Brewed in North America:[*]

Mergers, Marginal Costs, and Efficiency

Paul Grieco[†]   Joris Pinkse[‡]   Margaret Slade[§]

Penn State   UBC

August 2017

Abstract:

We show how production and pricing data can be used to estimate merger–related efficiencies using pre merger data, and we assess the changes in efficiency and marginal costs that are expected to occur post merger. To do this, we jointly estimate firm level returns to scale, technical change, TFP growth, and price–cost markups. We implement our empirical model using data for the North American (US and Canadian) brewing industry, and we use the estimated model of firm technology to evaluate the changes that are associated with the merger between Molson and Coors. We forecast nontrivial increases in returns to scale and declines in marginal costs and we verify those results by analyzing the impact of the merger retrospectively using post merger data.

Keywords: Mergers, Efficiencies, Marginal costs, Returns to scale, Technical change, Productivity

JEL Classifications: D22, D24, L49, L66

[†]Department of Economics, The Pennsylvania State University, and the Center for Auctions, Procurements, and Competition Policy

[‡]Corresponding author, m.slade@mail.ubc.ca, Vancouver School of Economics, The University of British Columbia

1 Introduction

Competition authorities and defendants in merger cases often use quantitative tools to forecast the impact of a proposed merger on market power and welfare. Their analysis usually proceeds in two stages: first they predict changes in prices holding marginal costs constant, and second, if competitive concerns are raised, offsetting efficiencies are evaluated by the merging parties. This two stage process is incomplete since prices and costs are jointly determined. Moreover, firms' marginal costs can change for two reasons: first, changes can be due to potential "synergies" realized by merging the firms (e.g., returns to scale) and second, they can be due to changes in the optimal level of output. However, most merger simulations ignore potential changes in the marginal cost of production.[1]

We propose a quantitative technique that can be used to forecast merger–related changes in returns to scale, marginal costs, and total factor productivity (TFP) growth and can be implemented with only pre merger data. Whereas the standard simulation model defines a merger as a consolidation of control rights — the ability to set prices — and uses estimates of demand to predict the impact of a merger on prices, we consider a merger to be a consolidation of assets and use estimates of the firms' technology to predict the impact of a merger on costs. We view this approach as an important first step to incorporating cost changes directly into merger analysis.

We use our model of production to evaluate the cost impacts of a merger between Molson and Coors that occurred in 2005 and united the second largest brewer in Canada with the third largest in the US. Over time, the Canadian and US markets had become more integrated due to trade liberalization in general and to the North American Free Trade Agreement in particular. For example, at the time of the merger, Coors Light was the largest selling light beer in Canada. The Molson Coors cross border merger rationalized production, marketing, and distribution, since the brands of both firms were already sold in both countries. Our goal is to quantify these effects. To do this, we perform ex ante forecasts using pre merger data. The advantage of using a consummated merger for this exercise is that we are able to verify our forecasts using data from subsequent periods.

Our model of technology is a standard production function in the spirit of Olley and Pakes (1996) and Ackerberg, Caves, and Frazer (2015). However, since we are interested in estimating returns to scale, we specify a more flexible functional form than the commonly used Cobb Douglas. This is important for our analysis since mergers imply a substantial change in the size of firms. Moreover, since our estimates of returns to scale vary across firms and time, technical change and TFP growth are not the same. We show how they differ and how each can be estimated.

Given data on revenues and expenditures, which are available from most standard firm level data sources, we can also estimate markups. In doing this, we make no assumptions about the way in which prices are determined. Instead, using the insight of De Loecker and Warzynski (2012), we estimate markups as wedges between output elasticities and expenditure shares. Moreover, we also use this insight to estimate marginal costs that, unlike those obtained from a demand model, do not rely on an assumption about the market game.

---

[1] Fan (2013) and Jeziorski (2014a), who model cost reductions due to mergers of local newspapers and radio stations respectively, are important exceptions. However, the applications are specialized and do not make use of data on production inputs.

As a prelude to a more detailed merger analysis, we compare efficiency, marginal costs, and markups over time and across sectors and countries.

Unfortunately, just as data on demand alone cannot be used to forecast changes in marginal costs, data on supply alone cannot be used to forecast changes in optimal output or prices. If one is interested in how prices and markups of different brands change in different geographic markets, one can combine our model of costs with a full demand model to conduct a more thorough merger analysis. We describe how these two sorts of simulation models can be solved simultaneously to obtain a complete picture of a merger.

Our approach unifies and extends three literatures: the use of Divisia indices to measure TFP growth that began with Richter (1966) and Jorgenson and Griliches (1967), the estimation of productivity in the presence of endogenous variable factors and selection bias that dates to Olley and Pakes (1996), and the calculation of markups without specifying how firms compete in product markets that was initiated by Hall (1988). Each of those ideas has been extended in many directions by numerous researchers.[2] However, by combining the three strands, we think that we provide a unifying supply–side approach to efficiency and market power measurement that can be complementary to the traditional demand–focused approach to merger analysis.

The heart of our approach is the estimation of a flexible firm–level production function. To estimate the production function, we use an unbalanced panel of publicly traded US and Canadian brewing companies between 1950 and 2012. During the first half of that period, all brewers were in the mass production sector. However, after 1980, the craft sector gradually began to grow in importance. The firms in our data are therefore heterogeneous because they produce in different countries and sectors and we hypothesize that there will be systematic differences in returns to scale and markups across those groups. Moreover, even within sectors and countries, there could be substantial cross sectional variation in efficiency and market power. We therefore estimate a translog production function that allows returns to scale and markups to vary across firms and time, and we use the estimated function to forecast efficiency changes.

We chose to use firm rather than plant data for two reasons. First, production and packaging costs account for only slightly over half of the costs of a large modern brewer. The remainder are attributable to advertising, marketing, sales, transportation, and administration, some of which are performed at the head office and others in different geographical markets. Second, if the parties to the merger were minimizing production costs at each plant prior to the merger, much of the merger related cost reduction will come from outside the plant. Finally, plant data on input expenditures and output are not readily available.

The use of standard firm level data sets such as Compustat is fraught with difficulties. For example, typically revenue from output and expenditures on inputs are recorded but not prices and quantities. We therefore constructed firm–specific input and output prices that allow us to measure physical inputs and output more accurately. In addition, with data on publicly traded firms, selection is especially important. Indeed, a firm can disappear from the data not only because it fails but also because it merges with another firm that need not be in the data (e.g., a firm that is not North American) or because it goes private. We therefore specify a model of selection that distinguishes among these possibilities.

We assume that, at least from the production point of view, beer is relatively homogeneous. Although

---

[2]De Loecker and Warzynski (2012) combine the second and third strands.

brands differ with respect to alcohol strength and flavor additives, those differences are small when compared to total production and marketing costs. In contrast, consumers have strong preferences over brands. This means that, at least to a first approximation, the product is homogeneous on the supply side but differentiated on the demand side. Our approach to efficiencies measurement is best suited for products that are relatively homogeneous in supply. Moreover, there are many other industries, such as gasoline, soft drinks, and dairy products like yogurt, in which products are homogeneous in supply but differentiated in demand. That said, other industries produce distinct products that cannot be easily aggregated. We therefore discuss how our model can be extended to incorporate multiple outputs.

Once the model has been estimated we are able to compute the relevant performance measures for each firm. We then simulate the impacts of a merger by analyzing the implications for these variables of combining the merging firms' fixed and variable inputs. We allow for multiple models of "synergies" between firms based on the merging firms' recovered technical efficiency. Finally, we verify those forecasts by comparing them to the estimated performance measures recovered from post merger data. To anticipate, our structural simulation of the Molson Coors merger forecasts that scale economies will become stronger and marginal costs will fall. When we compare our forecasts to ex post estimates, we find that, by the second year post merger, forecast efficiencies had been largely realized.

The next section, which describes the North American brewing industry, the Molson Coors merger, and the data, is followed by discussions of antitrust policy towards mergers in North America and the use of merger simulations. We then discuss related literature and present the theoretical and empirical models, and finally, our empirical findings and conclusions.

2 The North American Brewing Industry

2.1 Background and History

Brewing consists of combining hops, malt, barley, and other ingredients with water and allowing the liquid to ferment. When this has been accomplished, the fermented beverage is packaged into bottles, cans, and kegs and the packaged goods are shipped to market. Since there can be economies of scale in brewing and packaging, a medium sized brewer faces a tradeoff between having one large brewery, which involves lower production but higher shipping costs, versus several smaller breweries with lower shipping but higher production costs. In contrast, a large national brewer with several large breweries can achieve economies of scale in both production and distribution.

Brewing is the first phase in a three tier system that consists of production, distribution, and sales. In the US, with the exception of microbreweries, federal law prohibits integration of the three phases. In Canada, in contrast, distribution and sales are regulated by the provinces. However, in most provinces, the downstream phases are separate from brewing and are often handled by a provincial liquor control board. Nevertheless, in both countries, the brewer incurs the freight costs of shipping product to distribution points that are closer to markets.

Historically, at the national level the industry was relatively unconcentrated. In the US, antitrust vigilance

meant that the large national brewers grew internally. However, early mergers between small brewers led to the creation of regional brewers. Those mergers, which occurred in a shakeout that allowed the remaining firms to achieve economies of scale, were of two sorts: mergers to achieve synergies and growth (the fate of Pabst, for example) and mergers for asset stripping (the fate of, e.g., Stroh, Schlitz, and Heileman). Mergers for asset stripping, which were common during the wars of attrition in the 1970s, allowed failing firms to exit gradually. In contrast to the US, the national Canadian brewers expanded through mergers with smaller brewers. The Canadian situation was different because interprovincial trade was banned, which meant that the only way to become a national brewer was to acquire or establish a brewery in each province.

This trend towards increasing concentration has continued in recent years due to, for example, mergers with non North American brewers (e.g., Labatt/Interbrew in Canada and Anheuser Busch/Inbev (ABI) in the US). In addition, domestic mega mergers have further increased concentration in the industry post Molson Coors, most notably the Miller Coors joint venture that occurred in 2008.[3] Compared to this mega merger, the Molson Coors merger was rather small.

In the same period, the brewing industry witnessed technical changes that increased efficiency. For example, table 1 in Kerkvliet, Nebesky, Tremblay, and Tremblay (1998) contains estimates of the minimum efficient scale (MES) in brewing that were obtained by various researchers, and those estimates show how MES increased with time, particularly in the 1970s. The authors attribute that increase to the introduction of super breweries and to advances in packaging techniques, particularly in canning. In addition, improvements in shipping, such as the widespread availability of refrigerated trucks, allowed brewers to expand their geographic markets.

Between 1950 and 1980, North American brewing consisted almost entirely of mass production. However, at some time around 1980, the craft beer sector took off. The craft sector, which focuses on darker beers such as ales rather than the lagers that are the mainstay of mass production, has grown until, in 2015, it constituted 12% of the US market. Its popularity is due in part to a reaction against the light and rather flavorless beers that had come to dominate conventional brewing. During the same time period, the share of imported beers began to rise until today it is nearly 15%. Early on, most imported beers were European. Today, however, of the four most popular brands imported into the US, three come from Mexico.
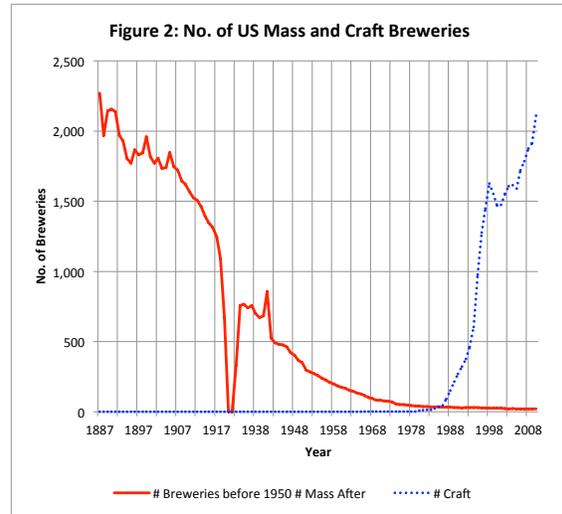
These trends in industry technology, sectoral growth, market structure, and imports are illustrated in Figures 1 through 5.[4] Figure 1 contains graphs of labor productivity — output per person — and materials intensity — materials use per unit of output — for the the North American industry as a whole. The figure shows that labor productivity increased dramatically over the entire period. Materials intensity, in contrast, fell somewhat after 1975, probably due to the rising popularity of light beer, which contains more water.

