GENERIC DRUG INDUSTRY DYNAMICS

David Reiffen and Michael R. Ward*

Abstract—Because of its unique institutional and regulatory features, the generic drug industry provides a useful laboratory for understanding how competition evolves. We exploit these features to estimate a system of structural relationships in this industry, including the relationship between price and the number of competitors, and between drug characteristics and the entry process. Our methodology yields a number of findings regarding industry dynamics. We find that generic drug prices fall with increasing number of competitors, but remain above long-run marginal cost until there are eight or more competitors. We also find the size and time paths of generic revenues, rents, and the number of firms are greatly affected by expected market size. Finally, we show how estimates derived from a system of structural equations can be used to simulate the effect of changes in an exogenous variable.

I. Introduction

Both the economics literature and the business press suggest that a typical pattern for a "new" industry [or what Jovanovic and MacDonald (1994) call an *invention*] is to have an initial phase in which a small number of firms each earn significant profits, followed by a phase in which rapid entry of new firms leads to increased competition and dissipation of some of those profits, often accompanied by a shakeout, whereby only a few large firms remain (especially if subsequent innovations increase the optimal scale). Although this pattern seems to characterize many industries, the length of time during which early movers retain their profits, how prices adjust during the entry process, and the degree of shakeout vary widely across industries (Gort & Klepper, 1982). Because the factors that influence the timing of entry and exit are idiosyncratic to each industry, empirical studies of this process tend to focus on a single industry (see, for example, Gisser, 1999; Klepper & Simons, 2000), and in a sense constitute a single data point, making generalizations tenuous.

Several characteristics of the generic drug industry result in it being a useful laboratory for understanding how competition evolves within a market. First, each chemical represents a distinct experiment. There are a large number of individual experiments within the same industry, providing multiple observations on similar dynamic processes.[1] Second, information about the market for each drug is observable to researchers.[2] For example, because a market begins when the patent on an existing drug expires, the date at which the market opens to competitors is known in advance and the potential revenue can be projected with some accuracy by both participants and researchers. Because entry occurs at observable points in time, the consequences of changes in the number of producers on pricing is measurable. Moreover, because entry requires Food and Drug Administration (FDA) approval, firms must sink significant costs to apply for approval prior to knowing when, or how many, rivals will enter the market. Hence, firms must determine if their expected post-entry rents are sufficient to justify the costs sunk prior to entry.

These features enable us to impose restrictions from economic theory that identify the key structural relationships describing the evolution of these markets. Two simultaneously determined relationships are the effect of available rents on the pattern of entry over time and the effect of changes in industry structure (namely, entry) on rents. The latter relationship can be estimated because the process of FDA approval takes the timing of entry decisions out of the hands of individual firms, so that the number of firms at any point in time is not determined by the current price, yet the price is affected by the number of competitors. Thus we can estimate the effect of the number of firms on current price, assuming that current industry structure is exogenous. Combining that with estimates of revenue, we are able to calculate the expected rents conditional on the number of competitors and the elapsed time since market opening

We develop an iterative estimator to determine the effects of rents on entry. A zero-expected-profit condition is exploited that equates the expected number of entrants to the ratio of total generic industry rents to sunk entry costs per firm. At the same time, the rents available to potential entrants will depend on the number of entrants. We estimate the probability of any number of competitors in each time period as a function of the available rents. We then use these estimates, together with the industry rents conditional on the number of firms and time, to calculate the available rents. Equilibrium is obtained when the rents predicted by the entry parameters equal the rents assumed in their estimation.

Our structural estimates yield a number of empirical findings. First, consistent with previous work, we find that generic drug prices fall with an increase in the number of competitors. Though estimating the relationship between market structure and prices is a necessary component of estimating our system of structural relationships, the esti-

Received for publication October 15, 2002. Revision accepted for publication October 28, 2003.

* Commodity Futures Trading Commission and University of Texas at Arlington, respectively.

We would like to thank our former employer, the Federal Trade Commission, for acquiring the data; David Balan, Denis Breen, Bill Crowder, George Deltas, Craig Depken, Haj Hadeishi, Ron Hansen, Dan Hosken, Paul Pautler, and Mick Vaman for their helpful comments on previous drafts of this paper; Dr. James Loverde for technical advice; and Sara Harkavy for assistance. The views expressed in this paper are solely those of the authors, and not necessarily those of their current or past employers.

[1] Bresnahan and Reiss (1991) also look at industries with multiple (in their case, geographic) markets in each industry. Their work focuses on characterizing how the static equilibrium varies across markets, rather than intramarket dynamics.

[2] We use the terms *market*, *generic drug*, and *chemical* interchangeably. All three terms simply refer to a prescription drug whose patent has expired. In particular, the use of the term *market* may not correspond to its antitrust meaning.

The Review of Economics and Statistics, February 2005, 87(1): 37–49

mated effect of entry on price is also of independent interest, for this relationship has been an area of ongoing interest in the industrial organization literature.[3] We calculate that prices for the initial generic monopolist are 20%–30% (or perhaps even more) above long-run marginal costs. Generic prices steadily decline with an increase in the number of producers and begin to approach long-run marginal cost when there are 10 or more competitors. Second, more firms enter, and enter more quickly, in markets with greater expected rents. Finally, we find that the flow of generic industry profits increases as revenues grow, but begin falling after 5 to 12 months, as entry reduces price-cost margins. In addition, we find that this pattern is accelerated in larger markets because competitors enter more quickly.

An advantage of estimating the set of structural relationships that constitute the equilibrium is that one can trace through the effect of changes in market characteristics on the equilibrium. This can be particularly valuable in evaluating the effects of alternative policies. To illustrate the use of these estimates to inform policy, we simulate the effect of an actual change in the competitive environment in this industry. In response to a scandal involving illegitimate approvals, the FDA increased its scrutiny of generic drug applications in mid-1989. Though the policy may have allowed the FDA to discover, and therefore reject, more substandard applications, it also raised the cost of obtaining FDA approval for qualified applicants. Our estimates provide a means of determining the effect of the higher entry costs on long-run generic prices.

II. Background

Before marketing a new chemical entity, a prospective manufacturer must obtain FDA approval. To obtain a new drug approval (NDA) from the FDA requires demonstrating that a drug is safe and efficacious, which is both expensive and time-consuming. It has been estimated that for the average drug that was obtained FDA approval in the 1990s, its producer had spent over $335 million (in 2000 dollars) on development, and an additional $467 million on clinical and other testing.[4] In addition, the clinical trial process took approximately 8 years.

Prior to 1984, producing a generic version of most existing drugs involved a similar application process. Although the generic producer did not face the cost of drug discovery, it still bore the costs of demonstrating the safety and efficacy of its version. The Waxman-Hatch Act in 1984 created an abbreviated new drug approval (ANDA) procedure that reduced the regulatory burden for generic producers by requiring only that they demonstrate bioequivalence to a drug that was already approved by the FDA. The ability to avoid safety and efficacy testing considerably reduced the cost of obtaining FDA approval. As discussed below, we estimate that the cost of applying for an ANDA (including the cost of the requisite testing) was approximately $603,000 in the early 1990s (and approximately $338,000 in the period immediately following passage of the Act).

Not surprisingly, this expedited approval process has increased the number of firms producing generic versions of previously patented drugs. Cook (1998) reports that for 13 major drugs with patents expiring between 1990 and 1993, 11 had generic entry within 2-months of patent expiration. In contrast, she notes that in Caves, Whinston, and Hurwicz's (1991) study of pre-Waxman-Hatch entry (between 1976 and 1982), only 2 of the top 13 drugs had generic entry within 1 year of patent expiration.

Entry still requires significant up-front expenditures, with a payoff that depends both on the FDA's decisions with respect to a firm's application, and the timing of FDA approval of rivals' ANDA applications. The time it takes the FDA to process applications can be both considerable and variable. In the vast majority of cases, the initial ANDA application is found deficient, requiring the applicant to conduct additional tests or submit additional material. Often, approval is granted only after the applicant has gone through two or three resubmissions. Hence, from the applicant's perspective, the time between initial submission and FDA approval is quite variable. Scott Morton (1996) calculates that between 1984 and 1994 the time between the initial application and approval of ANDAs averaged approximately 19 months, with considerable year-to-year variation. In addition, entry requires time to obtain an approved source of materials and adequate production facilities. In total, the applicant has to anticipate 2 to 3 years elapsing from the time it begins preparing to enter until it can begin selling a generic drug.

