

2-2014

Price and Quality Dispersion in an Offshoring Market: Evidence from Semiconductor Production Services

David Byrne
Federal Reserve Board

Brian K. Kovak
Carnegie Mellon University, bkovak@cmu.edu

Ryan Michaels
University of Rochester

Follow this and additional works at: <http://repository.cmu.edu/heinzworks>

 Part of the [Databases and Information Systems Commons](#), and the [Public Policy Commons](#)

This Working Paper is brought to you for free and open access by the Heinz College at Research Showcase @ CMU. It has been accepted for inclusion in Heinz College Research by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Price and Quality Dispersion in an Offshoring Market: Evidence from Semiconductor Production Services*

David Byrne
Federal Reserve Board

Brian K. Kovak[†]
Carnegie Mellon University
and NBER

Ryan Michaels
University of Rochester

February 2014

Abstract

We study price and quality differences across international intermediate input suppliers. Our uniquely detailed transaction-level data exhibit large price differences across suppliers of observationally identical products, but these price gaps narrow over the product life cycle. We disentangle pure (frictional) price dispersion from unobserved quality differences using a framework in which switching suppliers is costly. We construct a quality-adjusted price index using these results and show that pure price dispersion biases standard indexes upward. Finally, we demonstrate how to bound the true constant-quality price index using commonly available data and without making strong market structure assumptions.

*An earlier version of this paper was prepared for the conference “Measurement Issues Arising from the Growth of Globalization,” sponsored by the W.E. Upjohn Institute for Employment Research and the National Academy of Public Administration (NAPA), and funded by the Alfred P. Sloan Foundation. Thanks to Yong-Gyun Choi, Manuel Gomez, Steven Paschke, and Fabio Rueda for excellent research assistance. We would like to thank Bill Alterman, Wenjie Chen, Ken Flamm, Erica Fuchs, Ross Goodman, Susan Houseman, Ben Keys, Nina Pavcnik, Marshall Reinsdorf, Jodi Shelton, Lowell Taylor, Sonya Wahi-Miller, Kim Zieschung, seminar participants at Columbia, Carnegie Mellon, Essex Trade Workshop, Queens College, U.S. Dept. of Commerce, NBER CRIW Summer Institute, NBER ITI Meetings, Upjohn/NAPA conference, and the Rocky Mountain Empirical Trade Conference for helpful discussions and Chelsea Boone at GSA, Peter Dziel at LSI, and Len Jelinek at IHS iSuppli for help with data and background information on the industry. Remaining errors are our own. This paper reflects the views of the authors and should not be attributed to the Board of Governors of the Federal Reserve System or other members of its staff.

[†]Corresponding author: bkovak@cmu.edu

1 Introduction

Accurate measures of market prices are important in all branches of applied economic analysis. One of the most persistent challenges in the practice of price measurement is accounting for changes in product quality.¹ Such changes may involve products' characteristics, but may also include aspects of the overall transaction, such as customer service or timely delivery (Carlton 1983). In practice, it is often quite difficult to measure both physical product attributes and the less tangible characteristics of transactions. While these challenges have been well known for decades, they have recently taken on particular significance in markets for intermediate inputs. Such markets are characterized by increased internationalization of production chains and shifts toward relatively low-price suppliers in developing countries such as China (Hummels, Ishii and Yi 2001). Moreover, an increasing number of "factoryless manufacturers" have outsourced product fabrication activities altogether (Bayard, Byrne and Smith 2013, Bernard and Fort 2013). These developments have led to increased substitution across suppliers of intermediate inputs, both domestic and international. In this context, failure to accurately estimate quality-adjusted price differences across suppliers will lead to biased import quantity and productivity measures.

In this paper, we examine price and quality differences across intermediate input suppliers located in different countries. We utilize uniquely detailed data from the semiconductor industry to observe physical product characteristics, and develop a novel approach to inferring unobserved quality differences across suppliers. These methods allow us to identify constant-quality price dispersion, which is substantial in this market. The results inform our understanding of price-setting among international competitors and have practical significance for constructing price indexes.

We make three main contributions. First, we investigate price dispersion using a new database that reports transaction prices for offshore contract manufacturing services in the semiconductor industry. The database provides extremely detailed information on all relevant product characteristics. This information is far more detailed than the Harmonized System (HS) product codes reported in disaggregate trade data, which contain substantial product heterogeneity, even at the most detailed (10-digit) level.² This level of detail allows us to study price differences across suppliers of technologically identical products. The data exhibit substantial shifts in production across countries and show that suppliers with growing market shares typically charge much lower prices for observationally equivalent goods. Chinese producers on average charged 17 percent less than firms in market leader Taiwan for observably identical products, and increased their market share from 7.1 percent in 2000 to 21.8 percent in 2011. Moreover, the average price differences mask im-

¹The 1961 Price Statistics Review Committee (1961), chaired by George Stigler, declared that "If a poll were taken of professional economists and statisticians, in all probability they would designate (and by a wide majority) the failure of the price indexes to take full account of quality changes as the most important defect in these indexes."

²For example, conventional laptops and tablet computers are found in the same 10-digit category. Our entire analysis falls within one 10-digit HS code: 8542.21.80.05 - Electronic integrated circuits and microassemblies; parts thereof: Monolithic integrated circuits: Digital: Unmounted chips, dice and wafers: Silicon.

portant dynamics in price dispersion over the product life cycle. The price gaps between suppliers start out very large early in the life of a product generation and then converge over time.

Second, we develop a novel method using these dynamics to adjust observed price differences for *unobserved* heterogeneity. The latter remains a concern, despite our exhaustive information on physical product attributes, because there may be differences across suppliers in hard-to-measure aspects of transactions, such as customer service or design assistance, that are not reflected in physical product features.³ Our approach to this issue is informed by a model in which suppliers enter sequentially and buyers face a cost to switch suppliers, following Klemperer (1995). In this setting, the leading supplier initially charges a relatively high price to exploit its customers, who are partially locked in by the switching cost, and then charges a lower relative price as its existing customers phase out and its market power erodes. As a result, the influence of the switching cost on price dispersion abates, and late in the product cycle, price dispersion reflects only differences in quality across suppliers. This feature makes it possible to empirically proxy for the contribution of unobserved heterogeneity using observed prices late in the product cycle. We use this approach to show that pure price dispersion and unobserved heterogeneity are both quantitatively important in explaining price dispersion for semiconductor manufacturing services.

