of A_k is positive; it is increasing if $\rho_k > 1$.

The first order condition for A_k from the minimization problem (3.2) can be rearranged to obtain:

$$\ln(A_k) = \frac{2}{2+\rho_k} \ln(P_{k0}), \tag{3.8}$$

Equation (3.8) is a central results, and shows that technology responds to expected prices at the time at entry. A one percent increase in the expected price of input j would increase the optimal A_k by $\frac{2}{2+\rho_k}$ percent. I define the parameter $a_k \equiv \frac{2}{2+\rho_k}$, to simplify the expressions below.

I substitute equation (3.8) into equations (3.3) and (3.6), to express the costs shares for entrants and incumbents, in terms of initial expected and current prices:

$$s_{j0}^{I} = \alpha_j - \sum_k a_k \beta_{jk} \ln(P_{k0}^{I}) + \sum_k \beta_{jk} \ln(P_{k0}), \tag{3.9}$$

$$s_{j0}^{E} = \alpha_j - \sum_k a_k \beta_{jk} \ln(P_{k0}^{E}) + \sum_k \beta_{jk} \ln(P_{k0}), \tag{3.10}$$

where P_{k0}^{I} and P_{k0}^{E} are the expected prices at the time of entry for incumbents and entrants.

From equation (3.8), an entrant and an incumbent choose the same technology in the steady state. By equations (3.9) and (3.10), and the expressions for price elasticities, the two plants have the same demands and elasticities.

These equations show the short and long run effect of energy prices on cost shares. In the short run, an increase in the current price causes plants to substitute, with technology fixed. By comparison, an entering plant chooses its technology in response to initial expected prices. Because of the negative sign before the first summation in equation (3.10), technological change affects the cost share in the opposite direction as the short run response.

I use equation (3.8) to rewrite the cost function for an entrant as:

$$\begin{aligned} \ln c_{t0}^{E} &= \ln E + \sum_j \alpha_j \ln(P_{k0}) + 1/2 \sum_j \sum_k \beta_{jk} \ln(P_{j0}) \ln(P_{k0}) \qquad (3.11) \\ &\quad - \sum_j a_j \alpha_j \ln(P_{k0}^{E}) - \sum_j \sum_k a_j \beta_{jk} \ln(P_{j0}^{E}) \ln(P_{k0}) + 1/2 \sum_j \sum_k a_j a_k \beta_{jk} \ln(P_{j0}^{E}) \ln(P_{k0}^{E}), \end{aligned}$$

The cost function for incumbents is analogous, with P_{j0}^{I} substituted for P_{j0}^{E}. The summations in the first line show the short run effect of prices on costs, i.e., holding technology constant. The second line corresponds to the additional effect of technology adoption on costs. In the initial steady state, expected and current prices are the same, so entrants and incumbents have the same costs.

I use equations (3.9), (3.10) and (3.11) to calculate the long run own- and cross-price elasticities:

$$\varepsilon_{jj}^{E} = (1 - a_j)(\beta_{jj}/s_{j0} + s_{j0}) - 1 \tag{3.12}$$

$$\varepsilon_{jk}^{E} = (1 - a_k)\varepsilon_{jk}. \tag{3.13}$$

The condition $\beta_{jj} > 0$ is sufficient for the long run own-price elasticity to be larger (in magnitude) than the short run. The cost share equation explains this relationship, because if P_j increases and $\beta_{jj} > 0$, the cost share would rise in the short run. Technology adoption would cause the cost share to rise by less, in the long run, meaning that demand would decline by more than in the short run. From equation (3.13), the long run cross-price elasticity is smaller than the short run elasticity, since $a_k < 1$ (i.e., technology responds less than proportionately to expected prices).

I now consider the effect of imposing a carbon tax, beginning from the steady state. I assume that the tax is levied on energy producers, and raises prices in proportion to the carbon emissions from burning the energy sources.[4] I focus on four outcomes: costs, input demands, emissions and the sensitivity to additional price shocks.

Coal is the most carbon intensive fuel, so the tax would have the greatest effect on coal prices. Natural gas is the least coal-intensive, and oil is intermediary. I assume that the composition of fuels used to generate electricity does not change, so the effect of the tax on electricity prices is the weighed sum of the effect on fuel prices. Prices rise from $\mathbf{P}_0$ to $\mathbf{P}_1$.

The first result is that entrants' costs increase by less than incumbents. From equation (3.8), entrants select a larger $\mathbf{A}$, so $G(\mathbf{A})$ is greater. Since entrants could select the same $\mathbf{A}$ and have energy costs equal to those of incumbents, the fact that they select a larger $\mathbf{A}$ means that their energy costs must be lower. In other words, they pay higher entry costs to reduce their variable costs.

From equation (3.9), the cost share of a plant that entered before the tax, s_j^I, is given by:

$$s_{j1}^{I} = \alpha_j - \sum_k a_k \beta_{jk} \ln(P_{k0}^{I}) + \sum_k \beta_{jk} \ln(P_{k1}), \tag{3.14}$$

[4] In the following section I model a tax on emissions, where plants may respond by reducing emissions rates, in addition to the responses studied here.

where P_{j1} is the price of input j after the policy. Since the incumbent has already chosen its technology, the first summation in equation (3.14) is unaffected by the tax. The second summation depends on the new prices, so s^I_{j1} changes by a weighted sum of the price changes, with weights β_{jk}.

The jth cost share of a plant that enters after the tax is:

$$s^E_{j1} = \alpha_j - \sum_k a_k \beta_{jk} \ln(P^E_{k1}) + \sum_k \beta_{jk} \ln(P_{k1}). \tag{3.15}$$

Comparing equations (3.14) and (3.15), entrants choose a different technology from incumbents. Since $0 < a_k < 1$, technology adoption partially offsets the effect of the price shock on cost share j. For example, suppose the price of coal increases and other prices remain constant. If $\beta_{cc} > 0$, then the the coal cost share increases more for incumbents than for entrants, because entrants choose a technology with greater coal efficiency (a higher A_c). Thus, the choice of A_c represents an inward shift (and rotation) of the demand curve in response to the tax. As the cost function in the minimization problem shows, raising A_c corresponds to decreasing the effective price, lowering energy costs.

I now use the cost function parameters to calculate the effect of the policy on the outcomes of interest. After the price shock, a plant's demand for input j is, $X_{j1} = c_1 \cdot s_{j1}/p_{j1}$, where c_1 is given by equation (3.11), using the appropriate prices, and s_{jt} is given by equation (3.14) or (3.15), depending on whether the plant is an entrant or incumbent. The percent change in demand for entrants and incumbents, ΔX_j, relative to the demand in the absence of the policy, is:

$$\Delta X^I_j = \frac{c^I_1 \cdot s^I_{j1} \cdot p_{j0}}{c_0 \cdot s_{j0} \cdot p_{j1}} - 1, \tag{3.16}$$

$$\Delta X^E_j = \frac{c^E_1 \cdot s^E_{j1} \cdot p_{j0}}{c_0 \cdot s_{j0} \cdot p_{j1}} - 1. \tag{3.17}$$

From equations (3.12) and (3.13), for small price changes, if the short run demand change is positive, then the long run change is smaller, or negative. If the short run change is negative, the long run change is also negative, but may be smaller or larger in magnitude.

The effect of the policy on carbon emissions is the weighted sum of the changes in fuel and electricity demands, with weights equal to the rate of emissions per unit of input. The percent change in emissions, ΔC, is given by:

$$\Delta C = \sum_j \overline{C}_j \Delta X_{j.} \tag{3.18}$$

where $\overline{C}_j$ is the CO_2 emissions per BTU of input j. The uncertainty about relative short run and long run input changes means that it is also indeterminate whether emissions fall more in the long run.

Finally, I consider the effect of the policy on the sensitivity to additional energy price shocks. That is, I am interested in how much the tax affects the response of energy costs to a subsequent shock. I compare two types of incumbents: plants that entered with no carbon policy, and plants entering with a carbon policy. Recall that $\frac{d \ln c}{d \ln p_j} = s_j$, so if the tax causes plants to have a higher cost share for input j, they become more sensitive to subsequent shocks to that input.