Turning to the US market, Figure 2 contains graphs of the number of US breweries in each sector from 1887 to recent times. Since the distinction between mass and craft production was not meaningful in the early years, the solid line represents the total number of breweries prior to 1950 and the number of

---

[3]The ABI Miller merger that was approved in 2016 did not affect concentration in the US, since Miller brands are now owned by Molson Coors.

[4]The data for these figures are for the industry as a whole, not our sample of firms. The sources are as follows: Figure 1, the Beer Institute and Beer Canada; 2 and 5, the Beer Institute; and 3 and 4, Weinberg. These sources are discussed in the data appendix.

Figure 1: Trends in NA Labor Productivity and Materials Intensity

Figure 2: No. of US Mass and Craft Breweries

mass or conventional breweries thereafter. The figure shows that, whereas the number of breweries (in the conventional sector after 1950) fell from over 2,000 to just 20, the number of craft breweries rose from zero to over 2,000. However, in spite of the large number of craft establishments, during the period of the data, the craft sector accounted for at most 10% of production. Finally, the sharp dip in breweries in the 1920s was due to prohibition.

Figure 3: US 4 Firm Concentration Ratio

Figure 4: Market Shares of AB, Coors, and Miller
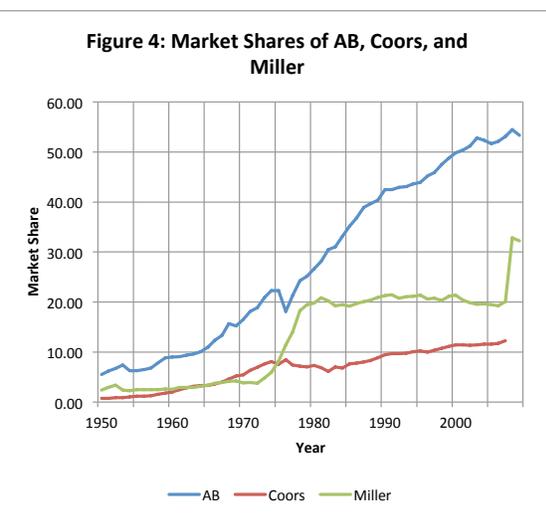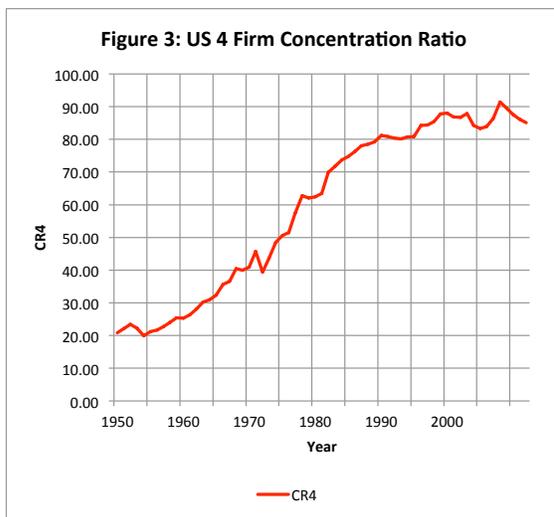
Figure 3 contains a graph of the US four firm concentration ratio between 1950 and 2012. It illustrates the dramatic upward trend in concentration over the time period of the data. Moreover, Figure 4, which graphs the market shares of three of the largest US firms, Anheuser Busch (AB), Miller, and Coors, shows that the shares of all three have increased with time.[5] Finally, figure 5 illustrates the degree of import penetration, which grew dramatically starting in the late 1970s.

---

[5]The dramatic increase in Miller's share in 2008 was due to the Miller Coors joint venture, when the US operations of Coors were deconsolidated from Molson Coors.

Figure 5: Imports Share of US Consumption
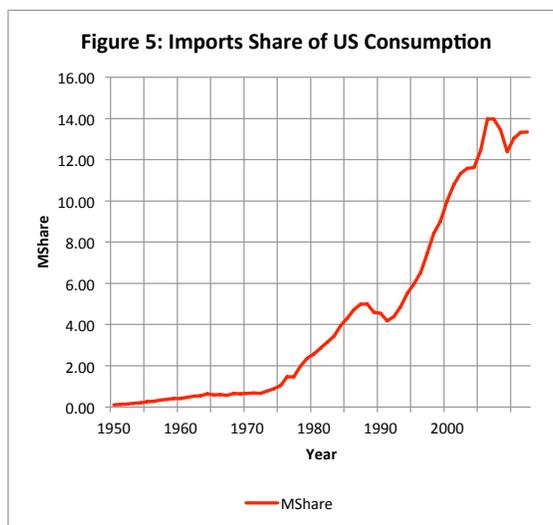
This brief snapshot of the industry suggests that industry concentration, returns to scale, and import penetration have increased over time. As a result, the market might have increasingly been characterized by both noncompetitive pricing and efficiencies associated with large size. Prior to our formal analysis of the merger, we assess those possibilities.

2.2 The Molson Coors Merger

The Molson Coors merger, which was announced in July of 2004 and consummated in early February of 2005, united the second largest brewer in Canada with the third largest in the US. Coors paid $3.5 billion to acquire Molson. However, instead of consolidating the two head offices, Canadian operations, including Canadian sales of Coors brands, remain headquartered in Montreal, Qc., whereas Coors operations, including US sales of Molson brands, continue to be headquartered in Golden, CO.

The Molson Coors merger was atypical among beer mergers in that rationalization of production was not a major motive. Indeed, Molson already produced Coors brands under license and Coors produced Molson brands. However, there were two separate entities, Coors Canada and Molson USA, that were responsible for production, distribution, and marketing in the foreign country, and those entities were eliminated by the merger.

Despite the fact that headquarters would not be consolidated, Coors' 2004 Annual Report forecast that, were the merger to occur, $175 million worth of synergies and merger–related cost savings would be realized. Furthermore, those efficiencies would principally be due to lower marketing and distribution costs, greater financial strength, facilitated geographic expansion, and increased tax benefits. In addition, the Coors brewery in Memphis, TN, would close at the end of 2006 whereas two new breweries, one in the Shenandoah Valley, VA and the other in Moncton, NB, would open in 2007. Although the brewery closure and openings have occurred as scheduled, it is not clear why they are merger–related.[6] The other efficiency claims, which we

---

[6]Our retrospective analysis of the merger makes use of data from a period that is prior to the brewery openings and closures.

7

assess, are more clearly tied to the merger.

In sum, although there were credible efficiency gains that could be expected, some of the usual gains from a merger, such as elimination of duplicate head offices, were not planned. On the other hand, there were large overlaps in markets but few overlaps in brewery locations, which are conditions that are conducive to efficiency gains.

2.3 Data

The data that we use is an unbalanced panel of North American — US and Canadian — brewers between 1950 and 2012, with much of the information coming from Compustat.[7] A firm is included in the Compustat North American data base if its shares trade on a US stock exchange. We drop those firms that are listed on US exchanges but are headquartered outside North America because they face different factor prices and economic conditions.[8] Firms with fewer than three consecutive time series observations were also eliminated. This process yielded 30 firms and 602 observations. Although our sample is not representative of the industry as a whole, it is representative of the firms that might be subject to scrutiny by competition authorities. Moreover, the sample of firms should be similar to the firms that participated in the merger.

We classify firms according to whether they are US or Canadian and whether they are mass or craft type of production. However, since all firms in the data are publicly traded, none are small microbreweries or brew pubs. For example, the craft brewer Boston Beer Company, which brews Sam Adams, is quite a large firm.

Our data consist of firm level prices and quantities of inputs — labor, materials, capital, and investment — and output — barrels of beer — for the years when a firm appears in the data.[9] Much of the Compustat data is in the form of revenues and expenditures rather than prices and quantities. Fortunately, we were able to obtain firm level data on production from various sources that are documented in appendix A, and we use that data to calculate output prices as revenue divided by production.[10]

We constructed input prices for each firm from a large number of data sources, and we use the price and expenditure data to obtain quantities. For example, an overall materials price index is calculated as a expenditure share weighted geometric mean of raw materials and packaging price indices, where the raw materials price index is calculated as an expenditure share weighted geometric mean of the prices of malt, hops, corn, rice, wheat, and sweeteners (sugar and corn syrup), and the packaging price index is constructed as an expenditure share weighted geometric mean of the prices of bottles, cans, and cartons. Furthermore, the materials input prices and shares differ by sector and country.

Since we have data on total expenditures, which, for example, includes management and workers involved in advertising, R&D, and distribution, as well as total expenditure on other inputs, our efficiency measures

---

[7]The data are described in more detail and their sources are listed in appendix A.

[8]None of the brewers with headquarters outside of North America that were dropped have a North American brewery. The products of those firms are either imported or produced under license by a North American brewer.

[9]The fact that we have prices of inputs and outputs is somewhat unusual. In fact, many researchers use revenues rather than output as the dependent variable.

[10]Revenue is price times sales, not production, where sales pertain to the brewer, not retailer. However, beer inventories are small since the period between brewing and the sell by date is quite short.

that are based on returns to scale encompass economies of density, procurement efficiencies, savings from specialization, and the elimination of duplication as the firm expands.

When estimating the production function, we found that labor and materials were highly correlated and it was not possible to disentangle their separate effects. For this reason, we created a single variable input, V (with price P_v) as an expenditure share weighted geometric mean of employment and materials. This procedure implies that the function of conventional inputs can be written as $F(V(L,M),K)$. When doing this we found that the revenue shares of three small firms were substantially greater than one, and we dropped those firms. This left us with firm-year 573 observations.

Capital, K, is our fixed input and we have assumed that V, which is a combination of employment and materials, is variable. Although it seems clear that materials are variable, the situation with labor is less obvious. Verbal communications with strategy officers of brewers led us to the following conclusions: Brewers have long term labor forces that are adequate for the minimum expected demand. Seasonal and other fluctuations are handled by hiring temporary workers. For example, since the demand for beer is largest in the summer, in that season, brewers hire students who work full or part time. Furthermore, since beer can be stored for a short period, very short term fluctuations can be handled through changes in inventories. We are therefore comfortable with the assumption that, at least with yearly observations, labor is a variable factor.

Finally, we determined the reason for exit — failure, merger for synergies, merger for asset stripping, or going private — for all exiting firms.

Table 1 contains summary statistics for the variables in the North American data. Although many of the variables do not have natural units, output and employment are well defined. Production averaged over firms and years is 11 million barrels per year and the average price is \$74 per barrel. In addition, the average wage is \$25,500 per year and average employment is nearly 5,000 people. The table also shows that there is substantial overall variation in all of the variables. In addition, although one cannot see this in the table, there is substantial cross sectional as well as time series variation in prices as well as quantities. The table also shows that approximately 20% of the observations are on Canadian firms and about 20% are on craft brewers.

3 Competition Policy Towards Mergers in North America

There are many similarities between the ways in which mergers are evaluated by competition authorities in the US and Canada. In both countries, pre merger notification is required for large transactions. Moreover, if a merger is challenged, there is a two step process in which the authority must establish the anticompetitive damage that is associated with a merger, whereas the merging parties must show all other aspects of the tradeoff, including the nature, magnitude and timeliness of efficiency gains and whether such gains are greater than and offset the anticompetitive effects. In addition, they must show that the gains are likely to occur and that they are specific to the merger. Cost savings can include economies of scale, scope, and density, and savings from specialization and the elimination of duplication.

In the US, the Department of Justice (DOJ) or Federal Trade Commission (FTC) has 30 days after the parties have filed a pre merger notification in which to decide whether to file a second request for more

information. After that, if the agency feels that there are significant competitive concerns, it can seek a preliminary injunction and, during the evaluation process, the parties must behave as arm's length competitors. The authorities make every effort to ensure a speedy decision and most litigated mergers are decided on the basis of preliminary injunctions rather than trials.

The 2010 US Horizontal Merger Guidelines prohibit a merger if "the effect of such acquisition may be substantially to lessen competition." This language does not mention efficiencies and makes it difficult to mount an efficiencies defense (Blair and Haynes, 2011). Nevertheless, merger practice in the US has gradually become more sympathetic towards considering countervailing factors (Kolasky and Dick, 2003). In practice, if the agency decides that there will be competitive harm, the merging parties can seek to establish that there are offsetting efficiencies. Nevertheless, US law has adopted a consumer welfare standard in which efficiencies can cancel anticompetitive effects only if they are likely to reverse the merger's competitive harm to customers. In practice this means that they must offset the price increases, and cost savings per se are given zero weight.