Third, we present two methods for measuring quality-adjusted price indexes in the presence of pure price dispersion and unobserved quality differences. The first method takes its cue from the theoretical framework summarized above and infers the effects of unobserved heterogeneity on prices using price differences late in the product cycle. We then directly adjust observed prices for these quality differences and generate a constant-quality price index for contract semiconductor manufacturing.⁴ While this approach is unique, the empirical findings echo those of Houseman, Kurz, Lengemann and Mandel (2011), showing that standard index construction methods used by statistical agencies systematically omit price declines occurring when buyers substitute toward producers with lower quality-adjusted prices.⁵ We estimate that this omission would result in upward bias of 0.4 percentage points per year in prices for the semiconductor manufacturing market.⁶ The second method demonstrates how statistical agencies can generate bounds for the true constant-quality index using their standard calculations in conjunction with an alternative index following

³Kaplan and Menzio (2014) measure price dispersion of retail goods using scanner data. Interestingly, they note that the role of store amenities, such as the quality of service, accounts for little of the dispersion. However, this dimension is likely to matter more in intermediates markets, such as ours, where products are typically customized for the buyer.

⁴Our quality-adjustment method uses only information on product characteristics and prices. This contrasts with recent proposals to improve price measurement by collecting scanner data on prices and quantities (Hausman and Leibtag 2009). The minimal data requirements of our approach are encouraging, since statistical agencies may not have the resources to collect additional data on quantities for price index construction.

⁵An example of an index that uses this standard method is the such as International Price Program (IPP) of the US Bureau of Labor Statistics (BLS).

⁶To get a sense of scale, if bias of this size applied to all imported intermediate inputs, U.S. multifactor productivity growth estimates would be biased upward by 0.066 percentage points per year, where the official growth in multifactor productivity was 1.5 percent per year during the sample period.

Reinsdorf (1993). This method is general, in that it does not rely on the market structure or price dynamics, and does not require detailed information on product attributes. Furthermore, it can be implemented using the data already collected by statistical agencies.

While our analysis focuses on semiconductor manufacturing, it can be applied more widely. We focus on a contract manufacturing industry characterized by factoryless customers, reflecting a market structure that is becoming quite common throughout the U.S. manufacturing sector (Bayard et al. 2013). The switching costs that underlie price dispersion in this market appear to be quite prevalent in other intermediate input markets. Monarch (2013) finds that 45 percent of US arms-length importers maintain their primary Chinese supplier across years. Using a dynamic model of an importer's choice of suppliers, he infers that the length of these relationships implies substantial costs of switching international suppliers.⁷ Given these similarities, we expect our results to be informative outside the semiconductor industry.

Our paper follows a large literature documenting the price measurement challenges posed by entry of low-priced suppliers.⁸ Focusing on the domestic retail context, these papers argue that official indexes exhibit an "outlet substitution bias" by omitting price declines that occur when large retailers enter a market. We focus on a similar problem in the international context, contributing to an emerging literature that examines the implications of globalization for the accuracy of official statistics.⁹ The most closely related prior work is Houseman et al. (2011), who study the effects of shifting sourcing on import price measures using BLS IPP micro data covering all of U.S. manufacturing. While our data are more narrow in industry coverage, they make possible highly credible quality adjustments based on observable characteristics, and the narrower scope permits inferences about unobserved quality differences based on industry structure. In spite of differences in data, industry coverage, and methodology, the papers find comparable biases, suggesting their general applicability.

Second, our paper contributes to ongoing work on price dispersion in intermediate input and wholesale markets.¹⁰ Perhaps the most related paper is Foster, Haltiwanger and Syverson (2012).¹¹

⁷ Eaton et al. (2013) also document long-lasting relationships between U.S. importers and Columbian partners, noting that attachments appear to emerge after importers have experimented with shorter-lived matches.

⁸See, for example, the Boskin Commission Report (Boskin, Dullberger and Gordon 1998), Greenlees and McClelland (2008), and Hausman and Leibtag (2009).

⁹Feenstra, Mandel, Reinsdorf and Slaughter (2013) study three distinct mechanisms through which globalization can bias standard import price measures upward: using a non-superlative index number formula, omitting the effect of tariffs on purchase prices, and omitting the effects of increased variety. Their mechanisms are independent of ours, suggesting that the overall upward bias to import prices is larger than that reported by either paper. Also see Houseman and Ryder (2010) for a summary of papers discussing measurement problems arising from globalization.

¹⁰The contract semiconductor production industry is highly concentrated, so mechanisms generating quality-adjusted price variation based on search or informational frictions are unlikely to apply (Salop and Stiglitz 1977, Burdett and Judd 1983, Sorensen 2000). This market also faces very low transportation costs, with high value to weight ratio products, so this is also an unlikely source of price dispersion (Syverson 2004).

¹¹Of course, research on price dispersion in the input market stretches back further, at least to the seminal work by Stigler and Kindahl (1970). Abbott (1992) is also an important predecessor. However, our paper and Foster et al. (2012) differ from this earlier literature by pursuing evidence for a particular mechanism driving dispersion.