Consider a 10 percent increase in the price of input j. The difference in the cost elasticities is:

$$s_j^T - s_j^{NT} = \sum_k (1 - a_k)\beta_{jk} \ln(P_{k1}/P_{k0}), \tag{3.19}$$

where the superscripts T and NT denote plants that entered with and without the tax, respectively. As this equation shows, the difference in cost shares is a weighted average of the percentage increases in prices. This difference can be positive or negative, and by the homogeneity restrictions, they cannot all have the same sign. This implies that the policy makes some shocks more costly and others less costly.

To summarize, the policy has different effects on entrants and incumbents, because incumbents cannot change their technology. As the preceding discussion has shown, the effect of the policy on costs, emissions and the sensitivity to subsequent shocks can be calculated directly from the cost function parameters. Costs increase by less in the long run, but it is theoretically uncertain whether the long run emissions changes are smaller or larger in magnitude than the short run.

3.2.2 Emissions Tax

I consider a tax on carbon emissions, which does not directly affect energy prices. Plants pay an amount proportional to how much carbon they emit. If they cannot change their emissions rates (i.e., C_j), the effects of the tax can be calculated directly from the estimated cost function parameters from the previous section.

To see this, consider a similar model, where a potential entrant solves the following maximization problem:

$$\max_{\{\mathbf{X}_t\}_{t=0}^{\infty};\mathbf{A}} \sum_{t=0}^{\infty} \frac{E(\mathbf{X}_t, \mathbf{A}) - \mathbf{X}_t \cdot \mathbf{P}_t - T\mathbf{X}_t \cdot \overline{\mathbf{C}}}{(1+r+\phi)^t} - G(\mathbf{A}) \tag{3.20}$$

$$\begin{aligned} s.t. E(\mathbf{X}_t, \mathbf{A}) &= \sum_j \alpha_j \ln(X_{jt}A_j) + 1/2 \sum_j \sum_k \beta_{jk} \ln(X_{jt}A_j) \ln(X_{kt}A_k) \\ G(\mathbf{A}) &= \frac{\sum_k \frac{1}{\rho_k} A_k^{\rho_k}}{r+\phi}. \end{aligned}$$

There are several differences from the previous model. First, the plant maximizes the difference between its energy, $E(\mathbf{X}_t, \mathbf{A})$, and its total costs. Energy is produced from the energy inputs via a translog production function, with the same parameters α_j and β_{jk} as the cost function in the previous section (by the duality of the cost function). Total costs are the sum of operating costs and entry costs. Operating costs are the sum of the expenditure on energy inputs, $\mathbf{X}_t \cdot \mathbf{P}_t$; and the amount of the carbon tax, $T\mathbf{X}_t \cdot \overline{\mathbf{C}}$, where T is the amount of the tax, in dollars per million metric tons of CO_2 equivalent, and $\overline{\mathbf{C}}$ is the vector of emissions rates, in million metric tons of CO_2 equivalent per BTU. Entry costs are equal to the energy technology costs, $G(\mathbf{A})$, as before. As in the previous model, existing and entering plants can respond to the tax by changing their energy input demands, and entrants can choose a different energy technology.

Consider a small increase in the carbon tax, beginning from the steady state, with $T = 0$. The short run change in emissions is given by $\frac{\partial C^I}{\partial T} = \sum_j \overline{C}_j (\frac{\partial X_j}{\partial T_j})^I$. That is, the change in emissions is the weighted sum of the change in energy demands, holding energy technology fixed. After rearranging the first order conditions for X_j, I obtain the following expression for the short run change in emissions:

$$\frac{\partial C^I}{\partial T} = \sum_j \sum_k \frac{\overline{C}_j \overline{C}_k X_j}{P_k} \varepsilon^I_{jk} \tag{3.21}$$

The long run change is given by a similar expression, substituting the long run price elasticities. As this equation shows, the short run and long run changes in emissions can be approximated using the price elasticities and initial conditions. The underlying argument is that reducing X_j decreases costs by the amount $P_j + T\overline{C}_j$. Since the emissions rate is fixed, the effect of increasing T is thus proportional to the effect of increasing P_j. The analysis for the changes in costs yields the same conclusion.

I now consider a more general model, where plants can adjust the emissions rates after entering.[5] I assume an explicit functional form for the cost of reducing emissions rates, which allows me to express the changes in emissions and costs in terms of the initial emissions rates, energy demands, and price elasticities. A potential entrant's maximization problem is given by:

$$\max_{\{\mathbf{X}_t\}_{t=0}^{\infty};\{\mathbf{C}_t\}_{t=0}^{\infty};\mathbf{A}} \sum_{t=0}^{\infty} \frac{E(\mathbf{X}_t, \mathbf{A}) - \mathbf{X}_t \cdot \mathbf{P}_t - T\mathbf{X}_t \cdot \mathbf{C}_t - \mathbf{H}(\mathbf{C}_t)}{(1+r+\phi)^t} - G(\mathbf{A}) \tag{3.22}$$

$$\begin{aligned} s.t. E(\mathbf{X}_t, \mathbf{A}) &= \sum_j \alpha_j \ln(X_{jt}A_j) + 1/2 \sum_j \sum_k \beta_{jk} \ln(X_{jt}A_j) \ln(X_{kt}A_k) \\ H(\mathbf{C}_t) &= \sum_k \frac{1}{2}(\overline{C}_k - C_{kt})^2 \\ G(\mathbf{A}) &= \frac{\sum_k \frac{1}{\rho_k} A_k^{\rho_k}}{r+\phi}. \end{aligned}$$

This is the same problem as in the model just discussed, except that the plant can select its emissions rates each period. The cost of choosing $\mathbf{C}_t$, $H(\mathbf{C}_t)$, is separable in the emissions rates. The parameter $\overline{C}_k$ is the baseline emission rate, which is the value of emissions per BTU, if the plant does not undertake any further emissions reductions. The first order condition for C_{kt} is:

$$TX_{kt} = \overline{C}_k - C_{kt} \tag{3.23}$$

[5] By comparison, in the model just considered, the costs of adjusting the emissions rates are infinite.

As this equation shows, C_{kt} is decreasing in the carbon tax and the demand for input k. In the initial steady state, with $T = 0$, all plants select $C_k = \overline{C}_k$.

Again, I focus on a small increase in the carbon tax, beginning from the initial steady state. Existing plants adjust their input demands and emissions rates in response to the tax; entering plants can also adjust their technology. The short run change in emissions, $\frac{\partial C^I}{\partial T}$, is equal to $\sum_j \overline{C}_j(\frac{\partial X_j}{\partial T})^I + X_j(\frac{\partial C_j}{\partial T})^I$. From equation (3.23), and using the fact that $T = 0$ in the initial steady state, this expression is equal to $\sum_j \overline{C}_j(\frac{\partial X_j}{\partial T})^I - X_j^2$. Thus, the change in emissions can be calculated from the change in input demands $(\frac{\partial X_j}{\partial T})$ and the initial emissions rates and demands ($\overline{C}_j$ and X_j); an analogous expression holds for the long run change. After manipulating the first order conditions for input demands, it is possible to show that the short run change in emissions is given by:

$$\frac{\partial C^I}{\partial T} = \sum_i \sum_j \frac{\overline{C}_j \overline{C}_k X_j}{P_k} \varepsilon_{jk}^I - X_i^2 \tag{3.24}$$

The long run change in emissions is given by a similar expression, substituting the long run elasticities. Thus, to first order, I can calculate the effect of an emissions tax using the cost function parameters from the previous section. The difference between the short run and long run changes in emissions is a weighted sum of the difference between the short run and long run price elasticities. Similar conclusions hold for the short and long run effects of the tax on costs, where costs are the sum of input costs and tax payments.