III. Modeling Industry Dynamics

Two features of the entry process in this industry are important to understanding industry dynamics. First, an entrant's timing of entry into the market is largely not under its control. Not only is the date of its approval by the FDA uncertain, but each applicant lacks knowledge of when, or how many, other ANDAs for that drug will be approved. Thus, potential entrants make their entry decisions simultaneously (although actual entry will typically be sequential). Second, an individual entrant's share of the aggregate generic profits will depend greatly on when it gains approval relative to other generic producers. Firms gaining approval earlier face fewer competitors initially, and are able to sell for a longer time. There is some evidence that earlier entrants earn greater profits even after rivals have entered.[5]

[3] See Bresnahan (1989) for a discussion and analysis of this literature.

[4] See DiMasi, Hansen, and Grabowski (2003). This figure represents the expected cost of a successful drug, in the sense that it is adjusted for the probability that a drug never obtains an NDA.

[5] In addition to anecdotal evidence from industry participants, Cook (1998) shows that sales are highly concentrated among firms in each

Together these two features create a "lottery" for prospective producers of a generic version of a drug. If a firm obtains early approval, it is likely to earn a positive return on its application-related costs, whereas firms obtaining approval later in the process are likely not to recover their sunk costs. Thus, in contrast to markets in which entry decisions are sequential and competition results in the last, or *marginal* entrant earning zero profit, here the number of firms adjusts until the *average* firm earns zero profit.[6] Specifically (assuming n identical applicants), the expected profit for each firm from applying for an ANDA is

$$\textit{Expected Profit} = \frac{1}{E[n]} \sum_{t=1}^{\infty} \beta^t \left(\sum_{i=1}^{n} \rho_{it} \Pi_{it} \right) - A = \frac{V}{E[n]} - A , \tag{1}$$

where Π_{it} is total generic industry profits at time t with i firms in the market, ρ_{it} is the probability that i firms are in the market at time t, A is the cost of applying for an ANDA, and β is the discount factor. V is defined as the present value of the stream of expected rents for all generic producers of a drug. The goal of the empirical analysis in this paper is to estimate the key parameters of equation (1). Specifically, we estimate the interrelationships that allow us to calculate the Π_{it} and ρ_{it} as functions of exogenous, drug-specific variables. The remainder of this section details the estimation procedure. In brief, each Π_{it} can be thought of as the product of two factors: total revenue and price-cost margins. Accordingly, we calculate the Π_{it} by combining the results of regressions of each of price-cost margins and revenues against explanatory variables, such as time since patent expiration. Given these estimates, we can then determine Π_{it} conditional on any given i and t. To calculate the probability that a given Π_{it} occurs (ρ_{it}), we estimate two structural relationships: the relationship between the number of applicants for ANDAs (n) and rents in a market (V), and the relationship between the timing of FDA approval and rents. Thus, for any given levels of rents (and given set of exogenous variables), we can use these two relationships to calculate ρ_{it}. Note that because total rents both determine and are determined by ρ_{it}, these relationship must be estimated simultaneously.

A. *The Effect of Generic Industry Structure on Profits*

Generic price-cost margins are estimated as a function of observable market characteristics, including the number of generic competitors. We are interested in a specific aspect of the relationship between margins and the number of competitors, an aspect that is not explicitly examined elsewhere: how the marginal effect of an additional competitor on a drug's prices changes with the number of firms that already have an ANDA for that drug. To address this question, we estimate a regression of the form

$$\frac{P_{ik}}{P_{bk}} = \alpha_0 + \sum_{i=1}^{N-1} \alpha_i D_i + \sum_j \gamma_j X_{jk} , \tag{2}$$

where P_{ik} is the price in the post-patent-expiration period when there are i generic firms with FDA approval producing chemical k, and P_{bk} is the price of the branded version of product k during the year prior to patent expiration.[7] D_i is a dummy variable that equals 1 when there are i generic producers of chemical k and 0 otherwise, and the X_{kj} are variables representing demand or cost shifters for drug k.

Using dummy variables for the number of generic producers imposes no specific structure on the relationship between price and the number of competitors. This contrasts with some previous work, in which a specific structure on the relationship is assumed (for example, an a priori functional form is imposed on the effect of more firms on generic prices).[8] Each such specification makes implicit assumptions about the pattern of price effects that can result from entry. For example, the implicit assumption made when the number of firms is used as an explanatory variable is that the effect of an increase by 1 in the number of firms is independent of the initial number of firms. By allowing the marginal effect of an additional firm to vary with the number of firms, we can examine questions such as the number of firms necessary to lead to approximately marginal cost pricing. Allowing the marginal effect to vary is also important to our goal of accurately measuring the rents associated with any specific number of generic competitors.

This relationship can be viewed as structural only if one views the number of firms at any time as exogenous. One standard criticism of empirical studies of the relationship between market structure and prices is that structure is not exogenous, but rather is determined by the profitability of

market; even in markets with more than ten firms, the top two generic producers typically sell more than 60% of the units. Bond and Lean (1977) and Berndt et al. (1995) provide several examples of drugs for which the first entrant had a substantial advantage.

[6] Consequently, in contrast to the markets examined here, in a market with sequential entry, changes in the profits earned by the first entrant will not change subsequent firms' incentive to enter. Another important difference between generic drug markets (where entry decisions can be viewed as simultaneous) and other markets is that an exogenous change in the number of competitors (for example, due to a merger several years after patent expiration) can lead to higher prices without inducing entry, even if firms outside the market have the same entry costs as the incumbents.

[7] We use the branded price before patent expiration, rather than the contemporaneous branded price, because the latter is likely to be determined jointly with the generic price (as noted in footnote 21, the empirical evidence on the importance of this relationship is mixed). In contrast, the branded price before there is any generic entry is likely to be independent of the number of generic producers in future periods.

[8] For example, in other studies of generic drug competition, generic price is assumed to vary linearly with j, the number of firms (Frank and Salkever, 1997); with j and j^2 (Caves, Whinston, & Hurwicz, 1991); or with j and $1/j$ (Wiggins & Maness, 2004). These papers are discussed at greater length in section V.

entering the market.[9] This criticism implies that the observed cross-sectional relationship between price and the number of firms is an equilibrium relationship reflecting market-specific differences, and not a structural one reflecting the effect of more competitors on price. That is, as equation (1) illustrates, the number of firms applying for ANDAs adjusts in response to the available rents. However, in the generic drug industry, the nature of the FDA review process makes it unlikely that the number of firms at a point in time is affected by current price, within the time series of prices for any one drug. Most ANDA applications are submitted before the generic market even exists, and the number of competitors at any point in time depends on the FDA review process (most applications must be resubmitted multiple times). Hence, though the eventual number of approvals for a drug is related to the aggregate rents, the actual number of FDA-approved firms at any point in time may plausibly be considered independent of the contemporaneous price. A potential endogeneity issue arises when aggregating across drugs because there are unobserved differences between drugs that might affect both prices and the number of entrants. We control for these between-drug effects by estimating a random-effects model. This model allows there to be differences across drugs in the average relationship between generic prices and pre-patent-expiration branded prices (see Greene, 2003).[10] Finally, we tested this potential endogeneity using a Hausman test and cannot reject the null hypothesis that market structure is exogenous in the pricing equation (Hausman, 1978).[11]

In principle, N in equation (2) could be the maximum number of entrants observed in the data. In practice, we take N to be the minimum number of entrants such that the price effect of further entry is negligible. The interpretation of α_0 is the ratio of the generic price when there are more than N generic producers to the branded price that prevailed before patent expiration, if all other independent variables were equal to 0. The other α_i, such as α_5, are the increments in the ratio over α_0 when there are i producers. Because α_0 reflects the ratio below which additional entry does not lead to lower prices, we view $(\alpha_0 + \Sigma\gamma_j\bar{X}_{kj})\ P_{b,k}$ as the long-run marginal production cost of drug k (where $\bar{X}_{kj}$ is the mean value of X_j for drug k). Under this assumption, $\alpha_i/(\alpha_0 + \alpha_i + \Sigma\gamma_j\bar{X}_{kj})$ is a measure of the price-cost margin with i generic producers.