In contrast to the US, the Canadian Competition Act act adopts a total welfare standard that balances efficiency gains against anticompetitive effects. In other words, whereas the US gives zero weight to cost savings per se, Canadian authorities take them into account. Nevertheless, the standard of proof is higher for efficiencies.

The agencies have in general favored narrow geographic markets due to their asymmetric treatment of costs and benefits. In particular, in the US, if efficiency is given any weight at all, it is given less weight than harm.[11] This means that the agencies focus on consumers, demand, and substitution possibilities, and clearly consumers in New York do not purchase beer in California. However, we focus on firms and their technologies. In particular, we wish to see if mergers change returns to scale, and scale economies are inherently at the level of the firm. Moreover, brewers produce all of their brands in just a few breweries — Coors had just two — and coordinate production nationally. Finally, broad pricing strategies are centrally determined.

4 Related Literature

4.1 Merger Analysis

4.1.1 Merger Simulations

Competition authorities in both countries must assess anticompetitive effects and, in many instances, they have done this with the aid of merger simulations that involve only pre merger data, (e.g., Nevo, 2000; Pinkse and Slade, 2004; Ivaldi and Verboven, 2005). Since these are the sorts of tools that competition authorities commonly use, we describe them in detail before discussing simulations that require data that is not usually available.

A typical merger price simulation is a model of firms that produce a number of differentiated substitute products and engage in price competition.[12] Firms or players choose the prices of the products that they

---

[11]For more on asymmetric merger standards, see Crane (2014).

[12]In contrast to these price simulations, Jeziorski (2014b) assesses mergers where advertising is assumed to be the choice variable.

Table 1: Summary Statistics

Variable	Description	Mean	St Dev	Minimum	Maximum
Y	Beer production	11.19	23.31	0.005	132.5
PY	Beer price	74.33	69.08	13.19	319.7
EMP	Employment	4.83	7.65	0.006	46.61
WAGE	Average wage	25.50	20.61	2.64	93.88
MAT	Materials use	656.3	1383.3	0.580	7933.7
PMAT	Materials price	0.717	0.474	0.275	2.14
V	Variable input	169.7	172.2	3.56	981.0
PV	Price of V	1.30	0.655	0.539	8.29
I	Investment	9.19	20.84	0.001	124.78
K	Capital stock	62.36	155.2	0.050	875.1
KPRICE	Price of capital	0.576	0.397	0.175	1.54
DC	Indicator for Canada	0.197	0.398	0	1
DS	Indicator for craft	0.190	0.393	0	1

573 observations

Table 2: Summary Statistics for US Market

Variable	Description	Mean	St Dev	Minimum	Maximum
CR4	US 4 Firm Concentration Ratio	0.455	0.209	0.199	0.880
MShare	Firm's Share of US Production	0.072	0.112	0.0014	0.531
IShare	Import's Share of US Consumption	0.019	0.027	0.001	0.140

382 observations

own, taking into account the choices of other players. A market structure is a partition of the product space where the *j*th element of the partition is the set of products that the *j*th firm owns. A merger is modeled as a reduction in the number of players — a coarser partition. In particular, the products that were produced by two firms are now produced by one, and a single player can choose their prices. Since firms internalize the increase in demand for their own products that result from their own price increases, but do not internalize the comparable increase in demand for rival products, the merged firm has a unilateral incentive to raise prices. The magnitude of this incentive, however, depends on substitutability among products, and accurate price forecasts depend on accurate estimates of own and cross price elasticities.

In order to implement a simulation, one must estimate the demand and marginal cost for each of the N products.[13] Typically, marginal costs are not estimated using input data, but are instead derived from estimates of the demand function and the optimality conditions of some assumed model of competition (e.g., Nash–Bertrand pricing).[14] In other words, they are the costs that rationalize the assumptions on demand and equilibrium. The merger simulation then consists of altering the merged firms' objectives to reflect the fact that merged firms will maximize joint (not individual) profits and solving the resulting first order conditions for product prices while holding marginal costs for each product at their recovered level.

Static equilibrium simulations are sensitive to a number of factors including the equilibrium assumption, the specifications of demand, and the assumed model of competition. Some researchers have tested their equilibrium assumptions and have not rejected them for the markets studied (e.g., Nevo, 2001; Slade, 2004b). Unfortunately, however, estimated demand elasticities and marginal costs can be very sensitive to the specification of the demand and cost functions, as shown in Slade (2009). Finally and most importantly for our purposes, with simulations that are commonly used by competition authorities, costs either do not change post merger or they are exogenously varied. In other words, efficiencies are not considered—or are considered only secondarily—and price changes are therefore apt to be overestimated.

4.1.2 Retrospective Merger Analysis

Some researchers have estimated merger models that require data of the sort that is not usually available to competition authorities, either because one must observe a large number of changes in assets (i.e., mergers) or because post merger data is essential. Nevertheless, that literature is related because it attempts to assess merger–related competitive harm and/or efficiencies.

A number of interesting things can be done with data that include both pre and post merger periods. For example, Miller and Weinberg (2017) assume that pre merger competition is Bertrand but introduce a conduct parameter post merger, which allows them to test if the merger facilitated tacit collusion (coordinated effects). Although models with conduct parameters were heavily criticized (e.g., Makowski, 1987; Corts, 1999), recent theoretical work has argued that they can be rationalized by a particular dynamic equilibrium concept (Black,

---

However, the mechanics of the model are similar.

[13]A number of demand specifications have been used. For example, Nevo (2000) and Jeziorski (2014b) estimate the random coefficients model of Berry, Levinsohn, and Pakes (1995); Pinkse and Slade (2004) use the distance metric model of Pinkse, Slade, and Brett (2002); and Ivaldi and Verboven (2005) use a nested logit.

[14]Marginal costs are usually specified as constant parameters or functions of observable cost shifters in the first order conditions.

Crawford, Lu, and White, 2004). In addition, Miller and Weinberg allow the parameters of the marginal cost function to vary with the merger, which allows them to evaluate short–run efficiencies.

In contrast, Jeziorski (2014a) looks at merger–related efficiencies both long and short run. Like Jeziorski (2014b), the stage game is Nash in advertising. However, the one–shot game is embedded in a model of dynamic discrete choice and inference is based on revealed preference as in Bajari, Benkard, and Levin (2007). Specifically, when the model predicts a merger that does not occur, it is assumed that estimated efficiencies are too large, and when a merger is not predicted but occurs, it is assumed that efficiencies are too small. In this way, both marginal and fixed costs can be uncovered. Unfortunately, estimation requires data on a very large number of mergers.

4.2 Productivity Measurement

There is a large macroeconomic literature that is devoted to estimating aggregate productivity growth and much of that research was inspired by the work of Solow (1957). Typically, researchers start with a constant returns to scale Cobb Douglas production function and equate productivity growth with the percentage change in output minus a share weighted percentage change in inputs — the Solow residual. Moreover, it is common to estimate the production function by OLS, which makes the exercise very simple.

The potential bias in OLS estimates of production functions, however, has long been recognized (see e.g., Marschak and Andrews, 1944). This bias results from the possible correlation between input levels and firm level productivity shocks. Specifically, when firms experience a large productivity shock, they might respond by using more variable inputs. Applied economists have devised alternatives to OLS that attempt to circumvent this problem. Most use either a variant of the method developed by Olley and Pakes (1996) (OP) and extended by Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015), and Gandhi, Navarro, and Rivers (2016) or the GMM methods proposed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (2000). We have chosen to focus on the former since those methods are based on a full behavioral model that uncovers unobserved productivity as a function of observed decisions. Specifically, they employ an inverse function of input choices to control for unobserved productivity in the production function and to overcome the endogeneity problem associated with OLS estimation that is discussed above. Since this literature is summarized in Ackerberg, Benkard, Berry, and Pakes (2007), we do not discuss it in detail here.

Most researchers in the OP tradition follow the macroeconomic literature and assume the production function is Cobb Douglas. However, even within narrowly defined industries, output elasticities can be heterogeneous across firms, which is ruled out in the Cobb Douglas case as long as firm heterogeneity is modeled as Hicks–neutral technology differences.[15] As a result, when estimating a Cobb Douglas production function, one can equate productivity growth with technical change — a shift in the production function holding inputs constant.[16] However, when the technology is flexible, the growth of TFP also depends on

---

[15]Recently, Kasahara, Schrimpf, and Suzuki (2015), Zhang (2015), Balat, Brambilla, and Sasaki (2016), and Hoderlein, Fox, Hadad, and Petrin (2016) have considered modeling non–Hicks neutral sources of unobserved heterogeneity, although this is outside the scope of our paper.

[16]A Cobb Douglas can exhibit nonconstant returns to scale. However, with that function, scale economies are the same for all

economies of scale, which can vary by firm and over time. We therefore borrow from the index–number productivity–measurement literature to derive a measure of TFP growth that captures both technical change and changes in returns to scale. Diewert and Nakamura (2007) survey the index–number literature.

4.3 Markup Measurement using Production Data

In an early paper, Hall (1988) demonstrates that markups can be estimated from a production function. His research relaxes one of the frequently used TFP growth assumptions — competitive output markets — but maintains another — constant returns to scale. Under constant returns, revenue shares are equal to cost shares, which sum to one. It is therefore possible to divide both sides of the production function by the single fixed factor, K,[17] leaving the variable inputs on the right–hand–side, which means that it is not necessary to measure the user cost of capital. In general, however, noncompetitive pricing causes the Solow residual to deviate from the rate of growth of TFP. In particular, when price exceeds marginal cost, input growth is associated with disproportional output growth. Hall uses this insight to devise a test for deviations from competitive pricing and to show how average markups can be estimated.

De Loecker and Warzynski (2012) extend Hall's method by relaxing the assumption of constant returns to scale. Their insight is that the output elasticity of a variable input equals its share of revenue only when price equals marginal cost. Furthermore, under imperfect competition, the two are equal if revenue is evaluated using the shadow price of output — marginal cost (MC) — instead of the market price (P_y). Imperfect competition therefore, drives a wedge between the two that determines the markup, P_y/MC. in addition, instead of implementing their calculations in a Solow framework like Hall, they formulate a model of production that is based on the model of Olley and Pakes (1996) and its extensions, thus addressing the endogeneity problem discussed above. This extension allows them to recover firm and time–specific markups rather than industry averages.

5 Model

5.1 The Production Function

5.1.1 Specification

We adopt a fairly standard Olley and Pakes (1996) framework and suppose that a vector of variable inputs X and a fixed input K are used to produce a homogeneous output Y according to the production function,[18]

$$Y_{jt} = A_{jt}F_{s(j)c(j)}(X_{jt},K_{jt})e^{\eta_{jt}}, \tag{1}$$

firms and therefore the correction would not change productivity comparisons.

[17]Hall (1988), like most others in the markup literature, assumes that there is a single fixed factor.

[18]We use X as a vector of variable inputs in this section, despite the fact that in our application we use a scalar variable input V, to highlight the fact that our method easily extends beyond a single variable input. We restrict the application to a scalar variable input only due to collinearity in our dataset.

where j is a firm, t is a year, A is the state of technology, F is a function of the conventional inputs, the subscripts $s(j)$ and $c(j)$ are the sector and country to which j belongs, and η is a shock that is conditionally mean independent of current and past inputs and the current state of technology. In what follows, we drop the s and c subscripts from the production and investment functions.

Equation (1) can then be written as[19]

$$y_{jt} = a_{jt} + f(x_{jt}, k_{jt}) + \eta_{jt}, \qquad a_{jt} = \beta_0 + \beta_t t + \omega_{jt}, \tag{2}$$

where all variables are in natural logarithms and the state of technology consists of a constant, a trend, and unobserved productivity, ω. In addition, we assume that the state variables ω and k evolve according to

$$\omega_{jt} = g(\omega_{j,t-1}) + \xi_{jt}, \qquad k_{jt} = (1 - \delta_{j,t-1})k_{j,t-1} + i_{j,t-1}, \qquad i_{jt} = i(k_{jt}, \omega_{jt}, t). \tag{3}$$

In other words, ω is a first order Markov process, whereas k decays at the depreciation rate δ and is augmented by investment i, which is a function of the state.