Note that the separability of cost function $H(\mathbf{C})$ and the quadratic functional form are important for deriving these results. Under these assumptions, there is a linear relationship between emissions rates and input demands. Since the price elasticities measure how much the input demands change, they also capture how much the emissions rates change. This allows me to calculate the effect on costs and emissions from imposing an emissions tax.[6]

[6] For other functional forms, it is generally necessary to estimate the parameters of $H(\mathbf{C})$ to calculate the effects of the policy. In addition, these calculations are local approximations, and it is unclear how accurate they are for large changes in the emissions tax.

3.3 Empirical Strategy

I now discuss the estimation of the cost function. Beginning with equation (3.14), I add industry-region-year interactions and an error term. The cost share of input j for plant i in year t is:

$$s_{ijt} = \alpha_j - \sum_k \beta^0_{jk} \ln(P_{ik0}) + \sum_k \beta_{jk} \ln(P_{kt}) + \gamma_{lry} + \varepsilon_{ijt}, \tag{3.25}$$

where β^0_{jk} is the effect of the expected price of input k on cost share j; P_{ik0} is the expected price of input k when the plant entered; P_{kt} is the price of input k in year t; and γ_{lry} denotes a full set of industry-region-year interactions, for industry l, region r, and year t.

In this specification, I generalize the functional form for the cost of selecting a technology (i.e., $G(A)$), and allow the cost of increasing the technology for one input to depend on the level of technology for the other inputs. In other words, the marginal cost of increasing A_k is a function of A_k and all other A_js. Consequently, the optimal A_k depends on the expected prices of all inputs. This means that it is possible for the long run cross-price elasticities to be larger than the short run, or have opposite signs.[7]

The change in the cost share due to a change in the current price of input k, holding the initial prices constant, is β_{jk}. The response of the cost share to the expectation of price k at the time of entry is $-\beta^0_{jk}$. Consequently, a permanent increase in P_k would cause an incumbent's cost share to increase by β_{jk}, and an entrant's cost share to increase by $\beta_{jk} - \beta^0_{jk}$.

The empirical strategy relies on the translog specification, and the distinction between entering and existing plants. As shown in equation (3.25), the translog cost function, in addition to imposing few restrictions on the data, provides a set of linear equations for estimating the parameters.

The interpretation of these parameters arises from the assumption that plants choose their technology when they enter. After entering, the plant's technology is fixed, so the response of the cost shares to current prices, holding initial prices fixed, corresponds to the short run elasticities.[8]

[7] This specification does not identify the parameters in $G(\mathbf{A})$, but as the previous section shows, it is not necessary to recover them to calculate the effects of the policy.

[8] Because the expected price is constant for a given plant, I cannot include plant fixed effects in the estimation. Consequently, the short run parameter is identified by both within- and cross-plant variation. Estimating the

Technology can only respond to initial expected prices. The long run elasticity is identified by comparing two plants with the same current prices, and different expected prices. The correlation between their cost shares and the initial prices is the additional effect of price-induced technology adoption on input demands.

As is common with the translog cost function, I estimate the parameters by stacking the cost share equations, dropping one equation (distillate) and imposing the symmetry and homogeneity restrictions, both within and across equations. I jointly estimate the cost function, given by equation (3.11), plus industry-region-year interactions and an error term:

$$\begin{aligned}\ln c_t^E &= \sum_j \alpha_j \ln(P_{ik0}) + 1/2 \sum_j \sum_k \beta_{jk} \ln(P_{ij0}) \ln(P_{ik0}) - \sum_j \alpha_j^0 \ln(P_{ik0}^E) \\ &\quad - \sum_j \sum_k \beta_{jk}^0 \ln(P_{ij0}^E) \ln(P_{ik0}) + 1/2 \sum_j \sum_k \beta_{jk}^{00} \ln(P_{ij0}^E) \ln(P_{ik0}^E) + \gamma_{lry} + \varepsilon_{ijt},\end{aligned} \tag{3.26}$$

The parameters α_j^0 and β_{jk}^{00} are the coefficients on the expected prices and the interactions of the expected prices, respectively. The cross equation restrictions require the use of a Seemingly Unrelated Regression (SUR) model. The next section describes the data sources and construction of the variables. Note that the initial expected price is the price when the plant enters.

The empirical strategy incorporates several simplifications. First, I assume energy prices are exogenous; omitted variables correlated with the prices would bias the estimates. For example, there may be reverse causality, where technological change or industry-specific shocks affect energy prices. However, the fuel and electricity prices are aggregate prices by state and year, so they do not include the effects of industry-specific shocks to energy demand. As discussed in Linn (2005), much of the price variation during this period was due to government regulation and other factors exogenous to manufacturing activity. In addition, the industry-region-year interactions control for industry by region shocks, reducing the threat of reverse causality arising from industry-specific or aggregate changes in energy demand.

Another possible omitted variable is plant heterogeneity; e.g., some plants may have greater

model in first differences yields very similar short run elasticities (shown below), so cross-sectional heterogeneity does not appear to bias the results.

total factor productivity than others. If more productive plants are more likely to enter and survive, sample selection could bias the results. This does not appear to be a major concern, however; removing plant fixed-effects by taking first differences yields quite similar short-run price elasticities to those obtained with the baseline specification.[9]

Another concern is that I proxy for initial expected prices with prices at the time of entry. If prices follow random walk processes, then this is valid to first order; mean reversion would imply that the estimates would be biased towards zero (since an entrant would only respond to permanent price changes). That is, the differences between the long run and short run responses would be biased towards zero. As Linn (2005) argues, the random walk assumption appears reasonable for aggregate energy prices, so a useful check would be to perform the analogous exercise for individual energy sources.

Despite the relative flexibility of the translog cost function, it imposes several restrictions. First, the cost function is only a second order approximation. If the second order terms, β_{jk}, vary with energy prices, and current prices are correlated with expected prices, this could lead to spurious results.

Second, the estimates of β_{jk} are identified by the response of the cost shares to current prices, holding expected prices constant. It is possible that incumbents can adjust their technology, in which case β_{jk} would include the effect of this technological change. I would overestimate the short run effect of prices on cost shares, causing me to understate the difference between the short run and long run effects of the policy. However, this should not affect the long run estimates.

The third potential limitation of the translog specification is that I impose the homogeneity and symmetry restrictions. These are necessary for the translog to be a local second order approximation to an arbitrary cost function. The baseline model is also homothetic; relaxing this assumption does not affect the results.

The model includes the assumption that other input prices do not affect fuel and electricity demands, i.e., that energy is separable from other inputs. Below I discuss the results from

[9] As noted above, I cannot estimate long-run elasticities when taking first differences, because the initial energy prices are constant over the life of the plant. The similarity of the first difference results appears to be due to the industry-region-year interactions, which absorb much of the cross-plant heterogeneity; estimating equation (3.25) with only industry-year interactions yields elasticities quite different from the corresponding first difference estimates.

estimating a more general cost function, which includes other variable inputs (workers and materials). The estimates are similar, but I do not emphasize them because the other prices may be endogenous.

Finally, most plants do not use every fuel (all plants in the sample use electricity). This could lead to biased estimates of β_{jk}, because the effect of a price change may be nonlinear. The industry-region-year interactions partially address this concern. From equation (3.25), I estimate a different cost share for each industry-region-year cell. Consequently, if all plants in a given industry, region and year do not use a given fuel, those observations are not included in estimating the other coefficients.

There are many cells, however, in which some plants use a given fuel and others do not, so the interactions may not completely solve the problem.[10] This issue is a possible direction for future research.

3.4 Data

3.4.1 Variable Construction

I construct total costs, cost shares and real energy prices to estimate equations (3.25) and (3.26). The data is from the Census of Manufactures (CM), 1972-1982; the Annual Survey of Manufactures (ASM), which includes a fuel supplement, 1975-1981; and the Department of Energy's State Energy Price Report (SEPR), 1975-1981. I calculate the dependent variables directly from the ASM: coal, natural gas, distillate fuel oil, residual and electricity expenditure shares, and total energy costs. The sample includes all ASM plants with positive electricity expenditure, which entered after 1975. There are 61,599 observations for 22,061 plants.

The independent variables, the real fuel and electricity prices, are constructed from the SEPR and CM. The SEPR contains energy prices, in current dollars per million BTUs, for each state and year. I deflate the nominal prices by shipments deflators constructed from the CM.