The other relationship required for calculating V conditional on the ρ_{it} is the relationship between generic revenue and market-specific variables. Our estimation of this relationship is of the form

$$\ln(P_{kt}\,Q_{kt}) = \tau_0 + \tau_1 \ln(P_{bk}\,Q_{bk}) + \sum_{j=2}^{J} \tau_j X_{kj}, \qquad (3)$$

where $P_{kt}\,Q_{kt}$ is total monthly generic industry revenue at time t in market k, $P_{bk}\,Q_{bk}$ is the branded firm's average monthly revenue during the year prior to patent expiration, and the X_{kj} are other variables that might affect generic revenue. The X_{kj} will include many of the same variables as equation (2).

B. *The Effect of Industry Profitability on Entry*

The model we use to examine entry decisions treats each of M firms as homogeneous in regard to their ability to enter and produce a generic drug. We assume that generic rents are not sufficient to allow all M potential entrants to profitably enter any market, but that they are sufficient to allow one firm to earn profits in any market. We also make the natural assumption that each firm's profits from producing a drug are decreasing in the number of rival producers of the drug. We conceptualize the entry decision as each firm choosing independently and simultaneously whether to enter each market. This reflects the reality that each generic producer must independently decide whether to enter a market, at a point usually 2 to 3 years prior to patent expiration. The symmetric (mixed strategy) Nash equilibrium in this case will consist of each firm i choosing to enter market k with some probability μ_{ik}, where that μ_{ik} may depend on the expected rents in the market.[12] The symmetric Nash equilibrium in each market consists of a μ_k that is common to each firm, and that has the property that each firm optimally chooses it, given that all of its rivals have chosen that same μ_k. The μ_k in the Nash equilibrium yields zero expected profits, the logic being that if an entry probability generates sufficiently few expected entrants so that each entrant expects to earn positive profits, then any firm would be better off unilaterally changing its strategy to entering with probability 1 (and entering with probability 0 if expected profits are negative). Comparing across drugs, the equilibrium μ_k will be increasing in the expected rents associated with that drug, so that we expect to see more entrants for higher-V_k drugs.

One feature of this stylized game is that, because each firm's decision whether to enter is independent of all other firms' decisions, the equilibrium distribution of the number of entrants will follow a binomial distribution. We use the Poisson distribution as an approximation of the binomial to

[9] This criticism dates back at least to Demsetz (1973). For more formal analysis, see Bresnahan (1989).

[10] Other studies have allowed for drug-specific effects by including market-specific dummy variables. Either assumption allows calculation of the average effect of increasing the number of competitors in a market.

[11] Following Frank and Salkever (1997) and Caves et al. (1991), we use time since patent expiration and pre-patent branded revenues as instruments for the number of generic firms. Because we do not have enough instruments to estimate equation (2), our endogeneity tests employ several common functional forms of the number of competitors.

[12] An alternative equilibrium concept is employed by Berry (1992), who assumes that firms' entry costs differ. Given variation in entry costs, a pure-strategy equilibrium can emerge, in which only low-entry-cost firms enter.

derive the density function of the number of entrants in market k as

$$f(n_k) = \exp(-\mu_k)\,\mu_k^{n_k}/n_k!, \tag{4}$$

where μ_k is the equilibrium entry probability. The zero-expected-profits condition implies that $E[n_k]A = E[V]$. Because $E[n_k] = M\mu_k$ with the Poisson distribution, this yields $E[V_k]/M\mu_k = A$; that is, applications costs are equal to expected rents divided by the expected number of entrants. This implies that holding M and A constant, there should be a direct relationship between V_k and μ_k. There is reason to believe, however, that application costs increased substantially in 1989, when it was discovered that some ANDAs had been fraudulently obtained, and that the FDA reacted by increasing its scrutiny of applications (Scott Morton, 1996). We attempt to capture this in a dummy variable, *Stringent*, that equals 1 for the period after mid-1989 and 0 otherwise. Consequently, we estimate the relationship between μ_k and the cost and benefit of applying as

$$\mu_k = V_k \exp(\phi_1 + \phi_2\, Stringent_k) \tag{5}$$

from a cross section of 31 drugs. This relationship characterizes how the number of entrants adjusts to changes in the costs and benefits of FDA approval. It also provides us with a means of estimating the time series of entry within each market, because the expected number of producers at each point in time depends on the total number of applications, as detailed below.

For any given number of applicants, the pattern of entry will depend on the FDA review process. Our second entry equation characterizes the timing of entry, conditional on the total number of entrants. To reflect the stochastic nature (from the applicants' perspective) of the FDA review process, we model the rate of entry as a proportional hazard function in which the proportionality parameter is possibly affected by rents available and FDA regime. Specifically, we posit a probability λ of any firm that has not yet been approved obtaining an ANDA during month t. We estimate the following relationship for λ_k:

$$\ln \lambda_k = \theta_1 + \theta_2 V_k + \theta_3\, Stringent_k. \tag{6}$$

We postulate that λ_k may be increasing in V_k, because firms apply earlier in high-V_k markets and/or have a greater incentive to file accurately. Because the value of λ may also depend on the regulatory environment, equation (6) includes *Stringent*, our postscandal dummy variable. These parameters are estimated from data on the time to entry for all entrants in each of 31 generic drugs.

Combining equations (4) and (6) allows us to calculate the time path of expected entry, as a function of rents and the FDA regime. Specifically, we use the estimate of λ_k from equation (6) to determine the survivorship function, where surviving means the applicant has not yet been approved. This function is defined in terms of the hazard proportionality parameter as $S_{kt} = \exp(-\lambda_k\, t)$. Then, using equation (4) and the binomial formula, we calculate the probability that i firms have ANDAs in market k in period t as

$$\rho_{ikt} = \sum_{n_k} f(n_k)\,\frac{n_k!}{(n_k - i)!i!}\,(1 - S_{kt})^i S_{kt}^{n_k - i}. \tag{7}$$

C. *The Endogeneity of Rents and Identification*

Equations (2), (3), and (7) together make up the components of equation (1) and thus allow for the calculation of industry rents,

$$V_k = \sum_{t=1}^{\infty} \beta^t \left(\sum_{i=1}^{n} \rho_{ikt}\Pi_{ikt} \right)$$

$$= \sum_{t=1}^{\infty} \beta^t \left(\sum_{i=1}^{n} \rho_{ikt}\,\frac{P_{ikt} - c}{P_{ikt}}\,P_{kt}\,Q_{kt} \right). \tag{1'}$$

However, equation (7), governing the entry process, also depends on the magnitude of the expected available rents through equations (5) and (6). Larger expected rents V generate larger probabilities of entry, μ, shifting the probabilities ρ toward more firms at any point in time, which by equation (1′) tends to reduce expected rents V. Because V is endogenous, via equation (1′), we develop an iterative process to estimate the parameters of equations (5) and (6).

The mixed-strategy simultaneous-move Nash equilibrium suggested above will represent a stable fixed point in the mapping of V onto V under certain conditions. Specifically, the system of equations consisting of equations (1′), (5) and (6), along with subsidiary relationships embodied in those equations [for example, equation (2) within equation (1′)], will have a fixed point (V^*, λ^*, μ^*) if equations (2) and (3) indicate that per-firm profits are decreasing in the number of firms, and equations (5) and (6) indicate that λ and μ are such that the expected number of firms at every point in time is increasing in V. To see why (V^*, λ^*, μ^*) represents a fixed point, consider an alternative V, $V^a > V^*$. Because $V^a > V^*$, the λ and μ based on V^a will lead to more firms at each point in time if the second stability condition holds. Consequently, if the first stability condition holds, the V resulting from this λ and μ will be less than V^a. Hence, V's above V^* map to lower V's, and V's below V^* map to higher V's.