Ideally, to capture changes in market structure and other market level variables, the investment function would include year fixed effects that are interacted with the state. However, with a long time series that is not feasible. Instead, we include a trend, t and assume investment is smooth in t. Figure 3 shows that this is not a bad approximation. Indeed, market concentration increases over time in a rather smooth fashion. In addition, the investment function could include factor prices, and we experiment with that formulation.

Finally, we make a strict monotonicity assumption on the investment function — that i is monotonic in the unobservable ω — which implies that the investment function can be inverted and we can write[20]

$$\omega_{jt} = h(k_{jt}, i_{jt}, t). \tag{4}$$

To anticipate, there are two equations that are estimated:

$$y_{jt} = \beta_0 + \beta_t t + f(x_{jt}, k_{jt}) + h(k_{jt}, i_{jt}, t) + \eta_{jt}$$

and

$$y_{jt} = \beta_0 + \beta_t t + f(x_{jt}, k_{jt}) + g[h(k_{j,t-1}, i_{j,t-1}, t-1)] + u_{jt},$$

where $u_{jt} = \xi_{jt} + \eta_{jt}$. We adopt the Ackerberg, Caves, and Frazer (2015) framework and do not use the first equation to identify any production function parameters. In practice, we estimate the two equations jointly via GMM as suggested in Wooldridge (2009). Details of the estimation are discussed in section 6.2. Since our dataset is unbalanced, we will also control for selection of firms due to exit. Our procedure is a straightforward extension of Olley and Pakes (1996) that accounts for the fact that some firms exit due to mergers and acquisition, instead of failure; we postpone the details of this procedure to Section 6.2.

---

[19] It is not always possible to write (1) as (2). However, it is possible with the production functions that are commonly used.

[20] In a somewhat different context, Pakes (1994) derives monotonicity as an outcome of optimizing behavior.

5.1.2 Identification

Ackerberg, Caves, and Frazer (2015) and Gandhi, Navarro, and Rivers (2016) have raised identification concerns about both the Olley and Pakes (1996) and Levinsohn and Petrin (2003) approaches.[21] Gandhi, Navarro, and Rivers (2016) propose one solution to the problem: they obtain identification by adding a variable input demand equation to the model. However, they implicitly assume that firms are price takers in output markets, an assumption that is violated in our framework.[22]

Both of their concerns are related to the fact that standard assumptions made in this literature suggest that the model variables $x_{jt}, k_{jt}, i_{jt}, k_{j,t+1}$ are functions of a lower–dimensional subspace. We make slightly different assumptions. First, we assume that managers choose variable inputs x_{jt} prior to choosing investment i_{jt}, and that they are imperfectly informed of their productivity when making the variable input choice. Specifically, managers observe ω_{jt}^* when choosing variable inputs, while they observe $\omega_{jt} = E(\omega_{jt}|\omega_{jt}^*) + \upsilon_{jt}$ when choosing investment. We further assume that υ_{jt} is independent of the firms' observable variables when choosing variable inputs. This means that x_{jt} is a function of $(k_{jt}, t, \omega_{jt}^*)$ and not (k_{jt}, t, ω_{jt}), which addresses the identification concern pertaining to the choice of x_{jt}.[23] This assumption is reasonable for the beer industry where variable input decisions are likely to be taken at much higher frequency than investment decisions, so the information set is unlikely to be identical for each choice.

We further assume that not all investments are equally successful in the sense that capital accumulates according to $k_{j,t+1} = (1 - \delta_t)k_{jt} + i_{jt} + \varepsilon_{jt}$, where the noise term is idiosyncratic.[24] Note that with this admittedly strong assumption i_{jt} is a strictly monotonic function of ω_{jt} for any given k_{jt}, but $k_{j,t+1}$ is now a function of k_{jt}, ω_{jt} and noise. With these additional assumptions, none of the four aforementioned variables is a function of the others.

Lastly, both supply and demand simulations identify the effects of mergers without observing mergers. Our structural model of production relies on functional form and variation in input usage to predict output at counterfactual inputs. Demand simulations, analogously, rely on functional form and variation in product characteristics to be able to predict demand elasticities for products at counterfactual prices. It is therefore important to have observations on firms that are similar to the merged firm in order to avoid out of sample predictions.

5.2 The Performance Measures

Given estimates of the production function, we can construct five performance measures: returns to scale (RTS), technical change (TECH), the rate of growth of total factor productivity (TFPG), marginal cost (MC), and the price/cost markup (PCM). Three are fairly standard, and we devote more attention to the TFP growth

---

[21]Gandhi, Navarro, and Rivers (2016, appendix A), contains a detailed, clear, and elaborate description of the two main concerns.

[22]Their method could be extended to the case where firms face a log–linear demand curve, but this would imply uniform markups in the industry and so is not ideal for analysis of mergers.

[23]Alternatively, we could assume that x_{jt} is not chosen optimally due to some frictions or optimization error. However, this error would contaminate our estimation of efficiencies, marginal costs and markups below. These alternatives are discussed in the working paper version of Ackerberg, Caves, and Frazer (2015).

[24]Comparing this equation to (3), we see that the idiosyncratic depreciation rate in the latter is $\delta_{j,t-1} = \delta_{t-1} + \varepsilon_{j,t-1}/k_{j,t-1}$.

16

and marginal cost, which are less so.

5.2.1 Returns to Scale

Returns to scale are defined as the proportionate change in output that is due to an equiproportionate change in all inputs. If we define $Z = (X^T, K^T)^T$, z the vector of logarithms of Z, and $\tilde{f}(\lambda, Z) = \log F(\lambda Z)$, then our measure of *local* returns to scale is

$$\text{RTS}_{jt} = \frac{\partial \tilde{f}}{\partial \lambda}(1, Z_{jt}) = \iota' \frac{\partial f}{\partial z}(z_{jt}), \tag{5}$$

where ι is a vector of ones and the proportionate change in inputs is $\lambda - 1$. Alternatively, returns to scale is the sum of the output elasticities. With a Cobb Douglas production function, RTS is constant across firms and time. When the technology is flexible, however, RTS varies with Z.

Returns to scale is normally a long–run concept in the sense that all inputs must vary proportionally. Given time to build, proportional growth is not likely within a single period. However, a merger changes K through acquisition, not investment, and we can evaluate the consequences for RTS.

5.2.2 Technical Change

Technical change is a temporal shift in the production function, holding all inputs constant. With the production function (2), technical change is

$$\text{TECH}_{jt} = a_{jt} - a_{jt-1} = \Delta a_{jt} = \beta_t + \omega_{jt} - \omega_{jt-1} = \beta_t + \Delta \omega_{jt}, \tag{6}$$

which consists of a deterministic and a stochastic component.

5.2.3 TFP Growth

Total factor productivity growth is defined as the rate of growth of output that is not attributable to the rate of growth of inputs. There are many ways to calculate the rate of growth of inputs. We borrow from the index–number literature that measures TFP growth as an expenditure share weighted sum of the rates of growth of individual inputs.

Jorgenson and Griliches (1967) consider the problem under the assumptions of constant returns to scale and competitive pricing in input and output markets. They define the rate of growth of inputs as $\text{d}z/\text{d}t = \sum_k r_{kt} \, \text{d}z_{kt}/\text{d}t$, where r_k is the kth input's share of revenue,[25] and show that TFP growth is $\text{d}y/\text{d}t - \text{d}z/\text{d}t = \text{d}a/\text{d}t$. That is, the rate of growth of TFP is equal to the rate of technical change. However, we assume neither constant returns nor competitive output pricing, so the two will, in general, be different.

In the context of a cost function, Denny, Fuss, and Waverman (1981) relax the assumptions of constant

---

[25]Under the Jorgenson and Griliches assumptions, expenditure and revenue shares are equal.

returns and competitive pricing in output markets and show that the growth in TFP is

$$\frac{\mathrm{d}\log \mathrm{TFP}_C}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t} + \left(1 - \gamma_{cy}\right)\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t} + \left(1 - \frac{1}{\mathrm{RTS}_C}\right)\frac{\mathrm{d}y}{\mathrm{d}t}, \tag{7}$$

where

$$\gamma_{cy} = \frac{\partial \log C}{\partial \log Y} \qquad \text{and} \qquad \mathrm{RTS}_C = \frac{1}{\gamma_{cy}}.$$

With this formulation, TFP growth consists of not only technical change but also of a term that depends on returns to scale. For example, under increasing returns, $\mathrm{RTS}_C > 1$ and the rate of growth of productivity is greater than the rate of technical change as firms realize scale economies. Indeed, TFP growth is due to both a shift in the production function and a movement along that function.

The expression in (7) depends on economies of scale that are obtained from a cost function and we have measures from a production function. Unfortunately, the two are not the same. In particular, scale economies obtained from a production function are based on the assumption that all inputs increase at the same rate whereas those obtained from a cost function are based on the assumption that inputs increase along the expansion, or cost minimizing, path. Nevertheless, Caves, Christensen, and Swanson (1981) show that, at any point (i.e., when one evaluates derivatives, not discrete changes) the two measures are equal, which means that we can substitute RTS for RTS_C in (7).

Equation (7) involves infinitesimal changes and we have discrete changes. Our measure of TFP growth uses actual changes across periods rather that equiproportionate changes or changes along cost–minimizing paths (7),

$$\mathrm{TFPG}_{jt} = \Delta a_{jt} + \left[1 - \frac{1}{(\mathrm{RTS}_{jt} + \mathrm{RTS}_{jt-1})/2}\right]\Delta y_{jt}. \tag{8}$$

This can be viewed as a discrete time approximation of (7).

We use changes in TFP growth in response to an event such as a merger as our measure of dynamic efficiency gain, since TFP growth encompasses technical change.

5.2.4 Marginal Costs

Marginal costs are a key input to merger price simulations. Moreover, the change in marginal cost is used as the measure of efficiency gains in the calculation of upward pricing pressure of a merger (Farrell and Shapiro, 2010). Therefore, it is useful to point out that, in contrast to costs obtained from a demand simulation, our model can recover estimates of marginal costs that do *not* rely on assumptions about competition. We only assume that firms are price takers in input markets and that they minimize cost.

Suppose that the production function is $Y = \tilde{F}(X, K, t, \omega)$, and let the price of some variable input X_i be P_i. Firm and time subscripts have been dropped to simplify the exposition. The firm's cost minimizing problem is to choose variable inputs to minimize

$$\sum_i P_i X_i \quad \text{s.t.} \quad \tilde{F}(X, K, t, \omega) = 0.$$

18

If λ is the multiplier on the feasibility constraint then the i'th first order condition is

$$P_i - \lambda \frac{\partial Y}{\partial X_i} \equiv P_i - \lambda \tilde{F}_i = 0.$$

Multiplying both sides by the input output ratio yields

$$\frac{P_i X_i}{Y} = \lambda \tilde{F}_i \frac{X_i}{Y} = \lambda \phi_{yi},$$

where ϕ_{yi} is the output elasticity of the i'th variable input. Finally, λ is the shadow price of output, which is marginal cost. Substituting yields

$$\text{MC}_{jt} = \frac{P_{it} X_{ijt}}{\phi_{yijt} Y_{jt}}. \tag{9}$$

Equation (9) shows that ϕ_{yi} is X_i's expenditure share when it is evaluated at the shadow price of output, MC, not its market price, P_y. In a competitive industry, the two are equal. Under imperfect competition, in contrast, the two diverge. We use this relationship to assess marginal cost changes from pre to post merger.

Equation (9) yields a single marginal cost for the firm, not one for each plant. However, if firms are minimizing total cost, which includes not only production and packaging but also distribution and sales cost, they will equate marginal cost across plants.

5.2.5 The Price Cost Markup

Following De Loecker and Warzynski (2012), we *define* the markup (PMC) to be output price P_y divided by marginal cost, MC, and we use their method to measure the markup,

$$\text{PCM}_{jt} \equiv \frac{P_{yjt}}{\text{MC}_{jt}} = \phi_{yijt} \frac{P_{yjt} Y_{jt}}{P_{ijt} X_{ijt}} = \frac{\phi_{yijt}}{r_{ijt}}. \tag{10}$$

In other words, the price cost markup equals the elasticity of output with respect to the variable input divided by the input's share of revenue, r_i.

We use PCM as a measure of market power as it is a proxy for the related noncompetitive distortion. Given that we are using a firm–level output price and estimating a firm–level production function, this markup is best understood as a firm–level average. Clearly, markups can vary by location if there is either significant variation in transport costs or substantial variation in demand elasticities across geography. However, as noted above, if firms are minimizing total cost including distribution and sales cost, they will equate marginal cost across plants. Moreover, since we are interested in the firm, using the average price — a volume weighted average of the prices that the firm sets — in our markup calculation seems appropriate.