[10] The cost calculations reflect the assumptions of the model, that price changes have a linear effect on costs. For example, the increase in the price of coal caused by the policy raises costs by the same amount, whether or not a plant in the sample initially uses coal. In some cases this causes negative predicted cost shares after the policy is enacted; this issue requires further work.

These prices vary by 4-digit industry, state and year, and are aggregated from 7-digit by state product prices, as described in Linn (2005).

Table 3.1 presents summary statistics for these variables. Panel A shows the cost shares for the five energy sources, unweighted and weighted by total energy costs. Electricity is the predominant source of energy, followed by natural gas, oil and coal. The weighted means show that plants with greater energy requirements use more fuels and less electricity. In parentheses are the standard deviations of residuals obtained by removing industry-region-year means from the reported cost shares. The residuals correspond to the dependent variables used below, since the regressions include industry-region-year interactions. There is considerable variation within industry-region-year cells for all energy sources.

The next row in Panel A shows the fraction of plants that use each fuel. All plants in the sample use electricity, but there is considerable heterogeneity in the choice of fuels. The final row in Panel A shows the distribution of plants with zero expenditure for each fuel. The fraction is the number of non-users located in industry-region-year cells in which no plants use the fuel, divided by the total number of non-users. That is, if the fraction were equal to one, then there would be no cells containing both users and non-users. The fact that most of the fractions are significantly less than one suggests that cost shares may respond nonlinearly to prices, as discussed above.

Panel B shows average fuel and electricity prices. Standard deviations, after removing industry-region-year means, are in parentheses. Each plant's prices are weighted by total energy costs. Over this period, coal prices gradually decline, while the other prices are steady through 1979, then rise sharply. The industry-region-year means remove much of the price variation, but the remaining variation is substantial. In particular, the price changes I use to calculate the effects of a carbon tax are well within the sample range for all inputs, except coal.

3.4.2 Effect of a Carbon Tax on Prices

I calculate the effect of a carbon tax on energy prices. I use a tax of \$10 per metric ton of CO_2 equivalent, which is similar to recent policy proposals. I assume that the effect of the tax on a given energy price is proportional to the emission rate of the energy source from combustion, in metric tons of CO_2 equivalent per million BTUs. For electricity I use the average emissions in the U.S., which depends on the fuel composition electric utilities use to generate electricity. I calculate the percentage changes caused by the tax, from the 2002 average U.S. prices of the sources. I calculate a 59 percent increase in the price of coal, a 9.2 percent increase for natural gas, a 10.7 percent increase for distillate and residual, and a 5.7 percent increase for electricity.

Note that there are several simplifying assumptions in these calculations. First, the policy affects the energy prices paid by plants in the manufacturing sector. A tax levied on emissions from manufacturing could have a different effect. As I argue in section 3.2.2, if emissions rates are fixed, or the cost or reducing emissions rates increases quadratically, the effects of an emissions tax could be approximated using the cost function parameters. Thus, the conclusion that technology adoption significantly reduces costs and emissions for an energy tax is likely to hold for an emissions tax.

Second, I assume that the fuel mix of electricity generators does not respond to the tax. This paper considers the partial equilibrium response to a carbon tax; it would be necessary to account for technological change in electricity generation to determine the overall effect of the tax.

Finally, the calculations only include the effect of a tax on burning fossil fuels. Many policy proposals include emissions from other sources, such as industrial processes which generate greenhouse gases. These emissions comprise less than 5 percent of total manufacturing emissions (EPA, 2005); accounting for this channel would not greatly affect the results.

3.5 Results

I discuss the costs and benefits of implementing a carbon tax. The estimated own-price elasticities are larger in the long run than the short run. The long run increase in costs is about 60 percent of the short run change. Emissions decline considerably more in the long run than in the short run.

Plants become more sensitive to subsequent natural gas or distillate shocks, and less sensitive to coal, residual and electricity shocks.[11] The increase in costs would induce some plants to

[11] This effect could be a cost or benefit of the policy, depending on the relative volatilities of the different prices, and the response of technology to expected volatility. Assessing whether this is a positive or negative result of the policy would require the estimation of higher-order parameters.

exit, implying an additional cost of the tax, if plants cannot resell some portion of their capital stock.

3.5.1 Estimation of Cost Function Parameters and Price Elasticities

Table 3.2 shows the estimated price elasticities for the baseline specification. Equations (3.25) and (3.26) are estimated by a SUR model, imposing all within- and cross-equation restrictions. Observations are weighted by total energy costs, to account for measurement error of small plants. The Appendix Table shows the estimated short run coefficients and regression statistics.

I calculate short and long run elasticities numerically.[12] I use the expression of the demand for input j, $X_j = c \cdot s_j/p_j$, and the prices and quantities for the year each plant enters. I use the appropriate equations for costs and cost shares to calculate the percentage changes in demands for a one percent change in prices. I calculate elasticities for each plant, then weight by the plant's share in consumption for the corresponding input (in BTUs). Thus, the elasticity corresponds to the total change in demand of entering plants in the sample, if prices were one percent higher than they actually were. Panel A of Table 3.2 shows the short run elasticities, the effect of a one percent price increase after the plants have entered. The long run elasticities are in Panel B, the percent changes in demand if prices had increased by one percent before the plants entered.

The own-price elasticities are significant and negative in both the short and long run. The short run estimate for distillate, -1.02 (standard error 0.26) is considerably larger than the other energy sources, while electricity demand is relatively inelastic. The short run cross-price elasticities are generally quite small, and often insignificant; holding technology fixed, substitution capabilities are fairly limited. To gain some intuition for the effect of the carbon tax, the increase in the price of coal would have little effect on the demand for other inputs, except for natural gas, which increases somewhat.[13]

[12] I do not use the equations for the short and long run price elasticities in section 2 because I do not estimate a_j. The numerical short run elasticities are nearly identical to the estimates obtained using equaitons (3.4) and (3.5).

[13] It is perhaps surprising that natural gas demand is not more sensitive to residual or distillate prices. This could be due to the fact that the natural gas shortage limited the ability to substitute away from oil during the oil shock; future work with data from the 1980s and 1990s should resolve this issue. Also note that natural gas and petroleum products are slightly stronger substitutes in the long run.

Figure 3.2 shows an analogous picture to Figure 3.1, using the estimated cost functions. I plot the short run demand curve of incumbents, $D^{SR}(A^{NT})$, by normalizing to one all prices except for coal. I allow the current price of coal to vary, and calculate coal demand, holding all other current, and all initial prices constant. Under this price normalization, when $P_c = 1$, coal demand is equal to its average cost share. The curve $D^{SR}(A^T)$ represents the short run substitution capabilities of a plant that entered after the policy, with the initial coal price equal to 1.59, and all other initial prices equal to one.[14] To plot the curve I keep all current prices constant except for coal. Finally, D^{LR} represents the long run relationship between the price of coal and coal demand, holding all other prices equal to one. Note that this curve intersects $D^{SR}(A^{NT})$ at $P_c = 1$ and $D^{SR}(A^T)$ at $P_c = 1.59$.

This figure shows the distinction between the short and long run response to the price of coal. Holding technology fixed, plants reduce consumption by about 16 percent. The long run reduction is about 40 percent larger than the short run. Note that this does not contradict the similarity of the long run and short run own-price elasticities for coal in Table 3.2. The latter correspond to the change in demand caused by a small change in the price, suggesting that the first order approximation is not very accurate for the large price changes caused by the tax. Consequently, the price elasticities are used mainly to compare estimated cost function parameters using alternative specifications.

Returning to Table 3.2, with the exception of coal, the own-price elasticities are significantly greater in the long run than the short run. Some of the long run cross-price elasticities are significantly larger than the corresponding short run elasticities. The increase in the price of coal caused by the tax would induce substitution towards natural gas and distillate, and away from residual and electricity in the long run.[15]

[14] I hold prices other than coal equal to one for simplicity, to focus on the short and long run responses to coal prices. By affecting other prices, the policy would also cause the short run curve to shift.