The actual calculation of the fixed point follows this same logic. In the first iteration, we calculate V_1 using equation (1′) based on arbitrary values of λ and μ, along with the parameters estimated in equations (2) and (3). We then estimate the λ and μ, using equations (5) and (6) with V_1 on the right-hand side. These are used to calculate $f(n)$, the density of n_k, and the ρ_{ikt} according to equations (4) and (7).

We combine the n_k and ρ_{ikt} from this iteration with the π_{ikt} calculated from equations (2) and (3) to calculate V_2. We then compare V_2 with V_1 and if the two values are sufficiently close, we view the process as convergent; that is, these values of λ, μ, and V are the equilibrium. If the predicted V is sufficiently different from the initial value, we repeat the process, using V_2 as the right-side value in reestimating equations (5) and (6), and then calculating V_3 based on the new λ and μ and the unchanged π_{ikt}. In this way, we iterate through a series of V_2 until we obtain convergence.

IV. Data

Our primary source for price and quantity data is Generic Spectra® from IMS Inc., a proprietary vendor of information to the pharmaceutical industry. The IMS data provide information on 31 drugs that went off patent in the late 1980s and early 1990s, and subsequently faced competition from generic producers (see table 4). It includes information on monthly price and quantity for the patent holder and generic entrants for 3 years subsequent to patent expiration and 3 years prior to patent expiration (for the patent holder). These data include prices derived from two distinct sources: product shipments and price surveys. For both sources the data are provided separately for each strength (for example, 50 mg) and form (for example, oral solid) of the drug.

The shipment-based data on revenues and quantities are derived primarily from shipments by distributors (who purchase from manufacturers) to pharmacies and other dispensers. A small proportion, perhaps 5%, of sales are made directly by manufacturers. The sales by distributors are captured by IMS directly monitoring the shipments of a high percentage of distributors (98% of all such shipments are contained in their sample). This is combined with estimates of direct sales of manufacturers, which are estimated from a sample of invoices. Our measure of price per kilogram is the average revenue for a particular strength and form derived by dividing total generic revenue by the number of kilograms of generic product. We calculate this price separately for all generic sales, and for sales by the first generic entrant.[13]

The second set of prices in Generic Spectra is obtained from a sample of pharmacies. It includes data on average transaction prices paid by pharmacies. According to IMS, the measured acquisition price would reflect all relevant discounts, with the exception of year-end quantity discounts provided by some manufacturers. We calculate acquisition prices for both the first and the average generic seller.

For drugs with multiple strength-form combinations (types), we constructed a price series using price data only on the type of the drug that generated the most generic revenue.[14] For all but two of the drugs the most popular generic type was also the best-selling type for the innovator firm. For the two exceptions (Metaproterenol and Albuterol), there was no generic version of the most popular type.[15] For all drugs, the price of the best-selling generic type, rather than the overall average price, was chosen in order to distinguish changes in the price of specific products from changes in the mix of strengths and forms. A disadvantage of this approach is that we ignore information on price changes for other types. The tradeoff seems to favor our approach if the manufacturer does not anticipate changes in relative sales volume mix, so that the forces changing relative shares are uncorrelated with those changing prices. On the other hand, if changes in relative demand are anticipated, then prices will move for the same reasons as relative demand, and price changes for one strength may understate or overstate the "average" change in prices.

Another issue we faced was what time period constitutes an observation. For the estimation of equation (2), describing how pricing reflects industry structure, we aggregate months with the same number of entrants into one observation. Treating multiple months with the same industry structure as separate observations could artificially inflate the statistical significance of changes in industry structure.[16] Our approach reduces the number of observations substantially, which tends to reduce the statistical significance of our results, but as table 1 indicates, we are still able to find significant pricing effects. For the estimation of equation (3), forecasting generic revenue, we treat each month as a unique observation, for two reasons. First, consumer adoption of generic drugs is likely to be affected by the passage of time independently of other factors. Second, we are less interested in testing the parameter values from these estimates than we are in obtaining accurate forecasts.

Our measure of the number of entrants is the number of FDA-approved generic producers. Data on the timing of entry were collected from the FDA publication *Approved Drug Products,* commonly referred to as the Orange Book. This lists the date each firm received its NDA or ANDA from the FDA as well as information enabling us to deter-

[13] To the extent that there are first-mover advantages, the first generic seller's product may be a more homogeneous good over time than the average generic seller's product. Note, however, that the first generic seller will, in many cases, be a reseller, not a generic manufacturer.

[14] Because the price data covered retail pharmacies, forms of the drugs that are not typically sold by pharmacies (for example, injectables) are excluded from the price analysis.

[15] For these two drugs, the most popular type of the branded production was an aerosol inhalant. Entry into the generic production of this type came several years after generic entry into the drug types reflected in our data. We believe that the delay in developing a generic aerosol, after the patent on the chemical had expired, was due to an unexpired patent on the aerosol delivery system. There also were unresolved issues related to demonstrating bioequivalence of generic aerosol products to the branded versions.

[16] As Mouton (1986) observes, using multiple observations with essentially unchanged exogenous variables leads to a downward bias in estimated standard errors. For this reason, we chose the conservative approach of taking only one data point for each number of competitors in each market.

TABLE 1.—RANDOM-EFFECTS PRICE REGRESSION RESULTS USING GENERIC SPECTRA® PHARMACY DATA

	Average Wholesale Price	First Wholesale Price	Average Revenue/Quantity	First Revenue/Quantity
Intercept	0.675* (0.107)	0.729* (0.144)	0.698* (0.174)	0.341* (0.120)
Multiple	0.036 (0.047)	−0.015 (0.070)	−0.034 (0.092)	0.023 (0.054)
Number of uses	−0.004 (0.012)	−0.014 (0.018)	−0.007 (0.023)	0.005 (0.013)
Initial number of substitutes	0.003 (0.005)	0.007 (0.008)	−0.002 (0.011)	0.001 (0.006)
Revenue growth	−0.033 (0.043)	−0.048 (0.065)	−0.029 (0.086)	0.049 (0.050)
Change in substitutes	−0.030+ (0.013)	0.008 (0.013)	−0.044* (0.009)	−0.028+ (0.012)
Time (times 10)	−0.005 (0.011)	−0.015 (0.012)	−0.049* (0.009)	−0.016 (0.011)
One firm	0.249* (0.054)	0.171* (0.059)	0.093* (0.044)	0.258* (0.054)
Two firms	0.181* (0.047)	0.143* (0.046)	0.131* (0.037)	0.272* (0.045)
Three firms	0.169* (0.042)	0.177* (0.046)	0.094* (0.034)	0.232* (0.042)
Four firms	0.152* (0.037)	0.138* (0.040)	0.106* (0.030)	0.236* (0.037)
Five firms	0.128* (0.037)	0.080+ (0.039)	0.068+ (0.029)	0.183* (0.037)
Six firms	0.112* (0.034)	0.098* (0.035)	0.059+ (0.026)	0.149* (0.033)
Seven firms	0.089+ (0.039)	0.105* (0.040)	−0.003 (0.029)	0.114* (0.037)
Eight firms	0.092* (0.034)	0.067+ (0.034)	0.017 (0.025)	0.092* (0.032)
Nine firms	0.090* (0.033)	0.072+ (0.033)	0.035 (0.024)	0.096* (0.031)
Ten firms	0.059 (0.037)	0.024 (0.037)	0.039 (0.026)	0.048 (0.035)
Adjusted R^2	0.368	0.165	0.380	0.539
Number of obs.	164	166	168	166

Dependent variable is the ratio of the generic price to pre-expiry branded price. Asterisks denote significance at the 1% level, and plus signs denote significance at the 5% level. Standard errors in parentheses.

mine if there were multiple branded products prior to patent expiration. Because the Generic Spectra® data are limited to 3 years of post-patent-expiration data, we limit our analysis of entry to ANDAs awarded within 3-years of patent expiration.[17]

Finally, we constructed two demand-side variables—*Uses* and *Subs*—to capture demand differences across drugs from American Hospital Formulary Service (AHFS) *Drug Information* (1996), augmented by the AMA *Guide to Prescription and Over-the-Counter Drugs* and *The People's Pharmacy*. For two classes of drugs (hypotensives and antibiotics), we also consulted a practicing internist. From these sources, we determined the ailments, or *indications*, for which each drug is used. To the extent possible, we included not only FDA-approved (or *labeled*) indications, but significant unlabeled uses as well. The number of different indications constitutes our *Uses* variable, and $Substitutes_0$ is the number of alternative drugs used to treat all the *Uses* at the time of patent expiration.[18] Finally, we create a variable *Change in Substitutes* that varies in time for each drug. It is defined as the number of new substitutes for that drug that had entered the market since patent expiration.