5.3 Forecasting Efficiency Gains for Mergers

We model a merger as a consolidation of assets, both variable and fixed. Our goal is to determine the immediate impact of the merger on firm efficiency. There can also be long–run effects of the merger as firms adjust

their investment and input demand strategies to react to the new competitive environment, however we do not attempt to capture these as they would require a complete model of demand and investment costs.

Suppose that we have estimated the production function (2) using a panel of firms that includes, but generally is not limited to, the merging firms. Suppose further that a merger between i and j is proposed in year t. Let the inputs of firm ℓ in period t be $Z_{\ell t} = (X_{\ell t}^T, K_{\ell t}^T)^T$, $\ell = i, j$. Then the merged firm m will inherit the inputs Z_{mt} equal to the sum of the inputs of the parties to the merger.

Although the merged firm inherits the inputs of its predecessors, it will not produce the sum of the predecessors' outputs. Indeed, if there are increasing returns, the merged firm will produce more compared to the pre merger situation. There are many reasons why this might be the case. For example, the merger could facilitate geographic penetration and it could mean that advertising and marketing efforts are more effective, and both factors would cause sales to rise.[26] However, we do not have a model of within firm allocation of inputs nor do we forecast changes in the effectiveness of each input class. Instead, we assume that the new firm, like a firm with inputs that are similar to Z_m, will use the inputs more efficiently. With our application, this could include reallocation of L and M within the firm as well as reallocation between L and M, holding V_m fixed.[27]

We must also deal with how the merger affects technical efficiency. We consider two measures of ω_m. The first, $\bar{\omega}$ is the output share weighted average of ω_i and ω_j, which is appropriate if technical know–how is difficult to transfer across firms.[28] The second, $\omega_{\max}$, is the maximum of ω_i and ω_j, which is appropriate if know–how is fully transferable. We expect the actual ω to lie between the two. We do not forecast the evolution of ω since the question that we ask is what would happen if the merger were to occur today — at the time of the announcement.

Given these assumptions, one can reevaluate equations (5), (6), (8), and (9) using the new inputs and technical efficiency to obtain forecasts of the performance of the merged firm. In addition, our formula for the overall forecast of static efficiency gains of the merger is

$$\text{EFF}_{\text{forecast}} = \frac{\hat{Y}_m}{Y_i + Y_j} - 1, \tag{11}$$

where $\hat{Y}_m = F(Z_m)e^{\omega_m} = e^{f(Z_m) + \omega_m}$.

This is our baseline forecast of efficiency gains. We chose to focus on (11) since it is feasible and the merged firm inherits the inputs of each of its constituents and can choose to use them. Nevertheless, there are several reasons why this estimate might be biased. First, the new firm might find it difficult to coordinate operations such as information systems, or more generally to optimize within firm organization.[29] When this occurs, the firm will be inside the production function, merged output will be less than forecast, and merged marginal costs will be higher. Second, the new firm might choose to reduce X and produce a smaller Y. If this

---

[26]Certainly, merging parties often claim that this is the case.

[27]If K_m is fixed and the marginal product of V is positive, V_m is the cost minimizing V for Y_m.

[28]Suppose, for example, that all plants are the same size and firm i has n_i plants, $i = 1, 2$. Then $\bar{\omega} = (n_1/n)\omega_1 + (n_2/n)\omega_2$, where $n = n_1 + n_2$. In other words, each plant uses its old technology, and $\bar{\omega}$ is the output weighted average.

[29]For a discussion of within firm maximization after a merger, see Michel (2016).

occurs, it is not possible to predict the direction of the cost change. Indeed, on the one hand, a lower X will lower marginal cost. On the other hand, however, a lower Y will raise cost through a reduction in economies of scale. Third, if K_m is greater than the costs minimizing K for $\hat{Y}_m$ and the firm is able to sell some assets fairly costlessly, it could produce $\hat{Y}_m$ at minimum cost. When this is possible, costs will be lower than forecast. Since it is not possible to sign the direction of the bias, and since the three effects could offset one another, we have chosen (11) as our measure of efficiency gain.

An obvious concern with this method is that demand effects of the merger can lead the merged firm to restrict output and therefore Z_m will be less than $Z_i + Z_j$. Ideally, our approach could be coupled with a demand estimation so that both demand and supply side effects of a merger could be simulated jointly. If one expects that prices will rise and output will be curtailed, it is straight forward to perform a sensitivity analysis that evaluates the efficiency consequences of output reductions of different magnitudes. This is similar to what Hausman, Leonard, and Zona (1994) do with a merger price simulation when they evaluate the pricing consequences of efficiency gains (marginal cost reductions) of 0, 5, and 10 percent.

Our approach also assumes that firms are price takers in input markets, which means that the merger does not affect the demand functions for variable inputs.[30] Inputs to brewing are mostly agricultural and energy products that are sold in world markets and, in addition, are not specialized to brewing. Although the market for hops is worldwide, it is used almost exclusively in brewing. However, the cost of hops is about 1% of total brewing cost.[31]

We also make the standard assumption that the merging parties are exogenously given, an assumption that reflects the problem that competition authorities face. In particular, they are concerned with analyzing the effects of a given merger, not predicting which mergers will occur.

Finally, in order to avoid out of sample extrapolation, one should have data on firms that are larger than the merging parties. Indeed, we identify the effects of a merger by assuming that the merged firm will look like firms of similar size in the data. This would not be possible if, for example, the two largest firms were to propose a merger. However, it is unlikely that such a merger would be allowed without requiring substantial divestitures, and divestitures could be evaluated.

With post merger data, we can observe actual changes in production and input usage. Our ex post measure of efficiency gain is therefore the percentage change in output minus the share weighted percentage change in inputs,

$$\text{EFF}_{\text{realized}} = \frac{Y_m}{Y_i + Y_j} - \sum_k s_k \frac{Z_{mk}}{Z_{ik} + Z_{jk}}, \tag{12}$$

where i and j are pre, m is post merger, and s_k is the average of input k's pre and post expenditure shares. We use this measure to verify our forecast efficiency gains. Note that, because this measure includes post merger input use, this verification assesses the reasonableness of our approximation of variable input use with the sum of pre-merger input use in the pre merger forecasts.

---

[30]With some mergers, such as mergers between hospitals, bargaining over input prices is crucial. However, one would not want to use our efficiency model in those circumstances.

[31]If the firm enjoys some market power in the input market, merger may enhance that market power and result in lower input prices. If monopsony power were a main feature of a merger, one could consider explicitly modeling a bargaining game between firms and upstream suppliers (e.g., Gowrisankaran, Nevo, and Town, 2015), however, this is beyond the scope of the current paper.

5.4 Extensions

5.4.1 Multiple Outputs

As we have noted earlier, our simulation model is suitable for products that are relatively homogeneous in supply and can therefore be easily aggregated to obtain a single output. Many industries, however, produce distinct products, not just brands of a similar product, and we discuss how our model can be extended to encompass that situation.

Suppose that the firm produces multiple outputs so that Y is a vector. With multiple outputs, one can replace the production function $Y - \tilde{F}(X, K, t, \omega) = 0$ in subsection 5.2.4 with the transformation function, $T(Y, X, K, t, \omega) = 0$. In particular, $T(.) = 0$ implies that X can produce Y and that Y is efficient (it is on the transformation frontier).

One can obtain product specific marginal costs in a fashion that is similar to the single–output case. The new problem is to choose X to minimize

$$\sum_i P_i X_i \quad \text{s.t.} \quad T(Y, X, K, t, \omega) = 0.$$

There are now K^y first order conditions for a given X_i, where K^y is the dimension of Y. Specifically, we consider the effect of variation in X_i on Y_k, holding everything else constant. The kth first order condition is

$$P_i - \lambda_k \frac{\partial Y_k}{\partial X_i} = P_i - \lambda_k \frac{\partial T / \partial Y_k}{\partial T / \partial X_i} = 0, \tag{13}$$

where λ_k is the shadow price of the kth output.

Equation (13) can be interpreted as follows. The partial derivative is the extra amount of Y_k that can be obtained from one extra unit of X_i, which is multiplied by the shadow price of Y_k to obtain its value. The equation therefore says that the firm sets the marginal cost of X_i, P_i, equal to its marginal benefit.

Multiplying and rearranging, we have

$$\frac{P_i X_i}{Y_k} = \lambda_k \left(\frac{\partial T / \partial Y_k}{\partial T / \partial X_i} \right) \frac{X_i}{Y_k} = \lambda_k \phi_{ki},$$

where ϕ_{ki} is the elasticity of Y_k with respect to X_i. Finally, the shadow price of Y_k is MC_k, which implies that

$$\text{MC}_{kjt} = \frac{P_{it} X_{ijt}}{\phi_{kijt} Y_{kjt}}.$$

Estimation requires choosing a functional form for T and a stochastic specification. Although not necessary, we suggest assuming that the transformation function is separable in inputs and outputs so that it takes the form $T(Y, X, K, t, \omega) = T^y(Y) - T^x(X, K, t.\omega) = 0$. Note that T^x is the same as the single output production function $\tilde{F}$ above. We simply replace the single output with the aggregator function $T^y(Y)$.[32] In addition, we do not change the assumptions concerning the evolution of the state variables. Finally, as with

---

[32] A natural extension of the production function that is specified in subsection 6.1 is to assume that both T^y and T^x are translogs.

the production function, one can append an error, η, that is conditionally mean independent of current inputs and past inputs and outputs. The empirical specification is then

$$T^y(Y) - T^x(X, K, t, \omega) = \eta.$$

One can estimate the multi–output transformation function using one or two–step GMM. However, additional instruments are required to correct for the endogeneity of the outputs. Furthermore, one must impose linear homogeneity on T^y.[33]

Finally, one can obtain the efficiency measures that are developed in section 5.3 from $F^x(X)$ as before. For example, returns to scale are measured by the sum of the output elasticities, where output is the aggregator function $T^y(Y)$. In addition, Δy in (11) must be replaced with the change in the aggregator function.

5.4.2 Combining Supply and Demand Simulation Models

Given data on demand as well as on supply, it possible to combine the two sorts of simulation models to obtain mutually consistent forecasts of merger–related efficiency and market power changes. We sketch this possibility here.

Suppose that two merger simulation models have been estimated for a product that is homogeneous in supply but differentiated in demand. The supply model takes the post merger output as given by the production function for the merged firm and determines post merger efficiency and marginal cost, whereas the demand model takes the pre merger marginal costs as given and determines post merger brand prices and outputs. To combine the two, we must solve two sets of first order conditions simultaneously, where the conditions from the supply model involve input choices, whereas those from the demand model involve price choices. This can be done through iteration.

Step one:

Given predicted inputs and output, solve the supply model for post merger marginal cost.

Step two:

If in step one it is found that marginal cost falls by $x\%$ post merger, reduce the marginal cost of each brand by $x\%$ and solve the demand model using the new costs. This step will determine a new aggregate output for the supply model. However, since we also need inputs, in the medium run we assume that time to build restricts K to be fixed and solve for the cost minimizing X for the new Y. As long as the marginal products of X are positive, the solution will be unique.[34]

Next, step one is repeated with new inputs and output. The process has converged once changes in marginal costs are sufficiently small across iterations.

---

[33]With the translog, one can impose linear homogeneity by imposing well known coefficient restrictions (see, e.g., Berndt and Wood (1975)).

[34]It may be possible to simulate long–run changes in both K and X, but this will require a dynamic model of investment costs.

6 Application

6.1 Specification of the Production Function

We must specify a functional form for f, the function of conventional inputs. We chose a translog because it is flexible[35] and because it nests the Cobb Douglas that is used by most researchers in this literature. That function is

$$f(v,k) = \beta_c \text{DC} + \beta_s \text{DS} + \beta_v v + \beta_k k + \beta_{vv} v^2 + \beta_{kk} k^2 + \beta_{vk} vk, \tag{14}$$

where all variables are in natural logarithms and DC and DS are country and sectoral dummy variables.[36]

We experimented with many other specifications. For example, we considered nonneutral technical change, i.e., with trending coefficients, $\beta_\ell = (\beta_{\ell 0} + \beta_{\ell t} t)$, $\ell = v,k$, but found that it was not possible to obtain robust estimates of the separate trends due to high correlation between v_t and k_t.[37] We also considered coefficients that varied by sector and country and nonlinear trends. However, we ran into problems of multicollinearity and overfitting and finally settled on the simple specification in equation (14). Although the specification is simple, it captures the features that we are interested in. In particular, it provides flexible estimates of RTS, TFP growth, and MC, that vary by firm, over time, and with output. Most other studies, whether production function estimation in the spirit of Olley and Pakes or merger simulation, do not incorporate this flexibility.