[15] As noted above, in the model in section 3.2, long run and short run cross-price elasticities are proportional. In the empirical model, however, the long run elasticities can be smaller or larger than the short run, or have opposite signs.

3.5.2 Effect of a Carbon Tax on Energy Demand, Energy Costs and Emissions

In Table 3.3 I calculate the effect of a carbon tax on energy demand, energy costs, emissions, and the sensitivity to subsequent price shocks. For each outcome, I calculate the short and long run effect of implementing the policy, again using prices and quantities for plants in the year they enter. The short run effect corresponds to implementing the policy the same year the plants enter, but after they have chosen their technology. That is, for each plant I compare the actual value of the outcome with the predicted value, assuming initial expected prices (P_{ik0}) are equal to the observed prices. Current prices (P_{kt}) are equal to the observed prices, multiplied by the effect of the policy on prices (see the bottom of Panel A). For the long run effect, P_{ik0} is the observed initial price, multiplied by the effect of the policy.

Panel A shows the response of fuel and electricity demand to the tax, in the short and long run, using equations (3.16) and (3.17). Coal, residual and electricity fall in the short and long run. The long run declines are significantly larger than the short run changes, reflecting the larger own-price and negative coal cross-price elasticities in Table 3.2. Natural gas rises slightly more in the long run, and distillate falls in the short run but rises in the long run.

Panel B shows the effect of the policy on energy costs and emissions. The short run effect on costs, about 8 percent, is considerably larger than the long run effect. From equation (3.18) the percent change in emissions is the weighted sum of percent changes of each fuel. The weights are the emissions rates for each fuel, obtained from the Department of Energy website. That is, I assume that the tax does not affect emissions per BTU of energy consumed. The long run decline in emissions is larger than the short run. Thus, the results in Panel B show that it is important to account for endogenous technological change in analyzing climate change policies.

3.5.3 Effect of an Energy Price Shock Following Policy Implementation

Panel C in Table 3.3 shows the effect of the policy on plants' sensitivity to additional energy price shocks. As discussed in the previous subsection, the tax causes substitution away from coal, residual and electricity, and towards natural gas and distillate. Consequently, I expect that the costs of plants entering after policy implementation should be more sensitive to natural gas or distillate price shocks, than without the policy. To study this possibility, I begin with the predicted long run costs of plants in the ASM sample (i.e., the long run costs in Panel B, where both P_{ik} and P_{ik0} include the effect of the policy). Each column in Panel C shows the effect of increasing the corresponding price by an additional 10 percent, after plants have entered. Columns 3 and 4, for example, represent an oil shock. For comparison, I calculate the effect of a 10 percent price shock on costs, in the absence of the policy.

In general, the imprecise estimates prevent strong conclusions. However, the calculations follow the expected pattern. Coal, residual and electricity price shocks are slightly less costly under the policy, while natural gas and distillate shocks are more costly. This is because the shift towards natural gas and distillate increases the sensitivity to those price changes. As noted above, I cannot determine whether these effects constitute a cost or benefit of the policy. It would be necessary to estimate the variances of the prices (i.e., it could represent a cost if natural gas is more volatile than coal), and the response of technology to expected variance.

3.5.4 Robustness

I now address several potential criticisms of the baseline results. First, it is possible that the estimates of equations (3.25) and (3.26) are biased by omitted variables. For example, energy prices may be correlated with unobserved plant productivity, i.e., total factor productivity (TFP).

Alternatively, plants may demand a fuel for different end-uses, such as natural gas, which can be used as a boiler fuel or for process heating. Suppose that the elasticities for these two end-uses differ, and plants decide whether to use natural gas for these purposes, based on relative energy prices. Then the estimated own-price elasticity for natural gas would include the effect of prices on the end-use decisions, in addition to the flexibility of a plant with a given set of end-uses.

In Panel A of Table 3.4, I address these two issues, estimating equation (3.25) in first differences. This removes unobserved fixed effects, including TFP. Furthermore, assuming that technology cannot be changed after entry, this approach estimates the effect of prices on energy demand, holding fixed plants' end-uses. In other words, these estimates correspond to a weighted average of the elasticities for different end-uses.[16] In general, the results are quite

[16] Since the energy prices at the time of entry are fixed over the life of the plant, the long run coefficients are

similar to the baseline. Several of the elasticities are somewhat smaller in Table 3.4, such as the distillate own-price estimate, but the discrepancies are not large.

As Table 3.1 shows, energy demand varies according to total energy requirements. For example, plants using more energy tend to consume more coal. This suggests that the homotheticity assumption may not be valid, and that elasticities may vary with plant size. I relax this assumption in Panel B, by including the log of total energy consumption (in BTUs) on the right hand side in equation (3.25). Note that I do not jointly estimate the cost function. This is a commonly used technique in working with the translog cost function, but the inference is obviously complicated by the endogeneity of total energy demand. Nevertheless, the short run elasticities are quite similar to those in Table 3.2, suggesting that the homotheticity assumption is not a major concern (for comparison, Panel A of Table 3.5 shows the short run elasticities from estimating equation (3.25), without costs on the right hand side).

Finally, the empirical analysis has assumed that energy is separable from other inputs. This allows me to estimate the cost function parameters without including in the regressions other input prices, which may be endogenous. I relax this assumption, and estimate a cost function that includes production workers, non-production workers, materials, and the same five energy inputs.[17] The dependent variables are the expenditure of each input in total variable input costs. The independent variables are the same five energy prices, plus state-industry-year log real prices of production workers, non-production workers, and materials.[18] The analogous cost share equations to (3.25), and the corresponding cost function, are estimated by a SUR estimator, imposing all homogeneity and symmetry restrictions. The short run price elasticities for the energy inputs are shown in Panel B of Table 3.5. The estimates are generally similar to

not identified. I can only compare the short run elasticities from this model with those shown in Table 3.2.

[17] The specification reported in Table 3.5 does not include the price of capital as a control variable, so it contains the assumption that variable input demands are separable from capital. Recall, however, that the regressions include industry-region-year interactions. Thus, variation in investment prices at this aggregation level or higher are absorbed. More detailed prices are unavailable, preventing further progress.

[18] The nominal production worker wage by state, industry and year is the average wage, in dollars per hour, of plants in the corresponding cell. Non-production worker hours are unreported, so I assume that they work the same number of annual hours as production workers. The nominal materials price by state, year and industry is computed from the CM materials files, 1963-1997 (the ASM does not contain detailed materials information). The price is the weighted average of the 6-digit materials prices of plants in the corresponding industry and state. The weights are the reported share in materials expenditure of the products, reported in the CM materials files. These prices are extrapolated linearly between the Census years to obtain prices in the intervening years. All nominal prices are deflated by the same output price as the energy prices.

those in Table 3.2. Several of the own-price elasticities are significantly larger, which suggests that there is some ability to substitute from energy to other inputs. The substitution patterns are nearly identical to the baseline results, and in most cases the estimates are statistically indistinguishable. The corresponding calculations for Table 3.3 are also fairly similar, which suggests that the separability assumption does not have a large effect on the results.

3.5.5 Effect of a Carbon Tax on Plant Exit

I now discuss the final implication of a carbon tax, that the increase in energy costs may cause plants to exit. To the extent that capital cannot be resold, this would imply an additional negative effect of the policy.

The strategy for calculating the effect of the policy on exit is straightforward. To determine the change in exit, I multiply the predicted short run increase in energy costs, 8 percent, by the elasticity of exit with respect to energy costs. I estimate the elasticity from a regression where the independent variable is the probability a plant survives, and the dependent variable is log energy costs. I expect the coefficient to be negative, since higher energy costs should make a plant less likely to survive.

More specifically, the estimation sample includes all plants in the Census of Manufactures that enter after 1963, for the years 1967-1992. The dependent variable is an indicator, equal to one if the plant survives to the next Census (five years later).