[17] A useful feature of the 3-year time frame is that the process of entering generally takes about 3 years. Hence, any firm that receives an ANDA in this time frame will have commenced the process prior to observing the ANDAs that were awarded to other firms.

[18] Although the IMS provides information on the "therapeutic class" (e.g. cephalosporin antibiotics) of each drug, these categories tend to be overinclusive in that all drugs in the therapeutic class would not actually be used for the same ailment, as Caves et al. (1991) and Lu and Comanor (1998) have noted. Scott Morton (1996) finds that the *therapeutic class* variable has little predictive power in her regressions. Lu and Comanor (1998) follow a similar procedure to that used here and find that their measure does have explanatory power.

V. Results

The relationship between the number of producers and generic prices, as characterized in equation (2), is discussed below in section V A. Because price effects both play a part in other results and are of interest in themselves, we examine the robustness of those results by using several alternative price series. That subsection also includes a discussion of relationships between generic prices and the number of competitors that have been found in other studies. Section V B presents the revenue regression depicted in equation (3). By combining the estimates from equations (2) and (3), we can calculate the aggregate generic profits conditioned on the number of generic producers and the elapsed time since generic entry occurred. Section V C presents results relating to the entry parameters, μ and λ. As V, μ, and λ are jointly determined, equations (5) and (6) were estimated using the iterative procedure outlined above. We find that convergence occurs in the 18th iteration.[19]

A. *Prices and Structure*

Findings: Table 1 reports our estimates of the effect of the number of competitors on price for four alternative price series. The first two regressions are based on the sampled transaction prices paid by pharmacies; the last two are based on the average revenue received by manufacturers and distributors. The regressions are run separately for the average generic price and for the first generic entrant's price.

[19] We define convergence to occur when the squared sum of differences between the V_t^0 in successive iterations is less than 0.0001 times the average V.

The pricing regression include six factors other than the number of competitors: *Multiple, Uses, Substitutes*$_0$, *Revenue Growth, Change in Substitutes,* and *Time*. The first three of these represent factors that might affect the pre-expiry branded price, and only vary in the cross section (that is, between drugs). Since the endogenous variable here is the generic price divided by the pre-expiry branded price, factors that make the branded price lower will raise the ratio. *Multiple* equals 1 if there were multiple branded products in the market prior to patent expiration, and 0 otherwise.[20] To the extent the branded firms compete, pre-expiry prices would be lower for any given level of demand, which in turn implies a higher ratio of generic price to pre-expiry branded price, other things equal. The branded price of a drug with more *Uses* should be higher before patent expiration, and hence the ratio of marginal cost to pre-expiry branded price could be lower. More substitute chemicals in the pre-expiry period (*Substitutes*$_0$) should reduce the pre-expiry branded price, and hence lead to a higher ratio of marginal cost to pre-expiry branded price. The fourth factor that only varies between drugs is *Revenue Growth,* which is the average monthly change in revenue during the year prior to patent expiration. This is a proxy for expected post-expiry demand growth, which may influence generic prices.

The two variables that change over the sample period are *Time* and *Change in Substitutes*. The monthly time trend, *Time,* reflects any effects related to the passage of time, say learning-by-doing cost reductions, rather than generic entry per se. Finally, *Change in Substitutes* is the number of new substitutes for the drug since patent expiration. An increase in the number of substitutes once the patent has expired will likely reduce the current branded price (but probably not the pre-expiry branded price) and generic prices. Hence, holding the number of substitutes at the time of patent expiration fixed, we would expect that an increase in the number of substitutes will reduce the ratio of generic price to pre-expiry branded price.

To interpret these results, first note that the intercept represents the ratio of generic price to the branded price when the number of competitors is large and all other independent variables are equal to 0. For example, the estimate of 0.675 for the coefficient on the intercept in column 1 of table 1 implies that on average, the generic price would be 67.5% of the pre-expiry price of the branded product when there are 11 or more competitors, if all other variables were equal to 0. As discussed above, adding the other variables (evaluated at their mean values) times their coefficients to the intercept yields an estimate of the ratio of marginal cost to branded price, which here equals 0.631 (with standard error 0.031). The interpretation of the other firm number coefficients, such as the coefficient on one firm [α_1 from equation (2)] is the increase in this ratio due to having fewer than 11 generic competitors. For example, the coefficient 0.249 on one firm in column 1 implies that the ratio of generic price to pre-expiry branded price will be 0.880 (= 0.249 + 0.631) when there is a single generic firm. Note that, as one would anticipate, the α_i generally decline with increasing number of competitors. Again, using the example of the coefficient estimates from the first column, the ratio of generic price to pre-expiry branded price falls from 0.880 with one generic competitor to 0.812 with two generic competitors, and continues to decline toward 0.631 as the number of competitors rises.

The implied marginal costs tend to be lower in columns 3 and 4, in which average revenue is on the left-hand side, than in columns 1 and 2. For example, in column 3 we find that the implied ratio of marginal cost to pre-expiry branded price when there are 11 or more firms is approximately 50% (compared with an estimated ratio of approximately 63% in column 1).

Although there are some differences across the four regressions in regard to the magnitudes of the pricing effects, the general picture is quite similar across equations. In every case, there is an economically and statistically significant difference between the price when there is a single generic competitor and the price when there are a large number of generic competitors in the market. A premium remains (and, in some cases, is even slightly larger) when there are relatively few (between two and four) generic producers, but the premium falls and eventually disappears. The coefficient on 10 firms is small (less than 0.06) in all four columns, and not statistically significant in any. In all but the third column, the coefficients for seven to nine firms are positive and statistically significant, but all but two are less than 0.1. Generally, the results suggest a negative relationship between price and the number of firms, and that the marginal effect of an additional firm tends to decline with increasing number of firms.

Of the non-competition-related variables, none that differ only in the cross section have a consistent effect across regressions, and none are statistically significant in any regression. In contrast, the two variables that vary in the time series tend to have consistent signs and are sometimes statistically significant. *Time* has a negative effect in all four regressions, and is statistically significant in one of those four. The coefficient of −0.005 in column 1 means that the ratio of generic prices to pre-expiry branded price falls by approximately 0.02 over the first three years following patent expiration. *Change in Substitutes* is negative and statistically significant in three of the regressions, meaning that generic prices fall by approximately 0.03 if a new substitute drug gets FDA approval.

[20] Multiple brands might exist before patent expiration if the patent holder licensed the patent to another producer during the patent-protected period. This might occur if the two parties had some disagreement regarding which firm held the patent rights and reached a licensing agreement in lieu of litigation, or if two firms held complementary patents. In our sample, 7 of the 31 drugs had multiple brands prior to expiration.

Comparison with Other Results: Estimating the relationship between price and industry structure has a long history in industrial organization economics. One general criticism of the approach is that it implicitly assumes structure is exogenous, whereas in most industries it should be viewed as endogenous. Perhaps because features of the generic drug entry ameliorate the endogeneity problem, the relationship between the price of a generic drug and the number of firms producing that drug has been examined in at least three previous studies. All three studies of which we are aware use annual price and quantity data from IMS, and find a negative relationship between the generic price and the number of generic competitors.[21]

All three studies impose a specific functional form [for example, in Frank and Salkever (1997), price is assumed to be linearly related to the number of firms], and consequently, the coefficient estimates can only be indirectly compared to ours. For example, our results from the first two column of table 1 imply that an increase in the number of generic producers from 1 to 10 will reduce wholesale generic prices by approximately 30%. The estimates in column 4 are larger, with predicted price declines ranging to 40%. The estimates in Caves et al. (1991) imply that when there is only one generic producer, the price is approximately 40% below the pre-expiry branded price, and declines by approximately 50% (to 70% below the pre-expiry branded price) when there are 10 producers. The estimates from Frank and Salkever (1997) imply that an increase in the number of generic producers from 1 to 10 would reduce the price by 45%. Finally, Wiggins and Maness (1996) estimate that increasing the number of sellers (including both generic manufacturers and distributors) from 1 to 10 would lead to a 48% decrease in the average generic price. In general, the previous studies yield predicted effects that are slightly larger than the range of estimates in table 1.