They explore the dynamic process of “building a customer base,” in which larger current sales increase future demand. Our framework, based on costly switching across intermediate input suppliers, generates a very similar demand structure; lowering the price to attract more customers in one period increases demand next period, as more customers are partially locked-in by the switching cost. Thus, our model reflects a particular mechanism driving the more general phenomenon Foster et al. document.¹²

Third, our estimates are consistent with findings in the trade literature of substantial differences in the quality of imports coming from different countries, even within narrow product classifications (e.g. Feenstra and Romalis (2013), Hallak and Schott (2011), Hummels and Klenow (2005), Khandelwal (2010), and Kugler and Verhoogen (2011)). However, our measurement strategy is distinct. Papers in this literature typically infer quality from differences in market share conditional on price. By using data with far more detail on product attributes than is available in disaggregate trade data used by these previous studies, we measure differences in the quality of the physical product via direct observation. Thus, we follow Schott’s (2008) suggestion to use “very detailed data about the hedonic attributes of goods produced and exported by China and developed economies” to reveal cross-country quality differences. We also go one step further by using price dynamics to infer differences in the quality of less tangible aspects of the transaction.

Lastly, the paper’s application to the semiconductor market situates it within a large literature on this industry. Although semiconductor pricing has been studied frequently, nearly all prior work examines the markets for commodity semiconductor products such as microprocessors and memory chips.¹³ There is little price variation across suppliers in these markets, so related studies instead focus on learning-by-doing to explain stunning rates of average price decline across suppliers.¹⁴ We examine the market for semiconductor production services provided by contract manufacturing firms called “foundries.”¹⁵ Due to a lack of detailed data, this market has been studied infrequently in the previous literature, and it differs in important ways from processor and memory markets.¹⁶ Most foundry products are custom designs tailored to particular applications rather than for general-purpose computing. The custom nature of each product partly drives the large costs incurred when shifting a product from one foundry to another. Thus, our approach focuses on price variation between suppliers rather than on the sources of price declines across all

¹²Drozd and Nosal (2012) present a theoretical model that uses a very similar mechanism to generate deviations from the law of one price in the context of international business cycles.

¹³For microprocessors, see Dulberger (1993), Grimm (1998), Doms, Aizcorbe and Corrado (2003), Holdway (2001), and Flamm (2007) and for memory see Flamm (1993), Grimm (1998), and Aizcorbe (2002).

¹⁴For models of learning by doing in semiconductor production, see Baldwin and Krugman (1988), Irwin and Klenow (1994), and Flamm (1996), all of which assume the law-of-one-price holds.

¹⁵Fort (2013) provides a thorough discussion of contract manufacturing in the U.S. economy.

¹⁶Aizcorbe (2002) constructs price indexes with the limited previously available data for these products and argues for using a combination of prices of processors and memory chips as a proxy for the devices produced by foundries when data are not available.

suppliers.¹⁷

The paper proceeds by first discussing the technology of semiconductor fabrication in Section 2, which outlines the key technological attributes relevant for price setting, discusses contract manufacturing in the semiconductor context, and contrasts this market with the more frequently studied memory and processor markets. In Section 3, we investigate price differences and price dynamics between overseas manufacturers using a hedonic framework to control for an exhaustive set of product attributes across transactions. In Section 4, we present a model of switching costs and sequential supplier entry, which rationalizes the empirical findings and suggests an approach to inferring unobserved quality differences. Section 5 discusses the implications for price index construction. It presents a quality-adjusted price index and proposes a straightforward approach to bound the true constant-quality index, even in the presence of pure price dispersion and unobserved heterogeneity. Section 6 concludes.

2 Industry Background

To understand the subsequent analysis, we need to briefly review three important features of the contract semiconductor manufacturing industry. First, there are distinct, measurable technological attributes of semiconductors that significantly affect their price. Our data are remarkable in part because they provide information on each of these technological characteristics, allowing us to control for them in our analysis. Second, it has become more common for firms designing semiconductor chips to outsource production to specialized manufacturing firms overseas. This business model results in the arm’s-length purchases of contract manufacturing services that we focus on and that have contributed to large shifts in the geographic distribution of output around the globe. Third, in contrast to general-purpose processors and memory chips that have been the focus of previous work, contract semiconductor manufacturing primarily involves custom designs produced in much smaller quantities. This custom aspect to production implies that buyers must make supplier-specific investments, which raise the cost of switching suppliers at a later date. We argue that this friction in switching suppliers sustains quality-adjusted price variation across suppliers.

2.1 Semiconductor Wafer Technology

Semiconductor fabrication involves creating networks of transistors on the surface of a thin piece of semiconducting material called a “wafer.”¹⁸ The process begins with the design and layout of a

¹⁷Since our dynamic model is solved by looking for subgame perfect equilibrium, integrating learning by doing would be substantially more complex than in the one-shot Cournot (precommitment) equilibria used in much of the previous work on semiconductor prices. See Fudenberg and Tirole (1983) for a discussion of precommitment vs. subgame perfect equilibria in this context and Besanko, Doraszelski, Kryukov and Satterthwaite (2010) and Besanko, Doraszelski and Kryukov (2011) for examples of subgame perfect equilibrium in models with learning by doing.

¹⁸Turley (2003) provides an accessible overview of semiconductor technology, manufacturing, and business.

new chip. Semiconductor designers use complex software suites to specify the functionality of the chip, convert that logic into the corresponding network of transistors, determine the physical layout of those transistors, and simulate the behavior of the proposed design for debugging purposes.

Semiconductors are manufactured in a facility called a “fab.” Transistors are created on the surface of the wafer through a photolithography process, in which successive layers of conducting and insulating materials are deposited on the surface of the wafer and chemically etched away in the appropriate places to form the desired pattern of transistors and necessary interconnections. Design layout software determines the etching pattern for each layer, which is projected onto the wafer through a “mask” containing the desired pattern, in a process similar to developing a photograph by projecting light through a negative. Each step of the etching process is repeated multiple times across the wafer, resulting in a grid pattern of many copies of the chip. Once all transistors and connection layers are complete, the chips are tested in a process called “wafer probe,” and any faulty chips are marked to be discarded. The wafer is cut up, leaving individual chips, called “die,” that are placed inside protective packages and connected to metal leads that allow the chip to be connected to other components.