The independent variable is the log of predicted total energy costs. Substitution and technological change affect energy costs, so I cannot use actual costs as an independent variable. Instead, I calculate predicted costs using the cost function parameters from equation (3.26), and the plant's current and initial energy prices. In other words, the independent variable is a weighted sum of current prices, initial prices, and the interactions. The weights are the estimated cost function parameters, and do not change over time; consequently, the predicted costs do not include the effect of technological change or substitution due to energy prices.

Table 3.6 shows the estimates from an OLS regression of survival on predicted costs, including plant age by year, and industry-region-year interactions.[19] The estimate in column 1

[19] The binary dependent variable suggests the use of Maximum Likelihood estimators (e.g. Probit), but they cannot deal with the large numer of parameters and fixed effects. This remains an unresolved issue.

implies that a one percent increase in predicted costs reduces survival (increases exit) by 0.02 percentage points, or by about 800 plants, given that there were about 400,000 plants in the manufacturing sector in 1997. Assuming that capital cannot be resold when the plants exit, this effect would imply an additional cost of the policy of about \$2.2 billion, which is about 35 percent of the estimated short run increase in energy costs.[20] This calculation suggests that the costs due to exit are substantial.

In column 2 the dependent variable is a dummy equal to one if the plant survives for 10 years. The estimated elasticity is quite similar to that in column 1. Columns 3-6 include additional controls: age-industry-year interactions and age-industry-region-year interactions. The estimates are fairly stable across specifications.

3.6 Conclusions

This paper measures the effects on the manufacturing sector of implementing a carbon tax, which alters relative energy prices. I estimate a short and long run cost function for the sector, by assuming that plants choose their technology when they enter, and cannot change it afterwards. Aplant selects its technology in response to expected prices at entry.

I use the cost function estimates to calculate the effect of the tax on energy demands, emissions, energy costs and the sensitivity to subsequent price shocks. I estimate the effect of costs on exit.

The main implication of this paper is that it is essential to incorporate the effects of price-induced technological change when assessing climate change policies. I have argued that the cost function estimates can be used to calculate the response to an emissions tax, so the conclusion about the importance of technological change is likely to hold for an emissions tax, as well. Since most climate change policies would cover the entire economy, a useful direction of future research would be to analyze technological change in other sectors of the economy.

[20] I assume that production is constant returns to scale, and calculate the value of capital services as the difference between revenue and variable costs (labor, energy and materials). I use data from the Manufacturing Productivity Database, for the year 1996. I discount future capital services by 10 percent, then divide by the number of plants in the 1997 Census of Manufactures, to obtain a value of capital of about \$2.8 million per plant.

3.7 References

Acemoglu, Daron, 2002. Directed Technological Change. *Review of Economic Studies*, 69, 781-809.

Binswanger, Hans, 1974. A Microeconomic Approach to Induced Innovation. *Economic Journal*, 84, 940-958.

Doms, Mark E., 1993. Inter Fuel Substitution And Energy Technology Heterogeneity In U.S. Manufacturing. *Center For Economic Studies Working Paper* 93-5.

Goulder, Lawrence H. and Koshy Mathai, 2000. Optimal CO_2 Abatement in the Presence of Induced Technological Change. *Journal of Environmental Economics and Management*, 39, 1-38.

Goulder, Lawrence H. and Stephen H. Schneider, 1999. Induced Technological Change and the Attractiveness of CO_2 Abatement Policies. *Resource and Energy Economics*, 21, 211-53.

Jorgenson, Dale W. and Peter J. Wilcoxen, 1992. Reducing U.S. Carbon Dioxide Emissions: the Cost of Different Goals. In *Advances in the Economics of Energy and Natural Resources*, vol. 7 John R. Maroney, ed. IAI Press: Greenwich, CT

Linn. Joshua, 2005. Energy Prices and the Adoption of Energy-Saving Technology.

National Commission on Energy Policy, 2004. Ending the Energy Stalemate.

Nordhaus, William D., 1994. *Managing the Global Commons: the Economics of the Greenhouse Effect.* MIT Press: Cambridge, MA.

Pizer, William A., Winston Harrington, Raymond J. Kopp, Richard D. Morgenstern and Jhih-Shyang Shih, 2002. Technology Adoption and Aggregate Energy Efficiency. *Resources for the Future Discussion Paper*, 02-52.

Popp, David, 2001. The Effect of New Technology on Energy Consumption. *Resource and Energy Economics*, 23, 215-39.

Popp, David, 2002. Induced Innovation and Energy Prices. *American Economic Review*, 92, 160-80.

Popp, David, 2004. ENTICE: Endogenous Technological Change in the DICE Model of Global Warming. *Journal of Environmental Economics and Management*, 48, 742-68.

Table 3.1

Summary Statistics

	Panel A: Plant Characteristics				
	Coal	Natural Gas	Distillate Fuel Oil	Residual	Electricity
Cost Share	0.011 (0.062)	0.197 (0.194)	0.054 (0.135)	0.034 (0.109)	0.704 (0.223)
Weighted Cost Share	0.035 (0.151)	0.257 (0.173)	0.051 (0.088)	0.087 (0.135)	0.570 (0.276)
Fraction Plants Using Input	0.030 (0.148)	0.614 (0.427)	0.214 (0.372)	0.112 (0.280)	1.000
Fraction Non-Users in Empty Cells	0.960	0.602	0.589	0.755	

	Panel B: Log Energy Prices				
Year	Coal	Natural Gas	Distillate Fuel Oil	Residual	Electricity
1975	0.117 (0.156)	-0.391 (0.124)	-0.237 (0.059)	0.072 (0.067)	-0.078 (0.174)
1976	0.035 (0.127)	-0.272 (0.123)	-0.282 (0.066)	-0.195 (0.072)	-0.128 (0.187)
1977	0.008 (0.110)	-0.167 (0.119)	-0.265 (0.075)	-0.152 (0.078)	-0.126 (0.205)
1978	0.029 (0.102)	-0.110 (0.158)	-0.281 (0.084)	-0.250 (0.078)	-0.078 (0.195)
1979	-0.019 (0.101)	-0.009 (0.144)	-0.039 (0.084)	-0.078 (0.094)	-0.053 (0.189)
1980	-0.083 (0.115)	0.186 (0.132)	0.225 (0.090)	0.158 (0.097)	0.081 (0.206)
1981	0.065 (0.115)	0.337 (0.128)	0.410 (0.100)	0.350 (0.105)	0.208 (0.196)

Notes: Standard deviation of industry-region-year residuals in parentheses (see text). Figures in Panel A are calculated from the Annual Survey of Manufactures, 1975-1981. Weighted cost shares use total energy costs as weights. Fraction of plants with positive expenditure is the number of plants with positive expenditure for the corresponding energy source, divided by the total number of plants in the sample. Fraction non-users in empty cells is the fraction of plants that do not use the input, located in industry-region-year cells in which zero plants use the fuel. Real prices are computed from the Census of Manufactures and Deparetment of Energy (1972-1982, see text), weighted by total energy costs.

Table 3.2

Own- and Cross-Price Elasticities

Panel A: Short Run

	Price				
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.45 (0.19)	0.29 (0.14)	-0.02 (0.05)	0.00 (0.11)	0.18 (0.14)
Natural Gas	0.20 (0.08)	-0.43 (0.17)	0.06 (0.03)	0.05 (0.05)	0.11 (0.14)
Distillate	-0.09 (0.13)	0.32 (0.16)	-1.02 (0.26)	0.42 (0.18)	0.37 (0.16)
Residual	-0.01 (0.12)	0.16 (0.14)	0.16 (0.08)	-0.55 (0.20)	0.24 (0.13)
Electricity	0.00 (0.12)	0.05 (0.14)	0.03 (0.08)	0.03 (0.14)	-0.11 (0.05)

Panel B: Long Run

	Price				
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.49 (0.16)	0.39 (0.16)	0.06 (0.05)	-0.19 (0.10)	0.04 (0.12)
Natural Gas	0.30 (0.15)	-1.01 (0.29)	0.09 (0.07)	0.21 (0.11)	0.40 (0.17)
Distillate	0.18 (0.14)	0.46 (0.18)	-1.79 (0.37)	1.04 (0.24)	0.12 (0.15)
Residual	-0.27 (0.12)	0.39 (0.17)	0.39 (0.11)	-1.00 (0.29)	0.23 (0.16)
Electricity	-0.10 (0.12)	0.27 (0.16)	-0.02 (0.08)	0.16 (0.12)	-0.31 (0.17)

Notes: Short run elasticities are calculated numerically, as described in the text. Coefficients are obtained by estimating equations (3.25) and (3.26) using a Seemingly Unrelated Regression (SUR) estimator, with all homogeneity and symmetry restrictions imposed (see text). The sample is the same as in Panel A of Table 3.1. Elasticities are calculated using each plant's cost shares, and weighting by the quantity used of the corresponding input. Standard errors in parentheses, calculated using the delta method.