B. *Revenue*

Equation (3) relates the total revenue derived from generic sales to other observable characteristics of the market. In contrast to equations (2), (5), and (6), we are not primarily interesting in testing any hypotheses about the individual parameters of equation (3). Rather, the main use of these results is in calculating V. The specification we estimate is the following:

$$\begin{aligned}\ln(GenericRev) = {} & \tau_0 + \tau_1 \ln(BrandRev) + \tau_2\, Multiple \\ & + \tau_3\, Details + \tau_4\, Forms \\ & + \tau_5\, Strengths + \tau_6\, Stringent \\ & + \tau_7\, \%Conv.Ins. + \tau_8\, Uses \\ & + \tau_9\, Subs + \tau_{10}(1/Time) \\ & + \tau_{11}(Time)\,.\end{aligned} \tag{3'}$$

The dependent variable is the natural logarithm of revenues from oral forms to generic manufacturers. The explanatory variables include: ln(*BrandRev*), the natural logarithm of average monthly total revenue for oral forms of the branded product(s) in the year before patent expiration; *Details*, the number of thousands of detail visits to physicians over the year two years prior to patent expiration; *Strengths*, the number of available strengths of the oral form of the drug; *Forms*, the number of available oral forms of the drug (for example, this would be equal to 2 if the product came in both an oral solid and an oral liquid); *%Conv.Ins.*, the percentage of individuals with health insurance who are covered by a fee-for-service structure, as opposed to some kind of managed care organization (MCO), and *Substiutes, Uses, Multiple, Stringent,* and *Time* as described above.

Table 2 presents the results of our estimation of equation (3′). Columns 1 and 2 set $\tau_{11} = 0$; columns 3 and 4 set $\tau_{10} = 0$. These alternative estimates allow us to examine whether allowing generic revenue to change linearly over time or in a specific nonlinear fashion (namely, as 1/time) yields a better fit. Columns 1 and 3 present pooled OLS estimates of the coefficient; columns 2 and 4 use random-effects estimation. Our coefficient estimates are fairly similar across specifications, and are largely consistent with expectations. For example, in all cases, τ_1 is between 1.04 and 1.1 and is not statistically different from 1 at the 5% level, indicating that a given percentage change in the branded drug's pre-expiry revenues increases generic revenues by a similar percentage. Past brand-name detailing, the most common form of prescription pharmaceutical promotion, increases subsequent generic demand (significantly in three of the four regressions), suggesting that promotions for the branded products spill over onto generic products. The availability of more forms decreases generic revenues, possibly because entrants do not always enter all forms. Similarly, in the pooled estimates, the availability of more strengths tends to decrease generic revenue from oral forms. One interesting finding in the pooled estimates is that conventional fee-for-service insurance decreases generic revenue. This could reflect the tendency for MCOs to have policies that encourage the use of generic drugs. An implication is that, if V is increasing in the percentage of patients covered by MCOs, and μ is increasing in V, it means that

[21] In contrast, the relationship between the number of generic producers and the branded price is less clear. Frank and Salkever (1997) and Aronsson, Bergman, and Rudholm (2001) find that generic entry has a relatively small positive effect on branded price. Caves et al. (1991) and Grabowski and Vernon (1992) find a small negative effect on branded price. Finally, Wiggins and Maness (2004) and Bhattacharya and Vogt (2003) find a large and significant negative effect of generic entry on branded prices.

TABLE 2.—REVENUE REGRESSIONS

Variable	Pooled OLS	Random Effects	Pooled OLS	Random Effects
Intercept	−0.371 (0.520)	−1.225+ (0.614)	−1.027 (0.659)	−2.081+ (0.883)
Log Pre-expiry Branded revenue	1.044* (0.045)	1.101* (0.139)	1.038* (0.050)	1.084* (0.155)
Number of details	1.904* (0.330)	1.156 (1.015)	1.955* (0.370)	1.300 (1.135)
Number of strengths	−0.067* (0.027)	0.015 (0.088)	−0.065+ (0.034)	0.014 (0.096)
Number of forms	−0.279* (0.044)	−0.258+ (0.149)	−0.280* (0.049)	−0.255 (0.164)
Percent with conventional insurance	−1.292 (0.797)	0.177 (0.804)	−1.672+ (0.958)	−0.111 (1.253)
Stringent FDA period	−0.476* (0.135)	−0.343* (0.125)	−0.504* (0.154)	−0.319+ (0.142)
Number of substitutes	−0.038* (0.009)	−0.063* (0.029)	−0.037* (0.010)	−0.050 (0.032)
Number of uses	0.087* (0.024)	0.063 (0.078)	0.086* (0.026)	0.041 (0.086)
Multiple brands dummy	−0.363* (0.095)	−0.450 (0.326)	−0.340* (0.104)	−0.399 (0.353)
Inverse time since patent expiration	−3.009* (0.184)	−3.113* (0.172)		
Time since patent expiration			0.029* (0.004)	0.032* (0.005)
Observations	982	982	982	982
Adjusted R^2	0.622	0.617	0.542	0.540

Standard errors are in parentheses and asterisk and plus sign superscripts indicate significance at the 1% and 10% levels.

greater MCO coverage leads to more entry and lower generic prices. As one would anticipate, generic revenues are lower in the postscandal period. In contrast to the pricing regression, we would expect both new and existing substitutes to work in the same direction, lowering generic revenue. For this reason, we include a single substitutes variable, which captures the number of contemporaneous substitutes. We indeed do find that the availability of more substitute chemicals for the same number of indications tends to decrease generic revenue. Finally, generic revenue increases over time in all specifications, but allowing generic revenues to vary in the nonlinear fashion (columns 1 and 2) seems to fit the data better.

C. *Entry*

As noted in section III, we are interested in explaining entry in two senses. First, we are interested in the cross-sectional relationship between the total number of firms applying for ANDAs for each drug in our sample and the available rents. Second, conditional on the total number of applicants in each market, we are interested in explaining the time series of entry.

Equation (5) relates the number of generic producers that ultimately enter each market to the available rents to generic entrants. Equation (6) relates the time series of entry for each drug to available rents and the regulatory environment, using a hazard function. The causality between available rents and the number of generic producers runs in both directions, so that equations (5) and (6) must be estimated using the iterative procedure described above. Using this procedure, we find that convergence occurs after 18 iterations. That is, the sum of squared differences in V between iterations generated by μ and λ is within 0.0001 of the average V used to estimate equations (5) and (6) in the 18th iteration. The structural estimates that result from this procedure are

$$\mu = V \underset{(0.09)}{\exp}(\underset{(0.09)}{1.08^*} - \underset{(0.14)}{0.58^*} \times \textit{Stringent}), \qquad (5')$$

$$\ln \lambda = \underset{(0.161)}{-2.08^*} - \underset{(0.035)}{0.040}\, V - \underset{(0.139)}{0.361^*} \times \textit{Stringent}, \qquad (6')$$

where the standard errors are in parentheses and asterisks indicate significance at the 1% level.[22] The correlation coefficient between the actual and predicted numbers of entrants across our sample is 0.58. This suggests our model is a reasonably accurate representation for such a parsimonious model (almost 1,000 observations and only five parameters) and suggests that the Poisson–hazard-rate analysis is a useful way of modeling the dynamic entry process.