Finally, the functions h, the inverse of the investment function that determines unobserved productivity ω, and g, the function of lagged ω that determines how ω evolves, are low order polynomials of their arguments, where the order will depend on the application. We tried including factor prices in h. However, those variables were not significant.

6.2 Estimation and Selection

The following two equations are estimated (see section 5.1)

$$y_{jt} = \beta_0 + \beta_t t + f(v_{jt}, k_{jt}) + h(k_{jt}, i_{jt}, t) + \eta_{jt} \tag{15}$$

and

$$y_{jt} = \beta_0 + \beta_t t + f(v_{jt}, k_{jt}) + g[h(k_{jt-1}, i_{jt-1}, t-1)] + u_{jt}, \tag{16}$$

where j is a firm, $u_{jt} = \omega_{jt} - E(\omega_{jt}|\omega_{jt-1}) + \eta_{jt} = \xi_{jt} + \eta_{jt}$.[38]

Following Ackerberg, Caves, and Frazer (2015), we do not assume that the coefficients of v and v^2 are identified by the first equation alone. Instead, we estimate the two equations jointly by GMM as suggested by

---

[35]A function is flexible if it provides a second–order approximation to an arbitrary technology. In particular, the matrix of elasticities of substitution between inputs is unconstrained.

[36]Recall that V is an aggregate of L and M.

[37]We found that the performance measures obtained from the two specifications, one with trends and the other without, were very similar with one exception. The levels of technical change and thus TFP growth obtained from the model with neutral technical change were lower. However the rankings of those variables across firms were similar, and we are not interested levels per se but only in comparisons.

[38]As we discuss in detail below, when accounting for selection, the expectation in u_{jt} must also control for the fact that we only observe productivities for those firms that remain in the data between $t-1$ and t. We address this issue below.

[Wooldridge](#) (2009). In particular, we specify different instruments for the two equations, the difference being that all variables in (15) and their interactions and lagged values are valid instruments for the first equation, whereas contemporaneous values of v and i and any variables formed from them must be excluded from the second instrument set. The error u_{jt} in the second equation is correlated with v_{jt} and i_{jt} because it contains the current productivity shock, ξ_{jt}.

As noted by Wooldridge, joint GMM estimation has several advantages over the two–step approaches that most researchers use. First, equation (15) contains identifying information even when it does not identify any production function parameters by itself. Second, the errors in the two equations are allowed to be contemporaneously correlated and heteroskedastic. Finally, no bootstrapping is required as analytic standard errors are easy to obtain.

With our data on publicly traded firms, selection is particularly important. Indeed, firms exit the data when they cease to trade on a North American stock exchange. This can happen for three reasons: they merge with another firm, they go private, or they fail. However, in spite of the fact that there are three methods of exit, there are really only two outcomes: success or failure. We therefore classify firms as 'successful' if they remain in the data, if they go private, or they undergo a merger with a synergy motive. Firms are classified as 'failing,' in contrast, if they go bankrupt of if they undergo a merger for asset stripping purposes (delayed failure).

Failing firms are not a random set but are instead those with low productivity. We assume that the other three outcomes, remaining in the data, going private, and undergoing a merger for synergies are independent of productivity. For example, motives for going private include avoidance of onerous reporting to regulatory agencies and revelation of sensitive information, obtaining freedom to concentrate on long term goals, and providing incentives to management, none of which is directly related to pre–merger productivity. Furthermore, motives for a successful merger include reducing transport costs, achieving economies of scale, and reducing unit advertising costs. These goals, all of which involve post merger changes, do not depend on the productivity of the pre–merger firms. We therefore assume that the selectivity problem only arises for failing firms and we include the estimated probability of failure in the function that determines the evolution of unobserved productivity, g.

We model exit in two stages, in the first success or failure is determined. In the second, if firms have been successful, they choose among the three successful outcomes.[39] Let the first outcome be represented by a dummy variable D_{j1t}, where $D_{j1t} = 1$ indicates success, and the second outcome by D_{j2t}, which equals 1 if the firm remains as an independent publicly traded firm and 0 if it either merges for synergy reasons or goes private. A firm fails ($D_{j1t} = 0$) when ω_{jt} drops below a threshold, $\omega_{jt} < \tilde{\omega}_{jt}(k_{jt}, i_{jt}, t)$ for some function $\tilde{\omega}_{jt}$. Hence, D_{j1t} is correlated with ω_{jt}.

Formally, let Ω_{jt} be the information that is available in period t[40] and define

$$\chi_{jt} = D_{j1t}D_{j2t}.$$

---

[39]We assume that the shareholders make these decisions. In other words, they not only choose whether to liquidate the firm but also whether to accept a merger offer by a publicly traded or private equity firm.

[40]More precisely: Ω_{jt} is the sigma algebra generated by all random variables in all periods up to and including period t.

In other words, $\chi_{jt} = 1$ if the firm remains in the data. We assume that, conditional on $\Omega_{j,t-1}$ and $D_{j1t} = 1$, the decision to remain ($D_{j2t} = 1$) is uncorrelated with ω_{jt}, i.e.

$$\text{Cov}(\omega_{jt}, D_{j2t} | \Omega_{j,t-1}, D_{j1t} = 1) = 0.$$

Then,

$$E[\omega_{jt}|\Omega_{j,t-1}, \chi_{jt} = 1] = E[\omega_{jt}|\Omega_{j,t-1}, D_{j1t} = 1, D_{j2t} = 1] = \frac{E[\omega_t D_{j2t} | \Omega_{j,t-1}, D_{j1t} = 1]}{E[D_{j2t} | \Omega_{j,t-1}, D_{j1t} = 1]}$$

$$= E[\omega_{jt}|\Omega_{j,t-1}, D_{j1t} = 1] = E[\omega_{jt}|\omega_{j,t-1}, \omega_{jt} \geq \tilde{\omega}_{jt}(k_{jt}, i_{jt}, t)] = \tilde{g}[\omega_{j,t-1}, \tilde{\omega}_{jt}(k_{jt}, i_{jt}, t)].$$

Since the conditional probability of failure is a sufficient statistic for the outcome of this process, the solution proceeds in the usual way. In other words, an estimate of the probability of failure enters the equation for the evolution of unobserved productivity.[41]

7 Results

7.1 Production Functions

When estimating the GMM specifications, we found that the preferred model for the evolution of unobserved productivity, *g*, is a random walk with drift, which Wooldridge (2009) calls a 'leading case.' This finding greatly simplifies the estimation because it implies that the moment conditions are linear in the parameters. We assume that the inverse function *h* is quadratic in its arguments.

Table 3 contains three specifications of the translog production function. With all three, the dependent variable is output, $y = \log(Y)$. The specifications differ according to the estimation method used. The first was estimated by OLS, the second by GMM, and the third is the full model, GMM with selection.[42] All three show that, all else equal, output has been increasing over time and that, conditional on inputs, production is lower in Canada and in the craft sector.

The bottom panel of table 3, which contains means and standard deviations of estimated returns to scale (RTS), markups (PCM), technical change (TECH), and TFP growth (TFPG) for each specification, shows that average returns to scale range between 1.09 with OLS and 1.20 with the full model.[43] Furthermore, all distributions of RTS are symmetric. However, due to differences in estimates of the nonlinear coefficients, the distribution of GMM estimates is bimodal with a large number of firms with constant or slightly increasing returns and another group of firms with strongly increasing RTS. It is worth noting that the RTS distribution produced by GMM follows the actual firm size distribution more closely than the unimodal OLS distribution.

Average markups range from 0.96 with OLS to 1.19 with the full model. Although the range in average markups is large, histograms of that variable show that the distributions are similar in shape with both skewed

---

[41] See Olley and Pakes (1996) or Yasar, Raciborski, and Poi (2008) for examples of how selection is incorporated.

[42] The p values for the J statistics indicate that the over–identifying restrictions for the GMM specifications are not rejected. In particular, instrument validity is not rejected.

[43] These estimates are representative of the firms in our sample, not the industry as a whole.

Table 3: OLS and GMM translog Production Functions

	(1) OLS	(2) GMM	(3) GMM Selection
t	0.00734***	0.00760***	0.00760***
	(0.00143)	(0.00123)	(0.00122)
DC	-0.0683***	-0.101***	-0.101***
	(0.0256)	(0.0245)	(0.0243)
DS	-1.238***	-1.202***	-1.203***
	(0.0733)	(0.0610)	(0.0605)
v	0.735***	0.610***	0.612***
	(0.0656)	(0.0544)	(0.0543)
k	0.502***	0.354***	0.321***
	(0.0462)	(0.0881)	(0.0944)
v^2	-0.0153	0.00387	0.00335
	(0.0143)	(0.0138)	(0.140)
k^2	0.0446***	0.0431***	0.0493***
	(0.00360)	(0.0142)	(0.0144)
vk	-0.0642***	-0.00448	0.00222
	(0.0151)	(0.0320)	(0.0317)
Constant	-1.781***	-1.550***	-1.565***
	(0.0975)	(0.111)	(0.112)
p value, J stat		0.95	0.96
Observations	573	545	545
RTS Mean	1.09	1.17	1.20
	(0.09)	(0.16)	(0.20)
PCM Mean	0.96	1.16	1.19
	(0.51)	(0.38)	(0.39)
TECH Mean	0.008	0.007	0.007
	(0.12)	(0.12)	(0.12)
TFPG Mean	0.011	0.011	0.011
	(0.13)	(0.12)	(0.12)

t Robust standard error in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Standard deviations in parentheses, bottom section

v is the log of the variable input, k is the log of the fixed input, DC = 1 for Canada, DS = 1 for craft

The J statistic is Hansen's test of the over–identifying restrictions.

to the right.

With all three specifications, technical change is about 0.7% per year whereas TFP growth is 1.1%. Technical change is dominated by the trend, which is what one would expect given that ω is assumed to be a random walk plus drift. TFP growth is higher than technical change because output per firm has been growing and most firms are characterized by increasing returns. Finally, the standard deviations of TECH and TFPG are very large relative to their means, indicating that there is significant technological dispersion across firms in the brewing industry.

Table 4: Distributions of the Performance Measures

		RTS	PCM	TFPG	TECH
Mean		1.20	0.007	0.011	1.19
Quantiles					
(%)	25	1.04	0.96	-0.037	-0.035
	50	1.18	1.11	0.006	0.014
	75	1.36	1.29	0.057	0.064

Based on the GMM Specification with Selection
RTS   is returns to scale
PCM   is the price cost markup
TECH is technical change
TFPG is total factor productivity Growth

Table 4 contains the distributions of the estimated performance measures. It shows that, in addition to the large successful brewers, there are many small brewers that are pricing below marginal cost. Furthermore, there are many with negative technical change and TFP growth, which explains the high exit rate.

7.2   Analysis of RTS, markups, Technical Change, and TFP Growth

The averages that appear at the bottom of table 3 conceal considerable variation in all of the variables over time and firms. For this reason, we dig deeper into the determinants of RTS, PCM, TECH, and TFPG. In particular, we relate those measures to a firm's country, sector, and size (as measured by k) in a regression framework. With this analysis, we do not interpret the coefficients as causal but simply test whether there is a systematic tendency for different groups of firms to perform differently.

Table 5 looks at differences in performance over time and by firm characteristics based on the full model (column 3) in table 3. It shows that, all else equal, scale economies have increased and markups have fallen with time. It also shows that scale economies are lower in the craft sector and higher in Canada, whereas markups are higher in the craft sector. In addition, large firms benefit more from increasing returns and have higher markups. In contrast to RTS and PCM, there are neither trends nor country, sectoral or size differences in technical change and TFP growth. These findings are not surprising however, since those variables are rates of change, not levels.