Table 3.3

Effect of a $10 Carbon Tax

	Panel A: Energy Demand				
	Coal	Natural Gas	Distillate	Residual	Electricity
Short Run	-20.37 (8.41)	5.75 (2.99)	-9.44 (3.12)	-5.33 (2.45)	-1.16 (2.15)
Long Run	-25.20 (10.13)	8.18 (4.69)	4.81 (2.07)	-31.44 (6.11)	-5.16 (2.61)
Percent Price Increase Due To Tax	59	7	10	10	6
	Panel B: Costs and Emissions				
	Costs	Emissions			
Short Run	7.88 (3.21)	-2.30 (2.05)			
Long Run	4.92 (1.93)	-5.33 (2.55)			
	Panel C: Percentage Increase in Costs due to a 10% Price Increase				
	Coal	Natural Gas	Distillate	Residual	Electricity
With Tax	0.19 (0.22)	3.28 (2.02)	0.77 (0.42)	0.90 (0.42)	5.11 (2.57)
Without Tax	0.24 (0.30)	1.73 (0.99)	0.23 (0.31)	1.83 (0.61)	6.87 (2.65)

Notes: Percent changes in energy demands are calculated using equations (3.16) and (3.17). The estimates are the weighted averages over plants in the sample, using observations of plants in their entry year. Short run elasticities are calculated by adding the price changes shown at the bottom of Panel A to actual current prices. Long run elasticities are calculated by adding the changes to actual current and expected prices. Weights are the plant's average share of BTUs in the sample, with and without the policy. Percent increase due to tax is calculated from 2002 energy prices in the State Energy Price Report, and emissions per BTU from the Department of Energy, using a tax of 10 dollars per metric ton of carbon dioxide equivalent (see text). Costs and emissions are calculated from equations (3.11) and (3.18), using the same prices as in Panel A. Observations are weighted by the plant's share in energy costs, and emisions per BTU for each input, obtained from the Department of Energy. Observations are weighted by predicted costs. Estimates in Panel C are the percentage increase in costs due to a 10 percent increase in the price of the corresponding input, calculated from equation (3.11).

Table 3.4

Own- and Cross-Price Elasticities, Robustness Checks

Panel A: First Differences

	Price				
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.41	0.15	0.01	-0.01	0.26
	(0.19)	(0.12)	(0.05)	(0.11)	(0.15)
Natural Gas	0.07	-0.62	0.07	0.01	0.17
	(0.13)	(0.16)	(0.07)	(0.09)	(0.15)
Distillate	0.02	0.37	-0.80	0.06	0.36
	(0.13)	(0.17)	(0.23)	(0.15)	(0.16)
Residual	-0.02	0.10	0.02	-0.34	0.25
	(0.12)	(0.14)	(0.06)	(0.16)	(0.13)
Electricity	0.06	0.10	0.02	0.03	-0.22
	(0.13)	(0.14)	(0.08)	(0.11)	(0.16)

Panel B: Relax Homotheticity

	Price				
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.44	0.28	-0.01	0.01	0.16
	(0.20)	(0.14)	(0.05)	(0.11)	(0.13)
Natural Gas	0.20	-0.43	0.07	0.03	0.13
	(0.13)	(0.18)	(0.07)	(0.09)	(0.14)
Distillate	-0.06	0.35	-1.06	0.42	0.34
	(0.13)	(0.17)	(0.27)	(0.17)	(0.16)
Residual	0.00	0.12	0.16	-0.54	0.26
	(0.12)	(0.14)	(0.08)	(0.10)	(0.13)
Electricity	-0.02	0.07	0.02	0.04	-0.12
	(0.12)	(0.14)	(0.08)	(0.11)	(0.12)

Notes: Elasticities and standard errors are calculated as in Table 2. In Panel A, equations (3.25) and (3.26) are estimated in first differences. In Panel B, the log of total energy demand, in BTUs is added to each equation in (3.25). The model is estimated without the cost function in Panel B.

Table 3.5

Additional Robustness Results

Panel A: Omit Cost Function

			Price		
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.45	0.29	-0.01	-0.01	0.19
	(0.20)	(0.14)	(0.05)	(0.11)	(0.14)
Natural Gas	0.20	-0.45	0.07	0.03	0.15
	(0.14)	(0.18)	(0.07)	(0.10)	(0.14)
Distillate	-0.08	0.37	-0.99	0.37	0.32
	(0.13)	(0.17)	(0.26)	(0.17)	(0.15)
Residual	-0.02	0.12	0.14	-0.54	0.30
	(0.12)	(0.14)	(0.07)	(0.20)	(0.14)
Electricity	0.00	0.08	0.02	0.06	-0.16
	(0.12)	(0.14)	(0.08)	(0.11)	(0.08)

Panel B: Include Other Inputs

			Price		
Energy Source	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	-0.53	0.15	-0.04	-0.10	0.04
	(0.09)	(0.04)	(0.01)	(0.02)	(0.05)
Natural Gas	0.18	-0.83	0.07	0.01	0.00
	(0.04)	(0.05)	(0.02)	(0.02)	(0.03)
Distillate	-0.17	0.23	-1.37	0.41	0.03
	(0.03)	(0.04)	(0.09)	(0.06)	(0.03)
Residual	-0.13	0.02	0.13	-0.74	-0.05
	(0.03)	(0.03)	(0.02)	(0.11)	(0.03)
Electricity	0.00	0.00	0.00	-0.05	-0.35
	(0.03)	(0.01)	(0.02)	(0.06)	(0.07)

Notes: Elasticities and standard errors are calculated as in Table 3.2. In Panel A, the share equations in equation (3.25) are estimated (see text). Total operating costs are equal to energy costs plus labor and materials costs. The dependent variables used in panel B are the shares of expenditure in total operating costs of production workers, non-production workers, materials, coal, natural gas, residual and electricity. The independent variables are the log prices of these inputs, and the price of distillate. The operating cost share equations, plus the corresponding cost function, are estimated by a SUR estimator, imposing all cross- and within-equation symmetry and homogeneity restrictions.

Table 3.6

Effect of Predicted Energy Costs on Plant Survival, 1967-1992

	Dependent Variable: Survival Indicator					
	(1)	(2)	(3)	(4)	(5)	(6)
Predicted Energy Costs	-0.019 (0.002)	-0.024 (0.003)	-0.012 (0.002)	-0.015 (0.003)	-0.018 (0.003)	-0.032 (0.003)
R^2	0.07	0.08	0.06	0.06	0.07	0.08
Number of Observations	879,238	625,916	879,238	625,916	625,916	625,916
5-year Survival	Yes	No	Yes	No	Yes	No
10-year Survival	No	Yes	Yes	Yes	Yes	Yes
Interactions Included	Age x Year, Ind x Year x Reg	Age x Year, Ind x Year x Reg	Age x Ind x Year	Age x Ind x Year	Age x Ind x Year x Reg	Age x Ind x Year x Reg

Notes: Standard errors in parentheses, clustered by industry-region-year cell. The sample in columns 1, 3 and 5 includes all plants in the Census of Manufactures, 1967-1992, which entered after 1963, and are located in industries and states covered by the 1975 Annual Survey of Manufactures fuel supplement. The sample in columns 2, 4 and 6 is constructed similarly, but ends in 1987. The dependent variable in columns 1, 3 and 5 is an indicator equal to one if the plant survives to the next Census. The dependent variable in columns 2, 4 and 6 is an indicator equal to one if the plant survives for the next two Censuses. Predicted energy costs for each plant is calculated using the cost function estimates from the regression reported in Table 3.2, and the plant's current and initial energy prices (see text).