Semiconductor fabrication technology has advanced over time in discrete steps, defined by wafer size and line width (also called feature size). Increases in wafer size allow larger numbers of chips to be produced on a wafer. During our sample period, fabs produced 150mm (roughly 6 inch), 200mm (8 inch), or 300mm (12 inch) diameter wafers. Although larger wafers cost more to produce, each wafer contains many more die, so the move to a larger wafer has generally reduced the cost per die by approximately 30 percent (Kumar 2007).

Line width is the size of the smallest feature that can be reliably created on the wafer. Decreased line width means that individual transistors are smaller. A 30 percent decrease in line width approximately doubles the density of transistors on a chip. This makes chips of a given functionality smaller, lighter, faster, and more energy efficient, and also makes it feasible to include more functions on a single chip. The number of transistors that can be produced on a chip has grown exponentially over time, following Moore’s Law.¹⁹ Figure 1 shows the maximum number of transistors per chip and the minimum line width used to produce Intel processors over the last 40 years (both plotted on logarithmic scales).

Current line widths are measured in microns (μm) or nanometers (nm). In our sample, line widths range from 45nm to more than 500nm. As a rule of thumb, Kumar (2007) estimates that moving a given chip design to a 30 percent smaller line width results in cost savings of approximately 40 percent, assuming the same number of defects in both processes. The primary drawback of smaller line widths is increased cost per wafer, particularly early in the technology’s life span. Masks are much harder to produce when creating smaller features. In addition, new process technologies

¹⁹Gordon Moore, Intel’s cofounder, famously observed that the number of transistors on a chip doubled every eighteen months (Moore 1965). This regularity later slowed to doubling every two years.

often result in higher defect rates and lower “yields,” the fraction of chips on a wafer that function correctly. In spite of these challenges, the benefits of increased die per wafer and better performance have outweighed the costs of yield reductions, which improve as the fabrication technology matures. Given the benefits of smaller line widths, semiconductor manufacturers have steadily moved toward newer technology. This is apparent in Figure 1 for Intel processors and can be seen even more clearly in Figure 2, which plots the technology composition of sales at Taiwan Semiconductor Manufacturing Company (TSMC), the largest contract semiconductor manufacturer.

There are a number of options regarding the chemical compounds used to create the transistors themselves and how the transistors are arranged to implement logical functions. The most common technology, called complementary metal-oxide semiconductor (CMOS), a silicon-based chemical process, accounted for 97 percent of worldwide semiconductor production in 2008.²⁰ We therefore restrict our analysis to CMOS and refer to each combination of wafer size and line width as a “process technology” (e.g., 200mm wafer, 180nm line width).

In our analysis we must define the set of technological characteristics that influence the price of a given wafer. To guide this choice, we have consulted pricing models used by engineers to estimate production costs. Kumar (2008) presents a wafer cost model based on wafer size, line width, and logic family. The commercial cost estimation firm IC Knowledge distinguishes wafer cost estimates by wafer size, line width, logic family, number of polysilicon layers, and number of metal layers.²¹ All of these discrete characteristics are observable in our data, allowing us to compare prices across suppliers of technologically identical products in the analysis below.

2.2 Offshoring and the Foundry Business Model

In the early 1970s nearly all semiconductor producers were vertically integrated, with design, wafer fabrication, packaging, testing, and marketing performed within one company. Firms that perform both design and wafer fabrication are referred to as integrated device manufacturers (IDM). By the mid-1970s, IDMs began moving packaging and test operations to East Asia to take advantage of lower input costs (Scott and Angel 1988, Brown and Linden 2006).²² In spite of offshoring these relatively simple steps in the production process, firms maintained the more complex wafer fabrication operations in the home country.

As wafer fabrication technology advanced, however, it became prohibitively costly for younger and smaller semiconductor firms to stay at the frontier of process technology. The cost of build-

²⁰Share of wafer starts reported in SICAS Semiconductor International Capacity Statistics. Other transistor arrangements, such as bipolar logic, and other chemical processes, such as gallium arsenide (GaAs) or silicon germanium (SiGe), generally focus on niche markets for high-frequency, high power, or aerospace devices, rather than the storage and computational logic products comprising the majority of the CMOS market.

²¹See <http://www.icknowledge.com/>. Polysilicon and metal layers are used in the construction of transistor “gates” and interconnections.

²²Igami (2010) and (2013) analyses the dynamics of offshoring in the hard disk drive industry that likely parallel those in the semiconductor packaging and testing industries during this time period.

ing a fabrication facility has increased nearly 18 percent per year since 1970 and now stands at \$4.2 billion (IC Knowledge 2001, Global Foundries 2009). Consequently, during the middle of the 1980s, younger and smaller firms began contracting with larger U.S. and Japanese IDMs to produce some of their more advanced designs in the latter’s existing facilities (Hurtarte, Wolsheimer and Tafoya 2007). Around the same time, new contract manufacturing firms sprang up overseas that were entirely dedicated to manufacturing wafers designed by other parties. These firms, operating principally in Asia, are known as wafer “foundries.” Taking advantage of these new overseas facilities, a number of young U.S. semiconductor firms began outsourcing all of their wafer fabrication. These factoryless goods producers, which have little or no in-house manufacturing capability, are called “fabless” firms (Bayard et al. 2013, Bernard and Fort 2013). In general, fabless firms perform chip design and layout, and use foundries and other contractors for mask production, wafer fabrication, packaging, and testing.

The fabless business model has grown quickly over the last 30 years. It now accounts for about a quarter of total semiconductor industry revenue, as shown in Figure 3.²³ Some of the most prestigious U.S. chip makers, such as Fortune 500 firms Broadcom and AMD, are fabless firms. Along with the shift from an integrated manufacturer to a foundry business model came a shift in production capacity toward Asia, where most large foundries are located. Table 1 shows how the share of worldwide foundry capacity has evolved in the last decade.²⁴ In 2000, the majority of foundry capacity was already in Asia, mainly Taiwan. Since then, the share of capacity in Asia as a whole has only increased modestly, but there has been a notable shift in capacity within Asia. In particular, China has more than tripled its share of foundry capacity, largely at the expense of the industry leader, Taiwan.