Table 5: Performance by Country, Sector, and Size

	(1)	(2)	(3)	(4)
	RTS	PCM	TECH	TFPG
t	0.00301***	-0.00550***	-0.000712	-0.000931
	(0.000538)	(0.000880)	(0.000548)	(0.000582)
DC	0.0703***	0.00698	-0.00668	-0.00499
	(0.00982)	(0.0349)	(0.0119)	(0.0119)
DS	-0.310***	0.539***	0.0375	0.0343
	(0.0221)	(0.0847)	(0.0262)	(0.0258)
k	0.000610***	0.000838***	0.0000261	0.0000480
	(0.0000438)	(0.0000446)	(0.0000284)	(0.0000302)
Constant	1.122***	1.189***	0.0203*	0.0300**
	(0.0107)	(0.0207)	(0.0117)	(0.0127)
Observations	573	573	545	545

standard errors in parentheses

DC = 1 for US, = 0 for Canada

DS = 1 for mass produced, = 0 for craft

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.3 The Molson Coors Merger Simulation

7.3.1 Static Efficiency and Marginal Costs

Table 6 contains the results of the Molson Coors merger efficiency and marginal cost simulations. The merger was announced in the middle of 2004, which means that only data for 2003 and earlier would have been available for the evaluation. The first part of the table contains ex ante forecasts of efficiency based on equation (11) using data from 2002 and 2003 and the two measures of unobserved productivity, $\bar{\omega}$, a volume weighed average of the pre merger ω's, and ω_{max}, the maximum of the two. The table shows that simulated efficiency gains are between 7 and 12%, depending on the specification.[44]

Simulated efficiency gains can be compared to realized gains. The second part of table 6, which contains estimates based on equation (12), shows that gains were only 4% in the year of the merger, considerably less than forecast. However, by the second post merger year, realized are very close to simulated gains. Given the potential biases that are outlined in subsection 5.3, this is quite promising and possibly indicates that the biases have a tendency to cancel each other. We do not look at later years because there was another period of adjustment in 2007 due to brewery openings and closures.

Turning to magnitudes, it is common to measure the value of efficiencies as the merger–related reduction in costs, and indeed Coors forecast a cost reduction of $175 million at the time of the merger. With our

---

[44]The standard errors shown in the table were obtained without taking account of the fact that the maximum is a nondifferentiable function. This is largely immaterial here since the quantities whose maximum is taken are not close. If they were then one of the methods proposed by e.g. Woutersen and Ham (2016) should be used instead.

Table 6: Molson Coors Merger Efficiency Simulation and Ex Post Verification

Ex Ante Forecasts of Efficiency Gains (%)	Year	Using $\bar{\omega}$	Using ω_{max}
	2002	7*	10*
		(3.7)	(5.6)
	2003	9**	12*
		(3.9)	(6.3)

Ex Post Estimates of Efficiency Gains (%)	Year		
	2005	4	
	2006	9	

Ex Ante Forecasts of MC Reductions (%)	Year	Using $\bar{\omega}$	Using ω_{max}
	2002	7*	9*
		(3.7)	(5.3)
	2003	9**	10*
		(4.4)	(5.3)

Ex Post Estimates of MC Reductions (%)	Year		
	2005	8	
	2006	10	

Ex ante efficiency gains are based on equation (8)
Ex post efficiency gains are based on equation (9)
$\bar{\omega}$ is an output share weighted average of ω_{Coors} and ω_{Molson}
ω_{max} is the maximum of ω_{Coors} and ω_{Molson}
MC stands for marginal cost
Marginal cost reductions are relative to an output share weighted average of MC_{Coors} and MC_{Molson} pre merger
Bootstrapped standard errors in parenthesis, * $p < 0.10$, ** $p < 0.05$
Ex ante forecasts use pre merger data

Ex post estimates use post merger data

production function based analysis, however, we measure potential efficiencies as the value of the extra output that can be produced from the same level of inputs. To do this, we compare the value of pre and hypothetical post merger output holding prices and input usage constant at pre merger levels. This is clearly an approximate gain since output price and quantity can adjust post merger. However, since those adjustments will be negatively related, the change in revenue should be less sensitive to our assumptions. This calculation indicates that the merger increases the value of output by $80 and $120 million, depending on whether $\bar{\omega}$, or $\omega_{\max}$ is used.[45] The large width of this interval underscores how the potential gain depends on whether the merged firm is able to transfer know–how across its operations.

The third part of table 6 contains the marginal cost forecasts. Those forecasts, which are based on equation (9) and use pre merger data, range between an 7 and a 10 percent decline. This calculation represents the decline in marginal cost holding inputs fixed, which would be appropriate for use in a calculation of the upward pricing pressure of a merger. In a full merger analysis, one would want to account for how changes in the competitive environment induce the firm to alter output, as this would in general affect marginal costs. Nonetheless, the ex post estimates are quite close to the forecasts. This appears to support our approach as doing a reasonable job forecasting first-order changes in costs.

A fall in marginal costs of 7 to 10% may seem large. However, inframarginal costs, as well as average variable costs, are expected to fall by smaller amounts. We highlight marginal cost as it is the relevant variable for a merger price simulation. Overall, we believe that these results are a strong indication that merger simulations that assume that marginal cost are held fixed are likely to overestimate post merger prices.

7.3.2 TFP Growth

We next turn to changes in the growth rate of TFP after a merger, which we interpret as a dynamic efficiency gain. Our measure of TFP growth is based on equation (8). However, we do not include the trend because it does not change with the merger. Our measure therefore has two parts: growth that is due to changes in returns to scale and growth that is due to changes in ω. If we use $\bar{\omega}$, which is merely a weighted average of pre merger unobserved productivities, only the first part matters. However, if we use $\omega_{\max}$, the maximum of the productivities of the merging parties, the second part contributes as well.

Using $\bar{\omega}$, our estimate of the change in TFP growth that is due solely to changes in returns to scale is 0.1%, which may seem small. However, this change would move an average firm from a 1.2 to a 1.3% annual growth rate, which is not negligible. On the other hand, if we use $\omega_{\max}$, the change is 0.3%, which is substantial. These results indicate that the TFP growth increases from the merger are largely dependent on the merged firm being able to adopt the "best practices" of its most technically proficient ancestor.

---

[45]Our estimates are less than the cost reductions forecast by Coors. However, they were including cost savings due to brewery openings and closures in their estimates.

8   Final Remarks

Most mergers simulations of the sort that are used by competition authorities hold marginal costs fixed or change them exogenously and are therefore unlikely to capture the impact of projected cost efficiencies. This is despite the fact that firms almost always posit that cost reductions are a key motivation for the proposed merger. We leverage production function estimation techniques to propose a very simple method for forecasting merger–related efficiencies. Our approach provides a structural model to forecast a merger's effect on returns to scale, marginal costs, and TFP growth as well as an overall measure of efficiency gain. These forecasts can complement existing merger price simulations that evaluate competitive harm due to changes in market structure.

Our model of technology is based on the production function/productivity model that was pioneered by Olley and Pakes (1996) and is summarized in Ackerberg, Benkard, Berry, and Pakes (2007). We employ a translog production function that is able to capture both changes in returns to scale and changes in technical proficiency that can result from a merger. We draw upon the index–number/productivity literature to derive a theoretically consistent measure of TFP growth. In particular, we modify the standard measure to incorporate differences in economies of scale (a movement along the production function holding technology constant).

To illustrate our approach, we apply our proposed simulation method to evaluate the merger between Molson and Coors that occurred in early 2005 using panel data from the North American brewing industry. Because we have data for years following the merger, we are able to compare our forecast to retrospective measures of efficiency changes following the merger. We find that our simulations yield fairly accurate forecasts of efficiencies.

While we are able to provide a first-order forecast of merger efficiencies, supply side simulation alone has several important limitations. The main drawback is that, absent a model of demand, our simulation cannot be used for welfare calculations. In addition, our efficiency analysis does not provide estimates of price changes. It should therefore be clear that supply and demand simulation models are complements and that a complete analysis of competitive harm and offsetting efficiencies should involve using both price and efficiency simulations in an iterative fashion. We suggest a method of doing this.

Of course, our procedure requires production data, which is not typically part of a traditional merger simulation. That said, unlike equilibrium simulation models, we do not require data on all of the major players. Since it is difficult for competition agencies to subpoena data from firms that are not involved in a merger, this can be important. Moreover, it is not necessary to define the market in order to simulate the merger. Rather, we need only a panel of firms sufficient to estimate the firm's production function.

There are also limitations that both sorts of simulations share. For example, like equilibrium price simulations, we cannot forecast merger–induced changes in collusive behavior, product positioning, or the number of brands offered. A merger is a complex transaction and one cannot hope to capture all aspects of that transaction with a simple model of supply or demand. Given that fact, one might question the wisdom of using quantitative methods to infer merger–related changes in efficiencies and markups. However, if one side in a contested merger presents econometric evidence to demonstrate impact, or lack thereof, it is wise for

the other to respond in kind. Otherwise, it is difficult to argue that more realistic assumptions are not just technical niceties but instead can alter conclusions in important ways.

We conclude by highlighting how our work complements two recent papers on the impact of mergers in the brewing industry. Ashenfelter, Hosken, and Weinberg (2015) used a reduced–form analysis to assess the Miller Coors joint venture that occurred in 2008 and found small but significant increases in both prices and efficiencies (2–3%) post joint venture that roughly offset one other. Our efficiency estimates for Molson Coors are larger, which is perhaps due to the fact that Ashenfelter, Hosken, and Weinberg focus principally on reductions in transport costs.

Miller and Weinberg (2017) also assessed the Miller Coors joint venture. In a manner similar to the traditional merger simulation literature, they estimate marginal cost effects of the merger by combining a structural model of demand with a model of differentiated product price competition between firms. Using data pre and post merger, they find that the joint venture entailed significant cost reductions. While the cost reduction is more relevant to our work, Miller and Weinberg's primary focus is to document the possibility that mergers can induce tacit collusion. They find that the merger changed the market game and they reject the assumption of post merger Bertrand competition that underlies most merger price simulations. However, our merger was very different from the Miller Coors joint venture. In particular, whereas Miller Coors involved firms that interacted at arm's length pre merger, Molson Coors was a cross border merger between firms that already produced each others' brands under license. We therefore think that conduct changes are unlikely in our context.

References

ACKERBERG, D., L. BENKARD, S. BERRY, AND A. PAKES (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6. North Holland, Amsterdam.

ACKERBERG, D., K. CAVES, AND G. FRAZER (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83(6), 2411–2451.

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58(2), 277–297.

ARELLANO, M., AND O. BOVER (1995): "Another Look at the Instrumental Variable Estimation of Error–Components Models," *Journal of Econometrics*, 68(1), 29–51.

ASHENFELTER, O. C., D. S. HOSKEN, AND M. C. WEINBERG (2015): "Efficiencies Brewed: Pricing and Consolidation in the US Beer Industry," *The RAND Journal of Economics*, 46(2), 328–361.

BAJARI, P., C. BENKARD, AND J. LEVIN (2007): "Estimating Dynamic Models of of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

BALAT, J., I. BRAMBILLA, AND Y. SASAKI (2016): "Heterogeneous Firms: Skilled Labor Productivity and Export Destimations," Johns Hopkins University.

BERNDT, E. R., AND D. O. WOOD (1975): "Technology, Prices, and the Derived Demand for Energy," *The Review of Economics and Statistics*, 57(3), 259–268.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 847–890.

BLACK, R. M., G. S. CRAWFORD, S. LU, AND H. WHITE (2004): "A Virtual Stakes Approach to Measuring Competition in Product Markets," Bates White, University of Arizona, and UC-San Diego.

BLAIR, R. D., AND J. S. HAYNES (2011): "The Efficiencies Defense in the 2010 Horizontal Merger Guidelines," *Review of Industrial Organization*, 39, 29357–68.

BLUNDELL, R., AND S. BOND (2000): "GMM Estimation with Persistent Panel Data: an Application to Production Functions," *Econometric Reviews*, 19(3), 321–340.

BRESNAHAN, T., E. BRYNJOLFSSON, AND L. HITT (2002): "Information Technology, Workplace Organization, and the Demand for Skilled Labor," *Quarterly Journal of Economics*, 117, 339–376.

CAVES, D., L. CHRISTENSEN, AND J. SWANSON (1981): "Productivity Growth, Scale Economies, and Capacity Utilization in U.S. Railroads, 1995–74," *American Economic Review*, 71, 994–1002.

CORTS, K. S. (1999): "Conduct Parameters and the Measurement of Market Power," *Journal of Econometrics*, 88, 227–250.

CRANE, D. A. (2014): "Rethinking Merger Efficiencies," *Michigan Law Review*, 110(3), 347–391.

DE LOECKER, J., AND F. WARZYNSKI (2012): "Markups and Firm Level Export Status," *American Economic Review*, 102(6), 2437–2471.