The estimates in equation (5′) imply that during the nonscandal period the expected number of firms applying for ANDAs increased by approximately 2.9 [= exp(1.08)] with every $1 million increase in the available rents. The standard error of the constant estimate is 0.09, so that we can be highly confident that the effect of a $1 million increase in V is to increase the number of ANDAs by between 2.2 and 3.5 during the nonscandal period. These estimates also imply that the effect of rents on entry was smaller during the period following the generic drug scandal. This is consistent with newspaper accounts, which describe the postscandal period as one of greater FDA scrutiny of applications. Equation (5′) implies that the expected number of ANDAs increased by only approxi-

[22] The calculation uses the coefficient estimates from the first column of table 1 for the price effects of entry, and the second column of table 2 for the revenue forecasts. Because of the small sample size and the use of a generated regressor, the reported standard errors may be smaller than the relevant standard errors. Estimates of the coefficients obtained through a bootstrapping technique fail to reject bias in the coefficient estimates [that is, the estimates in equations (5′) and (6′) are similar to the bootstrap estimates]. However, bootstrapping does yield somewhat larger standard errors.

mately 1.7 [= exp(1.08 − 0.58)] with every $1 million increase in V during this period. Under our zero-profit assumption, the reciprocals of these estimates imply that the entry costs rose from $338,000 before the scandal to $603,000 after, so that the effect of increased scrutiny was substantial.[23] Using these estimated coefficients, the model implies that for a market with a V of $3.97 million (approximately the average market in our sample), μ would be 8.9 (evaluated at the mean value for *stringent*). This in turn implies that the probability of exactly one application for an ANDA in this market is approximately 0.12% [that is, $f(1) = 0.0012$], and we have $f(2) = 0.0054$, $f(3) = 0.0161$, and so on.

Equation (6) relates the conditional probability of FDA approval in any given month (λ) to the rents and regulatory environment, using a hazard function. The negative coefficient on the size of the rents (V) suggests that higher V reduces the probability that a given firm gains FDA approval in a given month conditional on the total number applying. Although this result is surprising, we note that it is neither statistically nor economically significant (a $1 million increase in V only reduces λ by approximately 2%). In fact, it is sufficiently small that, given the positive relationship between μ and V, the number of firms at each point in time is increasing in V. Consistent with the premise of greater scrutiny during the postscandal period, the coefficient on the *Stringent* term suggests that approval likelihood fell during that period. This effect is both statistically and economically significant (λ is approximately 15% higher in the prescandal period).

To illustrate the implications of these equations for market dynamics, we use the estimates of λ and μ from equations (5′) and (6′) to calculate the probability of i firms having gained approval by time t, using the binomial formula [as shown in equation (7)]. Table 3 shows, for the average-size market in our sample (monthly pre-expiry revenues of $7.95 million, so that V = $3.97 million), the probabilities of i firms gaining approval in the first 6 months following patent expiration. These probabilities change over time, and by 24 months after expiration, the likelihood that five or more applicants are approved in such a market is approximately 89.2%. For the typical large market in our sample (monthly pre-expiry revenues of $21.5 million, V = $6.23) the likelihood of five or more approved firms by 24 months after patent expiration is nearly 100%, and the expected number by 24 months is over 10.

[23] Scott Morton (2000) and Ellison and Ellison (2000) also examine the cross-sectional relationship between the number of entrants and market characteristics. Their results are not directly comparable to ours: many of the drugs in their samples were low-revenue drugs that had zero entry. Methodologically, these studies differs from ours in that they estimate the reduced-form relationship between the number of entrants and characteristics, rather than jointly estimating the structural relationships between rents and entry.

TABLE 3.—RELATIONSHIP BETWEEN TIME AND THE PROBABILITY OF ENTRY

Time t since Patent Expiration (months)	Probability of Entry (%)					
	$i = 0$	1	2	3	4	5+
1	46.7	35.5	13.5	3.4	0.6	0.2
2	23.3	33.9	24.7	12.0	4.4	1.7
3	12.3	25.8	27.0	18.8	9.9	6.2
4	6.9	18.4	24.6	22.0	14.7	13.4
5	4.0	13.0	20.8	22.2	17.8	22.1
6	2.5	9.2	17.0	20.9	19.3	31.2

Table shows the probabilities of i firms having ANDAs by month t for the median drug in our sample.

VI. Applications

One conclusion from the previous section is that incentive effects (as measured by the available rents) are important in determining the number of applicants for ANDAs. It follows that factors that reduce the available rents can have a significant effect on the number of applicants. In this section, we discuss some of the applications of those estimates.

A. *Calculation of Rents*

The calculations of the probabilities of entry over time presented in table 3 were based on a fixed amount of rent. As emphasized above, the calculation required to fully determine entry treats rents as endogenous, because entry and rents are jointly determined. Our methodology yields values for rents and entry for each of the 31 drugs. The predicted outcomes reflect drug-specific values for the exogenous variables.

The resultant predicted number of entrants and rents is depicted in table 4. Table 4 compares the expected number of approved ANDAs within 3 years of patent expiration with the actual number of approvals for each drug. For most of these drugs, this procedure seems to yield a fairly accurate prediction of the number of ANDAs.

B. *Dynamics*

Table 4 indicates how the total number of applicants and aggregate rents vary with features of each market. The dynamics of entry will reflect the total number of applicants as well as the speed of approval, and hence the dynamics of pricing will differ across markets. To illustrate, we calculated the dynamics for the median market in terms of pre-expiry revenue in our sample, as well as markets 1 standard deviation above (large market) and below (small market) the median. In the hypothetical large market (pre-expiry monthly revenue of $21.1 million), the expected number of entrants reaches 6 by 8 months after patent expiration. This means that the initial entrant's expected price falls quite rapidly (using the estimates in table 1) in such a market, and that total monthly expected generic rents begin to decline by month 5. In contrast, in the hypothetical small market (pre-expiry monthly revenue of $2.94

TABLE 4.—PREDICTED NUMBER OF GENERIC PRODUCERS AND GENERIC RENTS

Brand Name	Generic Name	Date of First Generic Entry	Actual Number of ANDAs	Predicted Number of ANDAs	Brand Sales Prior to Patent Expiration ($ millions)	Predicted Total Generic Rents ($ millions)
Alupent/Metaprel	Metaproterenol	Jan. 1988	9	3.6	6.9	1.6
Asendin	Amoxapine	Aug. 1989	2	1.5	1.9	0.9
Ativan	Lorazepam	Aug. 1985	13	13.5	9.1	4.6
Atromid-S	Clofibrate	Aug. 1986	3	2.9	1.2	1.0
Blocadren	Timolol	May 1989	6	0.4	1.2	0.2
Calan/Isoptin	Verapamil	Apr. 1986	10	10.6	6.5	3.6
Catapres	Clonidine	Jul. 1986	13	11.5	7.3	3.9
Cleocin	Clindamycin	Oct. 1987	1	1.1	1.0	0.5
Clinoril	Sulindac	Apr. 1990	6	9.0	17.0	5.4
Depakene	Valproic Acid	May 1986	4	3.0	2.9	1.0
Desyrel	Trazodone	Oct. 1986	9	10.1	5.3	3.6
Duricef/Ultracef	Cefadroxil	Mar. 1989	3	7.2	10.0	4.1
Dyazide/Maxzide	Triamterene/HCTZ	Sep. 1987	6	17.6	32.2	7.4
Feldene	Piroxicam	Apr. 1992	9	7.6	28.1	4.6
Flexeril	Cyclobenzaprine	May 1989	4	8.7	11.4	5.1
Haldol	Haloperiodol	May 1986	17	10.1	7.6	3.4
Inderal	Propranolol	Jul. 1985	18	20.7	31.0	7.0
Keflex	Cephalexin	Apr. 1987	11	16.8	25.5	6.6
Loniten	Minoxidil	May 1987	5	8.1	3.8	3.1
Ludiomil	Maprotiline	Jan. 1988	4	2.0	1.8	0.9
Minipress	Prazosin	May 1989	7	9.4	10.6	5.5
Minocin	Minocycline	Aug. 1990	3	7.5	11.5	4.5
Moduretic	Amiloride/HCTZ	Jul. 1989	6	5.3	4.7	3.2
Nalfon/Nalfon200	Fenoprofen	Aug. 1988	15	5.3	5.9	2.7
Procardia/Adalat	Nifedipine	Sep. 1990	5	10.4	48.9	6.2
Sinequan/Adapin	Doxepin	Apr. 1986	11	6.8	5.5	2.3
Tegretol	Carbamazepine	Jun. 1986	6	8.8	5.9	3.0
Tenormin	Atenolol	Jul. 1991	12	11.1	39.0	6.7
Tolectin	Tolmetin	Dec. 1991	7	2.4	5.9	1.4
Valium	Diazepam	Aug. 1985	16	14.3	22.4	4.8
Ventolin/Proventil	Albuterol	Dec. 1989	14	7.7	33.6	4.6

million), the expected number of entrants by month 8 is just over 2, and hence the aggregate generic rents continue to increase each month for the first year. The fact that the expected number of total applicants is only 4.2 in such markets means that equilibrium margins may remain high "permanently."