2.3 Foundry Production vs. Memory and Processor Markets

Economists have devoted substantial attention to studying semiconductor production in an effort to uncover the sources of rapid constant-quality price declines observed for high-tech products such as computers (Berndt and Rappaport 2001) and communications equipment (Doms 2005, Byrne and Corrado 2012). Attention has focused on the most important semiconductor components of computers, namely microprocessors (Dulberger 1993, Grimm 1998, Doms et al. 2003, Holdway 2001, Flamm 2007) and memory chips (Flamm 1993, Grimm 1998, Aizcorbe 2002).

However, general-purpose microprocessors and memory chips account for a minimal share of the market studied in this paper. Foundries instead specialize in custom chips for specific models of electronic devices such as cellular phones, hard drives, automobiles, and many others. These

²³Note that the shares in Figure 3 are likely to understate the extent of fabless production activity because it counts only companies that derive 75 percent or more of their semiconductor revenue from fabless production. Many companies not counted as fabless, such as Texas Instruments, nevertheless rely heavily on foundries.

²⁴The sharp increase in European capacity from 2008 to 2010 marks the founding of Global Foundries, which was the fab division of former integrated device manufacturer AMD.

Application-Specific Integrated Circuits (ASICs) have been the subject of limited previous research and differ from memory and processors in important ways.²⁵ ASICs are produced in smaller batches, each model requires a substantial investment in design, and they are more likely to be produced using technology one generation or more behind the leading edge.²⁶ The most important characteristic for our analysis is the custom nature of each ASIC model. The uniqueness of each design generates substantial fixed costs of producing a given chip at a particular foundry, which we argue below drives a substantial portion of the price dispersion across suppliers of otherwise identical products.

3 Price Dispersion Across Countries

We use a new database of semiconductor production transactions to measure price dispersion across supplying countries. The database provides information on all technological characteristics that are relevant to product pricing. In a standard hedonic framework, we show that semiconductor wafers produced in China sell at a 17% discount compared to otherwise identical wafers produced in Taiwan. However, this average difference masks a dynamic process in which the Chinese discount for a particular technology starts out very large and then declines over time.

3.1 Semiconductor Wafer Price Data

Information on processed semiconductor wafer prices comes from a proprietary database of purchases from foundries, collected by the Global Semiconductor Alliance (GSA), a nonprofit industry organization. The dataset consists of 6,916 individual responses to the Wafer Fabrication & Back-End Pricing Survey for 2004-2010.²⁷ The survey accounts for a representative sample of about 20 percent of the semiconductor wafers produced by the foundry sector worldwide.

The GSA dataset is unique in the amount of detail it provides for contract manufacturing of a high-technology product. For example, it includes information on the technological attributes that industry analysts and engineers report as being the key price-forming characteristics of wafers. These are logic family, line width, wafer size, and layers, as discussed in Section 2. In addition, GSA reports the location (country) of the foundry for each transaction and the price paid. This information allows us to examine how average prices vary by foundry location after controlling for all relevant technological characteristics determining the nature of the manufacturing services being purchased. An important limitation of the GSA data is the lack of firm identifiers. That is, we see information on individual transactions, including the country in which the producing

²⁵Efforts to construct price indexes for the custom devices produced by foundries have been relatively limited, in large part due to data limitations (Aizcorbe 2002, Aizcorbe, Flamm and Khurshid 2002).

²⁶Note the long persistence of technology nodes in Figure 2.

²⁷The survey began in 2001, but was substantially revised in 2004, so the sample we received from GSA begins in the first quarter of that year.

foundry is located, but not the identity of the producing firm or buyer of wafer fabrication services. In what follows, we will focus our analysis on the two largest sources of contract semiconductor manufacturing services, Taiwan and China, which account for 49 percent and 22 percent of global production capacity respectively in 2010 (Table 1). During our sample period, the vast majority of each country's output was accounted for by a single firm, Semiconductor Manufacturing International Corporation (SMIC) in China and Taiwan Semiconductor Manufacturing Corporation (TSMC) in Taiwan.²⁸ Thus, for the two largest supplying countries, the lack of producing-firm identifiers is not too limiting.

As mentioned in Section 2, in an effort to focus our analysis on the ASIC market that dominates foundry production, we only analyze CMOS wafer transactions. Our sample thus omits niche markets for chips used in aerospace and high-power applications that require different production techniques. For the same reason, we also limit our sample to omit observations for products produced in countries that focus heavily on non-ASIC products: Europe, Japan, and Korea.²⁹ Compared to the countries in our sample, foundries producing primarily in the dropped countries derived a much larger share of their revenue from analog devices (41.5% vs. 11.0%), a larger share from discrete devices and memory products (19.0% vs. 7.7%), and a much smaller share from computational logic devices characterizing the ASIC market (35.3% vs. 70.3%).³⁰ The remaining countries in our sample accounted for 86.9% of foundry revenue in 2010.

Descriptive statistics for key variables in the GSA database are shown in Table 2.³¹ We observe an average of 223 transactions per quarter. All figures are weighted using data on shipments by country and technology from the Pure Play Foundry Market Tracker database produced by market research firm IHS iSuppli.³² The changing technological characteristics of the fabrication process are evident in the statistics for wafer size and line width. Pilot production lines for 300 mm wafers were first introduced in 2000, and the share for this emerging technology rises from 3 percent of contracts to 20 percent of contracts over the sample period. Similarly, in general, newer line widths increase in share over time: 65 nanometer technology reached volume production in the overall semiconductor industry in 2006 and slowly gained share in the foundry market, accounting for 8 percent by 2010; 45 nanometer contracts were just emerging in 2010. Meanwhile, older technologies, with line widths larger than 250 nanometers, dwindle in prominence from 40 percent in 2004 to

²⁸These statements are based on IHS iSuppli data, but our confidentiality agreement prohibits us from quoting precise figures at the firm level. Note also that although TSMC operated one small production facility in China during this time period, it accounted for a very small share of Chinese output.