DENNY, M., M. FUSS, AND L. WAVERMAN (1981): "The Measurement and Interpretation of Total Factor Productivity in Regulated Industries, with Application to Canadian Telecommunications," in *Productivity Measurement in Regulated Industries*, ed. by T. Cowing, and R. Stevenson, pp. 179–218. Academic Press, New York.

DIEWERT, W. E., AND A. O. NAKAMURA (2007): "The Measurement of Productivity for Nations," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6, pp. 4501–4586. North Holland, Amsterdam.

FAN, Y. (2013): "Ownership Consolidation and Product Charachteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 103(5), 1598–1628.

FARRELL, J., AND C. SHAPIRO (2010): "Antitrust Evaluation of Horizontal Mergers: An Economic Alternative to Market Definition," *B.E. Journal of Theoretical Economics*, 10(1), Article 9.

GANDHI, A., S. NAVARRO, AND D. RIVERS (2016): "On the Identification of Production Functions: How Heterogenous is Productivity?," University of Western Ontario.

GOWRISANKARAN, G., A. NEVO, AND R. TOWN (2015): "Mergers when Prices Are Negotiated: Evidence from the Hospital Industry," *American Economic Review*, 105(1), 172–203.

HALL, R. E. (1988): "The Relationship Between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 96(5), 921–947.

HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): "Competitive Analysis with Differentiated Products," *Annales d'Economie et de Statistique*, 34, 159–180.

HODERLEIN, S., J. FOX, V. HADAD, AND A. PETRIN (2016): "Heterogenous Production Functions, Panel Data, and Productivity Growth," Boston College.

IVALDI, M., AND F. VERBOVEN (2005): "Quantifying the Effects from Horizontal Mergers in European Competition Policy," *International Journal of Industrial Organization*, 23(9–10), 669–691.

JEZIORSKI, P. (2014a): "Estimation of Cost Efficiencies from Mergers: Application to US Radio," *RAND Journal of Economics*, 45(4), 816–846.

———— (2014b): "Effects of Mergers in Two–Sided Markets: The US Radio Industry," *American Economic Journal: Microeconomics*, 6(4), 35–73.

JORGENSON, D., AND Z. GRILICHES (1967): "An Explanation of Productivity Change," *The Review of Economic Studies*, 34(3), 249–283.

KASAHARA, H., P. SCHRIMPF, AND M. SUZUKI (2015): "Identification and Estimation of Production Functions with Unobserved Heterogeneity," University of British Columbia.

KERKVLIET, J. R., W. NEBESKY, C. H. TREMBLAY, AND V. J. TREMBLAY (1998): "Efficiency and Technical Change in the U.S. Brewing Industry," *Journal of Productivity Analysis*, 10(3), 271–288.

KOLASKY, W., AND A. DICK (2003): "The Merger of Guidelines and the Integration of Efficiencies into Antitrust Review of Horizontal Mergers," Wilmer Cutler Pickering Hale and Dorr Antitrust Series, Working paper 31.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–341.

MAKOWSKI, L. (1987): "Are 'Rational Conjectures' Rational?," *Journal of Industrial Economics*, 36(1), 35–47.

MARSCHAK, J., AND W. H. ANDREWS (1944): "Random Simultaneous Equations and the Theory of Production," *Econometrica*, 13(1), 143–205.

MICHEL, C. (2016): "Identification and Estimation of Intra–Firm and Industry Competition via Ownership Change," Universitat Pompeu Fabra.

MILLER, N., AND M. WEINBERG (2017): "Understanding the Price Effects of the MillerCoors Joint Venture," *Econometrica, forthcoming*.

NEVO, A. (2000): "Mergers with Differentiated Products: The Case of the Ready–to–Eat Cereal Industry," *RAND Journal of Economics*, 31(3), 395–421.

——— (2001): "Measuring Market Power in the Ready–to–Eat Cereal Industry," *Econometrica*, 69(2), 307–342.

OLLEY, G. S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297.

PINKSE, J., AND M. E. SLADE (2004): "Mergers, Brand Competition, and the Price of a Pint," *European Economic Review*, 48(3), 617–643.

PINKSE, J., M. E. SLADE, AND C. BRETT (2002): "Spatial Price Competition: A Semiparametric Approach," *Econometrica*, 70(3), 1111–1155.

RICHTER, M. K. (1966): "Invariance Axioms and Economic Indexes," *Econometrica*, 34(4), 739–755.

SLADE, M. E. (2004b): "Market Power and Joint Dominance in UK Brewing," *Journal of Industrial Economics*, 48(1), 133–163.

——— (2009): "Merger Simulations of Unilateral effects: What Can We Learn from the UK Brewing Industry?," in *Cases in European Competition Policy: The Economic Analysis*, ed. by B. Lyons. Cambridge University Press, Cambridge, UK.

SOLOW, R. M. (1957): "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 39(3), 312–320.

WOOLDRIDGE, J. M. (2009): "On Estimating Firm–Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104, 112–114.

WOUTERSEN, T., AND J. C. HAM (2016): "Confidence Sets for Continuous and Discontinuous Functions of Parameters," National University of Singapore and University of Arizona.

YASAR, M., R. RACIBORSKI, AND B. POI (2008): "Production function estimation in Stata using the Olley and Pakes method," *Stata Journal*, 8(2), 221—231.

ZHANG, H. (2015): "Non-Neutral Technology, Firm Heterogeneity, and Labor Demand," University of Hong Kong.

A   Data Appendix

Table 7: Data Sources

United States

Abreviation	Source
AR	Company Annual Reports
ASM	Annual Survey of Manufacturers, Census Bureau
BA	Brewers Almanac, The Beer Institute
BEA	Bureau of Economic Analysis, Department of Commerce
BLS	Bureau of Labor Statistics
BMI	Beer Marketer's Insights
Compustat	S&P Capital, Compustat
HSUS	Historical Statistics of the US, Cambridge University Press
NASS	National Agricultural Statistics Service, USDA
NBER	NBER-CES Manufacturing Industry Data Base
Weinberg	The Office of Robert S. Weinberg

Canada

Abreviation	Source
AR	Company Annual Reports
ASB	Annual Statistical Bulletin, Beer Canada
ASML	Annual Survey of Manufacturing & Logging, StatCan
CANSIM	Computerized data base, Statistics Canada
HSBC	Historical Statistics, Beer Canada
StatCan	Statistics Canada

International

Abreviation	Source
IMF-IFS	International Monetary Fund, International Financial Statistics

North American Data

Data for all firms consist of firm–specific prices and quantities of output, labor, materials, capital, and investment. In addition, the mode of exit is included for all firms that left the data. Data sources can be found in table 7.

Output

All of the firms are brewers and most produce only beer. Beer prices and quantities are therefore the primary output data. Firm revenues net of taxes are from Compustat (Compustat mnemonic SALES). Beer production up to 2009 for each US firm is from Weinberg,[46] and after 2009 from BMI. US firm–level beer prices are then calculated as revenue net of taxes divided by production. Canadian production data for the later years (1990s and beyond) are from company annual reports, and firm–level prices for those years are calculated as in the US. Due to a lack of production data, a Canadian industry average price was used for the earlier years. This is calculated as beer industry value of shipments from ASML divided by industry production from CANSIM. The production of Canadian firms in the early years is then calculated as firm net revenue divided by industry price.

A few firms (4) produced in multiple markets for a decade or so and those markets are quite diverse (e.g., from wine and soft drinks to sports and entertainment). However, in all cases, revenue from the other segments is a small fraction of the total. Moreover, in recent years, most brewers have sold their non–beverage assets in order to concentrate on brewing. Revenue by segment was obtained from the Compustat Product Segment data base. Segment prices for US non–beer manufacturing markets are from NBER, augmented with BEA prices after 2009. Canadian non–beer manufacturing prices are from CANSIM. Finally, the relevant CPI or PPI (US or Canadian from BLS or CANSIM) was used for the non–manufacturing segments, sports, entertainment, primary energy and retail. An output price index was then constructed as a revenue share weighted geometric mean of segment prices and output was calculated as net revenue divided by that price.

Labor

The number of employees (full time equivalents) in each firm is from Compustat (EMP) and the wage is calculated as expenditures on labor, also from Compustat (StaffExp), divided by the number of employees. When firm–level labor expenditures are missing, a beer industry average wage is used. The average wage for the US is calculated as industry expenditures on labor divided by the number of employees in the industry, both from ASM. For Canada, the industry data on employment and expenditures are from ASML.

Materials

A materials price index was calculated for four segments: US and Canadian mass production and craft. Materials prices differ by segment because both expenditure shares and factor prices differ. For each segment, a raw materials price index was computed as an expenditure share weighted geometric mean of the prices of malt, hops, corn, rice, wheat, and sweeteners (sugar and corn syrup). A packaging price index was created as a expenditure share weighted geometric mean of the prices of bottles, cans, and cartons. An overall materials price index was then obtained as a expenditure share weighted geometric mean of raw materials and packaging price indexes. This two–stage process was adopted to reflect the fact that substitution is easier within compared to across groups of inputs.

Expenditure shares for the US are from BA while those for Canada are from StatCan. When share data were missing, they were extrapolated (33 points, all for small breweries). US raw materials prices are from

---

[46]This data was given to us by Carol and Victor Tremblay.

HSUS and NASS and Canadian raw materials prices are from CANSIM. In some cases, Canadian prices are withheld due to confidentially. When this occurred, the US price in CAD was substituted, using exchange rates from IMF–IFS. Finally, packaging prices are producer price indices from BLS.

Compustat does not report materials expenditures directly. Following Bresnahan, Brynjolfsson, and Hitt (2002), materials expenditures are calculated by subtracting labor expenditures from total expenditures, where total expenditures are calculated as net revenue minus operating income before depreciation (OIBD). The quantity of raw materials is then materials expenditure divided by materials price.

Capital and Investment

Data on investment flows and capital stocks, both in dollars, are from Compustat (CAPEX and PPET). In a few cases where investment data were missing, investment was calculated from changes in capital stocks and depreciation (DAA) data or was obtained from annual reports. It is assumed that the capital equipment used by breweries trades in international markets and that all firms face the same investment prices. NBER brewery investment prices are used and, after 2009, are augmented using data from their sources. Canadian investment prices are USD prices in CAD, using exchange rates from IMF–IFS. Finally, for those firms that produce in multiple markets, an investment price index was created as an investment share weighted geometric mean of investment prices, using investment shares calculated from Compustat product segments data and NBER segment investment prices. Real capital and investment are constructed as dollar values divided by investment prices.

Currency and Deflation

The currency for each observation is that reported in Compustat. This means that, with the exception of Carling O'Keefe who report values in USD in the early years, data for US firms are in USD and for Canadian firms are in CAD. In practice, as long as all values for a given observation are in the same currency, the currency is irrelevant. This is true because, for example, when a quantity variable is created by dividing expenditures by price, expenditures and price are in the same currency. For the same reasons, it doesn't matter if monetary variables are in constant or current dollars. However, units of measurement matter when estimating marginal costs. For those estimations, Canadian dollars were converted to US and all were deflated by the US PPI, all items. Real USD were also used in performing the ex post merger assessment.

Measurement Error

As with most empirical studies, our data are not perfect and measurement error is a possibility. In particular, for some observations, we use an industry, rather than a firm, output price or wage. However, this problem is less severe for us than for much research in this tradition, where industry wide input and output prices are used to deflate all monetary variables (e.g., Olley and Pakes (1996)) or where a single overall price index is used to deflate all variables (e.g., Levinsohn and Petrin (2003) and Gandhi, Navarro, and Rivers (2016)). Nevertheless, we discuss the issue. Measurement error in y does not create a problem as long as it is not correlated with the explanatory variables. On the other hand, measurement error in an explanatory variable tends to bias the coefficients and t statistics towards zero. As can be seen by examining table 3, we do not have those symptoms and therefore do not try to correct the problem.

Exit

Exit from the data base occurs for three reasons: failure (a liquidation), purchase by another firm (a merger), or private purchase (a management or private–equity buyout). In addition, mergers can have two motives,

synergies or asset stripping. An internet and literature search was used to determine the type and reason for each exit in the data.

US Data

Additional data for firms in the US mass production sector consists of industry concentration ratios, firms' shares of production, and import shares of consumption. Concentration ratios, the output of the four largest firms divided by industry production, and firms' shares of production are constructed from the Weinberg data on output by firm and data from BA on industry production. Import shares are constructed from data on imports and consumption from BA.