C. *Evaluating Policy*

One useful aspect of modeling the equilibrium relationships in generic markets is that it enables one to determine how changes in the costs or benefits of obtaining an ANDA affect the path of entry over time, and hence the path of prices over time. This enables one to simulate the likely consequences of specific policy changes.

To illustrate, consider the FDA's 1989 decision to make its review of applications more rigorous.[24] Although the more rigorous review process may have screened out some fraudulent drug applications, our estimates indicate that the change in the review process reduced the number of nonfraudulent products approved as well. Specifically, equation (5′) indicates that the fixed cost associated with obtaining approval rose from $338,000 before the scandal to $603,000 due to the scandal. In an average market, this leads to a decrease in the number of expected applicants from 9.3 firms before the scandal to 6.9 afterward.

This in turn increases the expected price at any point in time. For the average-size market, prices rise by an average of 4.9% over the 3 years. The price increase times generic expenditures ($35.2 million over 3 years for the average drug) is likely to be a close approximation of the lost consumer surplus. This amounts to $1.84 million for the average drug, which can be viewed as the cost to consumers of increased FDA vigilance against subsequent fraud.

[24] See Reiffen and Ward (2002) for additional policy simulations using our estimates.

VII. Conclusion

This paper develops a methodology for estimating the structural relationships that describe generic drug industry dynamics. These estimates enable us to describe how a market in this industry evolves from monopoly pricing toward competitive pricing. Two elements of the methodology are noteworthy. First, because the exact nature of the relationship between price and the number of competitors is critical to our estimation, the structural assumptions made about this relationship will have a large influence on our results. To minimize the possibility of misspecification, we allowed the data to determine the nature of the pricing

relationship by using a general functional form. One interesting finding from this functional form is that the negative effect of increased competition on prices continues at least until the fifth firm enters, but is not likely to be important after the eighth firm enters.

The second noteworthy element of our estimation procedure is that we use a system of simultaneous equations to estimate the relationship between entry and profitability. We do this because it is likely that the causality between the number of entrants and the available rents runs in both directions. We estimate these relationships simultaneously using functional-form restrictions that follow from economic theory to identify the system, which is then estimated using an iterative process.

Our estimates indicate that the flow of generic industry rents increases for the initial 5 to 10 months after patent expiration but then falls as more entrants compete away price-cost margins. We find that more firms enter, and enter more quickly, in markets with greater expected rents. Finally, the size and time paths of generic revenues, rents, and the number of firms are greatly affected by measures reflecting the expected market size. A consequence of these relationships is that the extent to which prices approach competitive levels in a market depends upon, among other things, the potential revenues in the market. We estimate that for markets of sufficient size (as measured by pre-expiry revenue), entry will ultimately lead to near-competitive pricing. In contrast, in small markets, prices will remain above marginal cost without inducing additional entry. Finally, this analysis suggests that even in large markets, mergers between competitors can lead to higher prices, because even the sixth or seventh entrant can have an effect on price. Moreover, such price increases may not induce entry, even if potential entrants have the same entry costs as the incumbents and entry would restore pre-merger prices. The reason is that the potential entrant knows it will be competing with $n - 1$ existing firms as soon as it enters. The expected rents from being the nth entrant are substantially less than the expected rents from being one of n firms, each with an equal likelihood of being the first approved ANDA, the second approved ANDA, and so on.

REFERENCES

American Hospital Formulary Services, *Drug Information* (Bethesda, MD, 1996).

Aronsson, T., M. Bergman, and N. Rudman "The Impact of Generic Drug Competition on Brand Name Market Shares—Evidence from Micro Data," *Review of Industrial Organization* 19 (2001), 425–435.

Berry, S., "Estimation of a Model of Entry in the Airline Industry," *Econometrica* 60 (1992), 889–917.

Berndt, E., L. Bui, D. Reilly, and G. Urban, "Information, Marketing, and Pricing in the U.S. Antiulcer Drug Market," *American Economic Review* 85 (1995), 100–105.

Bhattacharya, J., and W. Vogt, "A Simple Model of Pharmaceutical Price Dynamics," *Journal of Law and Economics* 46 (2003), 599–626.

Bond, R., and D. Lean, *Sales, Promotion and Product Differentiation in Two Prescription Markets*, (Washington, DC: Staff Report to the Federal Trade Commission, 1977).

Bresnahan, T., "Empirical Studies of Industries with Market Power," (pp. 1011–1057), in Richard Schmalensee (Ed.), *Handbook of Industrial Organization* (Elsevier Science, 1989).

Bresnahan, T., and P. Reiss, "Entry and Competition in Concentrated Markets," *Journal of Political Economy* 91, 977–109 (1991).

Caves, R., M. Whinston, and M. Hurwicz, "Patent Expiration, Entry, and Competition in the U.S. Pharmaceutical Industry," *Brookings Papers on Economic Activity: Microeconomics* (1991), 1–66.

Cook, A., *How Increased Competition from Generic Drugs Has Affected Prices and Returns in the Pharmaceutical Industry* (Washington, DC: Congressional Budget Office Study, 1998).

Demsetz, H., "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics* 16, 1–10 (1973).

DiMasi, J., R. Hansen, and H. Grabowski, "Cost of Innovation in the Pharmaceutical Industry," *Journal of Health Economics* 22, 151–185 (2003).

Ellison, G., and S. Ellison, "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration," MIT working paper (2000).

Frank, R., and D. Salkever, "Generic Entry and the Pricing of Pharmaceuticals," *Journal of Economics and Management Strategy* 6 (1997), 75–90.

Gisser, M. "Dynamic Gains and Static Losses in Oligopoly: Evidence from the Beer Industry," *Economic Inquiry* 37 (1999), 554–575.

Gort, M., and S. Klepper, "Time Paths in the Diffusion of Product Innovations," *The Economic Journal* 92 (1982), 630–653.

Grabowski, H., and J. Vernon, "Brand Loyalty, Entry, and Price Competition in Pharmaceuticals after the 1984 Drug Act," *Journal of Law and Economics* 35 (1992), 331–351.

Greene, William H., *Econometric Analysis, 5th ed.* (Upper Saddle River, NJ: Prentice-Hall, 2003).

Hausman, Jerry, "Specification Tests in Econometrics," *Econometrica* 46 (1978), 1251–1272.

Jovanovic, B., and G. MacDonald, "The Life Cycle of a Competitive Industry," *Journal of Political Economy* 102 (1994), 322–347.

Klepper, S., and K. Simons, "The Making of an Oligopoly: Survival and Technological Change in the Evolution of the U.S. Tire Industry," *Journal of Political Economy* 108 (2000), 728–760.

Lu, Z., and W. Comanor, "Strategic Pricing and New Pharmaceuticals," this *REVIEW*, 80 (1998), 108–118.

Moulton, B., "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics* 32 (1986), 385–397.

Reiffen, D., and M. Ward, "Generic Drug Industry Dynamics," Federal Trade Commission working paper no. 248 (2002).

Scott Morton, F., "Barriers to Entry, Brand Advertising, and Generic Entry in the U.S. Pharmaceutical Industry," *International Journal of Industrial Organization* 18 (2000), 1085–1104.

——— "Entry Decisions in the Generic Pharmaceutical Industry," *Rand Journal of Economics* 30 (1996), 421–440.

Wiggins, S., and R. Maness, "Price Competition in Pharmaceuticals: The Case of Anti-infectives," *Economic Inquiry* 42 (2004), 247–263.