²⁹We also estimated all analyses including these countries, and the results were nearly identical (available upon request). As the inconsistent product mix in Europe, Japan, and Korea clouds the interpretation, we prefer the restricted sample presented here.

³⁰Authors' calculations based on market data from IHS iSuppli.

³¹See Appendix A for details on data sources and data cleaning.

³²Appendix A describes how weights were constructed. We also implemented all analyses weighting all transactions evenly (available upon request). The unweighted results exhibited somewhat larger price gaps across countries, larger declines in price gaps over time, and larger differences between indexes. Thus, in all respects, the weighted results presented here are conservative, and importantly are more representative of the industry.

33 percent in 2010. The number of metal and mask layers per wafer also rose somewhat over the period studied, reflecting a trend toward foundries handling increasingly complex designs.

3.2 Cross-Country Price Dispersion

Given these detailed data, we move to investigating the cross-country variation in wafer prices in a simple hedonic regression framework. First, we regress the log of the wafer price for each transaction on quarter indicators and indicators for the location of production, with Taiwan as the omitted category. The results in column (1) of Table 3 demonstrate that there are huge unconditional price differences across suppliers. Focusing on the two largest suppliers, wafers produced in China cost 25.7 percent less than those produced in Taiwan.³³

These large unconditional price differences can partly be explained by differences in product attributes. Column (2) adds indicators for each wafer size and each line width, layer controls, and the size of the transaction.³⁴ The signs of all regression coefficients are intuitive. The omitted category is for 200mm wafers with 180nm line width, produced in Taiwan. More advanced production technologies, with larger wafers and smaller line widths, command higher prices. For example, the coefficient 0.671 implies that a 300mm wafer is 96 percent more expensive than an otherwise identical 200mm wafer produced in the same quarter. Similarly, a wafer with 150nm line width was 18 percent more expensive than an otherwise identical wafer with 180nm line width in the same quarter. Wafers involving more layers are more expensive, as these require more raw materials and more steps in the production process. Larger orders also command a small bulk discount. All of these coefficients are very precisely estimated and highly statistically significant.

When controlling for this exhaustive list of technological attributes, the estimated price differences across suppliers change, indicating differences in the characteristics of products produced by various suppliers. The measured China discount falls in magnitude from 25.7 percent to 17.0 percent, indicating that China produced more trailing edge products than Taiwan on average. When controlling for technological attributes, substantial price differences remain across the other suppliers as well. Singaporean and Malaysian producers also exhibit discounts relative to Taiwan, while producers in the U.S. charge higher prices for otherwise identical wafers. These large price differences across supplying countries are precisely estimated and very robust to changes in specification. Column (3) estimates a more flexible specification, with indicators for each wafer size and line width combination, with almost no change in the coefficients on the country effects.

Even if the physical product attributes are identical across suppliers, there may be *unobserved* differences across suppliers in the quality of the overall production service. This is a long-known concern, dating at least to Carlton's (1983, 1986) seminal work. One potential example of such a

³³ $\exp(-0.297) - 1 = -0.257$.

³⁴Note that hedonic regression estimates reflect a joint envelope of demand and supply functions (Rosen 1974). Including an order size regressor to control for non-linear pricing is reasonable in this context, unlike when trying to separately identify a demand or supply curve. That said, all results are similar when omitting this control.

difference is the “yield,” the fraction of chips on a wafer that function properly. A supplier with systematically lower yields would likely provide a discount to compensate the customer for the fact that each wafer represents fewer usable chips as compared to a supplier with higher yields. However, yield differences are unlikely to explain the results in Table 3. First, the vast majority of yield improvements occur during the engineering phase of product development. This is the initial phase of production during which volume is gradually increased and equipment calibrations are fine tuned. We omit data from the engineering phase, instead considering this portion of the production process to be part of the large startup costs modeled in the next section. Second, managers at fabless firms report that they receive similar service from the major Taiwanese and Chinese foundries in terms of timeliness and yields.³⁵ Third, to the extent possible, we have analyzed yield data directly and found i) very little variation in yields across supplying countries and ii) no evidence of lower yields at foundries with lower prices.³⁶

Another potential source of differences in manufacturing service quality relates to the intellectual property, design tools, and engineering support each foundry provides to its customers to help facilitate the transition from chip design to the foundry’s manufacturing process. It is possible that Taiwanese foundries provide better tools than their Chinese competitors, which could partly explain the observed difference in wafer prices between the two countries. In the absence of data on the quality of these tools and other services, we are unable to examine this hypothesis directly.

However, while the hedonic analysis reported in Table 3 does not allow us to distinguish between pure price dispersion and unobserved heterogeneity, we will show how to use the *dynamics* of observed price dispersion to help distinguish between these interpretations. An approximately constant difference in quality between China and Taiwan suggests a constant difference in price. In contrast, as we show below, important frictions in the contract semiconductor manufacturing market imply dynamic evolution in price dispersion, in which the price gap for a particular technology falls over time. When a new process technology is introduced in the foundry market, industry leading firms in Taiwan begin production at least 8 quarters before Chinese foundries.³⁷ In Section 4, we present a model in which this kind of sequential entry coupled with costs of switching suppliers leads to i) pure quality-adjusted price dispersion across suppliers, consistent with the results in Table 3, and ii) closing price gaps between the leading and following suppliers as time elapses after the follower’s entry. The following subsection investigates the price dynamics in the data, and confirms this prediction of a closing price gap between China and Taiwan.

³⁵We had extensive conversations on this issue with a design manager at LSI Corp, a major U.S. fabless firm, with the CEO and Managing Director of a large Korean fabless firm, and with the GSA Supply Chain Performance Working Group, an advisory panel consisting of executives and managers from various fabless firms.

³⁶Appendix B presents our analysis of these data.