Appendix Table

Estimated Cost Function Parameters

Panel A: First Order Terms (α's)

	Coal	Natural Gas	Distillate	Residual	Electricity
	0.099 (0.001)	0.319 (0.002)	0.034 (0.001)	0.094 (0.001)	0.454 (0.004)

Panel B: Second Order Terms (β's)

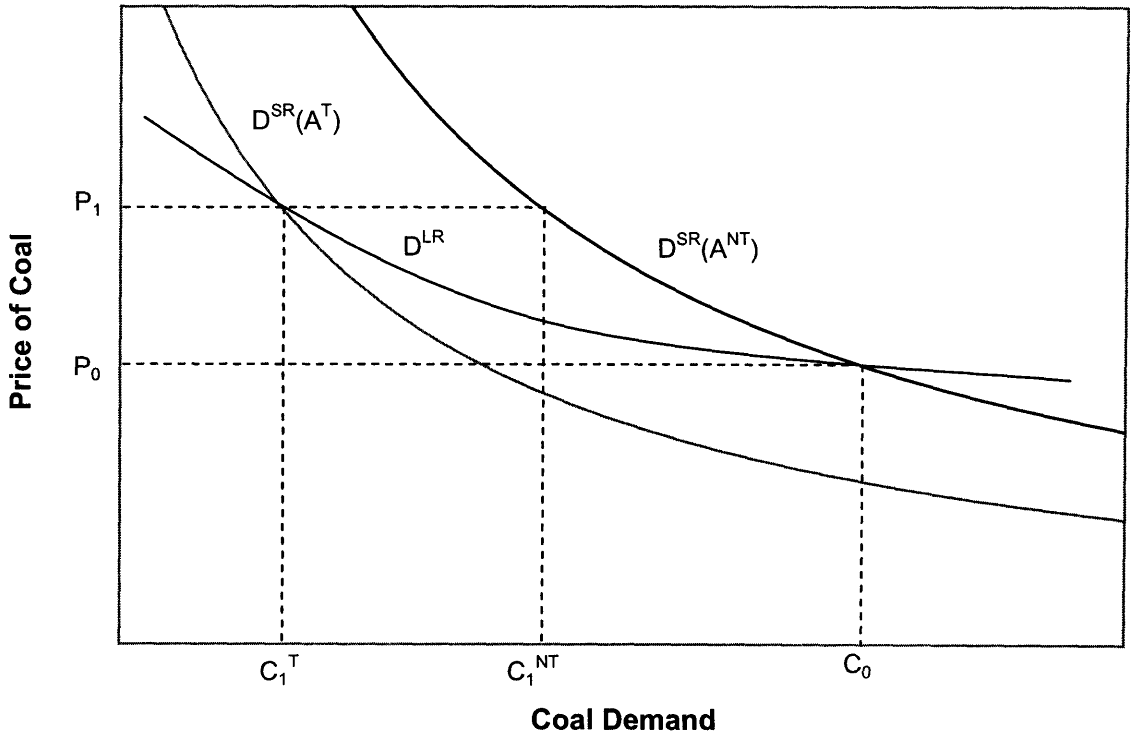
	Coal	Natural Gas	Distillate	Residual	Electricity
Coal	0.005 (0.001)	-0.010 (0.002)	-0.016 (0.002)	0.026 (0.002)	-0.005 (0.001)
Natural Gas		0.135 (0.010)	-0.008 (0.002)	-0.051 (0.004)	-0.066 (0.002)
Distillate			-0.014 (0.003)	0.070 (0.009)	-0.032 (0.004)
Residual				-0.036 (0.004)	-0.009 (0.002)
Electricity					0.112 (0.009)

Panel C: Regression Statistics

Number of Observations	61,599
Monotonicity	0.95

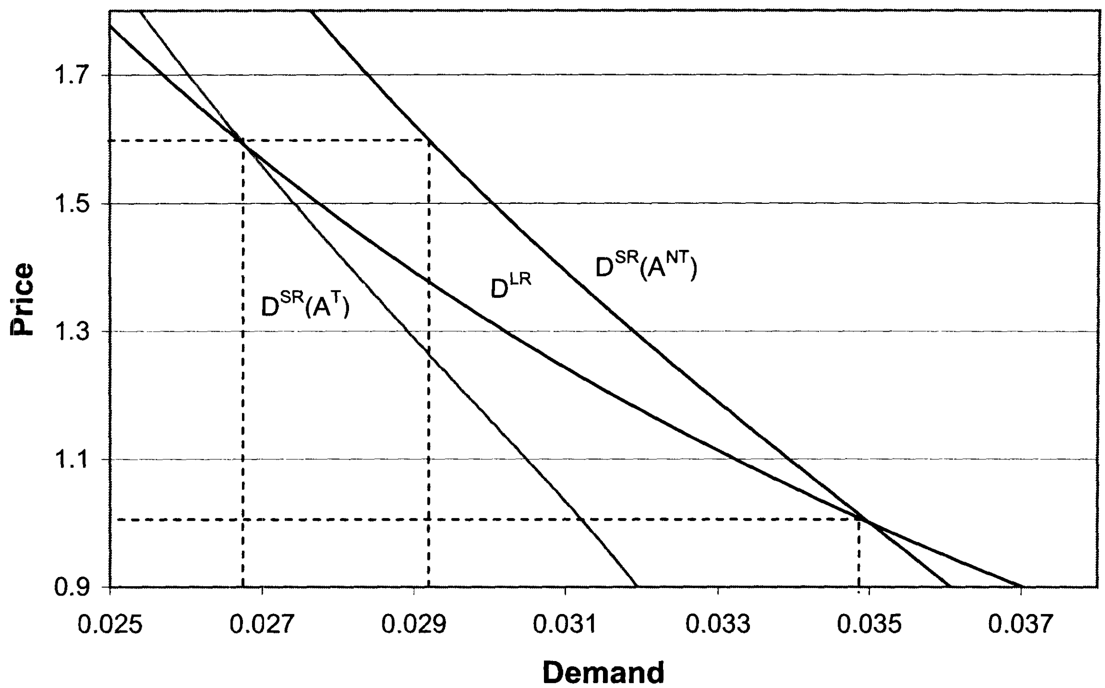
Notes: Standard errors in parentheses. Coefficients are from the estimation of equations (3.25) and (3.26) by a SUR estimator (see text). The dependent variables are log total energy costs, and the coal, natural gas, residual and electricity cost shares. Because of the symmetry restrictions, the 5 x 5 matrix of β is symmetric. Monotonicity is the fraction of observations for which all predicted cost shares are positive.

Figure 3.1: Short Run vs Long Run Effect of a Carbon Tax



Notes: The curve $D^{SR}(A^{NT})$ is the demand for coal, as a function of the price of coal, for plants that enter with no carbon tax in place, holding fixed technology, A^{NT}. The curve $D^{SR}(A^T)$ is the corresponding curve for plants that enter after the policy has been implemented, with constant technology, A^T. The curve D^{LR} is the demand for coal, as a function of the price of coal at entry. D^{LR} intersects $D^{SR}(A^{NT})$ at P_0, the price of coal with no tax, and intersects $D^{SR}(A^T)$ at P_1, the price of coal with a tax.

Figure 3.2: Estimated Short and Long Run Demand for Coal



Notes: The curve $D^{SR}(A^{NT})$ is the coal demand of a plant that enters with all input prices equal to one, holding technology fixed. Demand is calculated using the cost function parameters in equation (3.26); see text for details. $D^{SR}(^{AT})$ is the corresponding curve for a plant that enters with the price of coal equal to 1.59, and all other prices equal to one. D^{LR} is the coal demand of a plant that enters with the expected price of coal equal to the corresponding price, and all other prices equal to one; it intersects $D^{SR}(A^{NT})$ when the price equals one, and $D^{SR}(^{AT})$ when the price equals 1.59.