³⁷Authors’ calculations using IHS iSuppli data. Observable delays range between 8 and 12 quarters, though Taiwanese entry is censored for many technologies. Potential reasons for China’s delayed entry include a relative lack of expertise in introducing new process technologies into the foundry context and various restrictions on the export to China of advanced wafer fabrication equipment from manufacturers in the U.S., Europe, Japan, and Taiwan (EE Times 1998). In this analysis, we simply take the sequential entry as given.

3.3 Dynamics of Price Dispersion

To examine the dynamics of price dispersion, we first restrict attention to the two largest supplying countries, Taiwan and China, which account for 77 percent of total output during our sample period.³⁸ Column (4) of Table 3 demonstrates that this sample restriction has little effect on the China-Taiwan average price gap when flexibly controlling for process technology.

As in the hedonic specifications of Table 3, we seek to measure price gaps for technologically similar products, so we first calculate the difference in wafer prices charged by Chinese and Taiwanese producers, *within* each technology (wafer size x line width) x quarter cell. We also calculate China-Taiwan differences in average numbers of layers and order sizes within each cell to allow us to control for any differences in these continuous attributes. For each technology we measure the elapsed time since Chinese entry, and examine China-Taiwan price gaps for each technology in each quarter following Chinese entry.³⁹

The results are shown in Figure 4. The dashed gray line plots raw quarterly price differences for the process technology with the largest sales during our time period, 200mm wafers with 180nm line width. Because our data begin in Q1 2004, and China entered this market in Q3 2002, we first observe the price gap 6 quarters following Chinese entry. Although the pattern is somewhat noisy, it is clear that the average price gap closes substantially over the life of this technology, from an initial gap of around \$600 to around \$150 more than 5 years following Chinese entry. This pattern applies to other technologies with smaller sales as well. The black line in Figure 4 plots the price gaps averaged across all technologies in each quarter following Chinese entry and exhibits quite consistent declines in the gap between Chinese and Taiwanese prices.

By averaging across technologies, however, this pattern could reflect the changing mix of technologies over time. Only the newest technologies are observed just after Chinese entry, while older technologies first appear in the data many quarters after Chinese entry. To ensure that compositional changes are not driving the results, we regress the quarterly price gaps on technology fixed effects and the time since Chinese entry for the relevant technology. We thus identify *within-technology* variation in price gaps over time, which ensures that changing composition of technologies does not influence the findings.

The results are presented in Table 4. The dependent variable is China's price less Taiwan's price in a technology x quarter cell, so negative numbers represent lower relative prices in China. Column (1) includes the time since Chinese entry and technology fixed effects. There is a strong positive trend, indicating that China's price discount eroded over time within technology. Column (3) includes a quadratic term in the time since Chinese entry to capture the apparent concave shape of the price gaps in Figure 4. The closing price gap and concavity are confirmed in the

³⁸Other low-cost suppliers such as Malaysia are simply too small a part of the market to provide sufficient data to implement this analysis.

³⁹See Appendix C for details of measuring entry timing.

regressions. Columns (2) and (4) introduce controls for layers and order size, without much effect on the estimated trends. The dotted line on Figure 4 shows the estimated quadratic relationship from the within-technology regression in column (4) of Table 4, evaluated at the sample average of the technology fixed effects. The close similarity with the other series confirms that the observed closing price gap is in fact a *within-technology* phenomenon, and that the average China discount levels out at approximately \$150, 20 quarters after Chinese entry.

This dynamic pattern of price differences is unlikely to be driven by unobserved differences in products or services across Chinese and Taiwanese suppliers. The price gaps start out large and then converge for each new process technology, so constant differences across suppliers or differences that evolve over *calendar time* for all technologies cannot explain the observed pattern. Thus, steady improvements in the quality or reliability of China’s production service may explain price differentials *across* technologies, but is unlikely to account for the sharp, *within* technology dynamics we observe. This also rules out explanations related to brand recognition, customer service, intellectual property rights protection, tax policy, and other factors that might make Chinese producers more attractive over time.

Changes in unobserved quality can only explain the observed price dynamics if the quality improvements happen systematically within each technology. The most natural of these explanations relates to yields. In particular, Chinese foundries may enter each process technology with extremely low yields compared to Taiwanese foundries, and the yield differences may close over time, explaining the converging price gap. However, the price gaps shortly after Chinese entry are so large that if they were driven by yield differences, they would be observed even with the rough yield measures examined in Appendix B. As mentioned above, we find no anecdotal or quantitative evidence of lower yields at Chinese foundries as compared to Taiwanese foundries. Thus, unobserved heterogeneity in the quality of products or customer service across suppliers is unlikely to account for the observed price differences across suppliers.

Instead, the empirical patterns documented in this section likely reflect, at least in part, pure (quality-adjusted) price dispersion. In the next section, we discuss how this dispersion might arise in the semiconductor fabrication market.

4 Model of Supplier Price Dispersion

We have argued that quality-adjusted price dispersion likely exists in this market. In this section we rationalize this price dispersion and its dynamics in a model featuring substantial costs of switching suppliers. These costs are quite large in the semiconductor fabrication industry, and they provide the friction underlying equilibrium price dispersion. In addition, whereas the previous section documented the *presence* of quality-adjusted price dispersion, the model suggests a method for measuring its *magnitude*. In general, this is a challenging task. Notwithstanding our controls for

product attributes, the overall quality of the production service likely still differs across suppliers in ways that are difficult to observe. We show that under certain conditions, there is a straightforward correction for unobserved heterogeneity.

The following two subsections summarize the environment and confirm that the model rationalizes the pattern of diminishing price differentials observed in the data. In Section 4.3, we show that long-run price differentials fall to the point where they reflect only differences in unobserved quality. This suggests that one can use estimates of observed price differences later in a product's life cycle to adjust for unobserved heterogeneity in the quality of products or service. We implement this correction in the semiconductor context in Section 4.3.

4.1 Framework

The model has three periods. There are two types of agents in the market – buyers and manufacturers of an intermediate good. A cohort of buyers of mass one enters in each of the three periods. The period-1 cohort is present in periods 1 and 2, the period-2 cohort enters in period 2 and remains through period 3, and the period-3 cohort is present in period 3.⁴⁰ We assume that buyers have inelastic demand such that each buyer will purchase the intermediate good from one of the suppliers.

Even though buyers purchase the same physical input (i.e. the same wafer size, line width combination) from both suppliers, we assume there are details of the production process that have to be tailored to a buyer's unique design (e.g. the manner in which transistors are arrayed on the wafer). In other words, for a given wafer size and line width pair, there is heterogeneity across buyers in chip design, and some designs are more difficult to fabricate than others. Formally, we follow in the spirit of Klemperer (1995) and assume that design complexity, y , is distributed uniformly from 0 (lowest quality) to 1 (highest quality). This heterogeneity across buyers would be unobservable to an econometrician who has data only on the physical wafer size and line width. In this sense, the model allows for price dispersion that reflects unobserved heterogeneity.⁴¹

Turning to the manufacturers, Firm A (Taiwan) is the leader and is present in the market from period 1 onward. Firm B (China) is the follower; it joins the market in period 2. We assume Firm A is at the technology frontier. For example, in the wafer market, it is thought that Taiwan's fabrication plants have software tools that enable them to produce more complex designs with less assistance from the buyer. Accordingly, although Firm B can fabricate any chip, the customer must pay a cost to monitor and consult with this supplier. To be precise, we assume that buyers

⁴⁰The dynamics of price dispersion that we wish to study represent an out-of-steady-state phenomenon that is most easily handled in a finite-horizon framework. See Beggs and Klemperer (1992) and To (1996) for analyses of the steady-state equilibrium in infinite-horizon models.

⁴¹As we show below, this heterogeneity across designs is important to our analysis only insofar as it affects the quality of the fabrication service being purchased from a foundry. The relationship between chip design and final retail price has no implications for our analysis.

who purchase from Firm B pay a per-period premium that is increasing in design complexity, τy (where $\tau > 0$). The magnitude of τ indexes the extent of dispersion in the quality of manufacturing service.⁴²

The relative efficiency of Firm A renders it a clear advantage. However, we assume that Firm B faces a lower unit cost of production, c^B . Specifically, we assume that both firms have constant unit costs, and that $c^A > c^B$. This is intended to reflect the difference in input costs across suppliers in Taiwan and China.

When a buyer initiates production with a supplier, it must pay a startup cost, s . This cost has to be paid again if the buyer switches suppliers. If a buyer purchased from Firm A in period $t - 1$, it would pay a price, p_t^A , to remain with Firm A in period t ; or it may switch to Firm B , in which case it pays $p_t^B + \tau y + s$, where p_t^B is Firm B 's posted price this period, τy is the monitoring cost, and s functions like a switching cost.

A switching cost represents a significant and realistic trading friction in this market. A concrete example of such a cost is the mask set, which is used in fabricating the transistor network on the silicon wafer (see Section 2). Due to technical differences in production processes, masks cannot be transferred across suppliers and must be re-fashioned, at a price of over \$1 million, if production is moved to another facility. This led one industry association to state, "The time and cost associated with [switching] tend to lock customers into a particular foundry."⁴³ Other examples of startup costs include the many chemical and mechanical adjustments and calibrations to manufacturing equipment that are implemented during the engineering phase of production for a particular chip design. These adjustments must be redone when moving to a new production line.⁴⁴

Lastly, following much of the literature on costly switching, the model prohibits price discrimination. This restriction is roughly consistent with wafer supplier contracts, which limit a supplier's freedom in charging appreciably different prices across its customers.⁴⁵ Thus, we assume the price p_t^A (p_t^B) applies to all Firm A (B) buyers in period t .⁴⁶ Similarly, we assume contracts last only one

⁴²Changes in the relative yield of Chinese and Taiwanese producers during the life of a product would involve changes in τ . Given the empirical evidence against this phenomenon in Section 3.2, we assume τ is constant. This still leaves open the possibility that τ declines steadily *across* product generations as China becomes more experienced. But since τ likely remains reasonably constant *within* any one product generation, our assumption seems appropriate.

⁴³This quotation is taken from the Common Platform Alliance, an industry group consisting of a few large chip manufacturers such as IBM and Samsung that advocate for a "common platform" that would standardize aspects of semiconductor production technology. However, this alliance has not yet had a material impact on standardizing mask sets (McGregor 2007). We are grateful to Ross Goodman for insightful discussions on this topic.

⁴⁴An additional example is negotiating a new supplier agreement, which requires a vast amount of technical information to be exchanged. These talks can absorb much of the attention of senior management (c.f. Allan (2002)).

⁴⁵We have obtained a handful of these contracts. A representative agreement in terms of the handling of price discrimination, between Marvell Technology and Taiwan Semiconductor Manufacturing Corporation (TSMC), states that "TSMC shall calculate an average price for such processes in use at all of TSMC's plants," and if the buyer's price "deviates, up or down, by more than three percent from the [average price]," the buyer's price will be adjusted in the direction of the average price. Note that this agreement does not commit TSMC to a particular price path over time. The contract merely restricts price discrimination in a given period, consistent with the model's assumptions.

⁴⁶We suspect that the model's predictions do not hinge on this choice. Relative to the baseline with no discrimi-

period and are enforceable within each period, based on the observation that supplier contracts in this industry enumerate measurable requirements for buyers and suppliers and specify sanctions if either party fails to meet the specified requirements.⁴⁷

We wish to use the model to help guide our efforts to construct a quality-adjusted price index that accounts for pure price dispersion and unobserved heterogeneity across suppliers. As a first step, we want to ensure that the model is in fact broadly consistent with the key features of semiconductor prices described earlier, a task to which we now turn.

4.2 Solution

We solve for a subgame-perfect, pure-strategy Nash equilibrium using backward induction. Details of the solution are presented in Appendix D; here we briefly review the solution procedure and describe the results.

We conjecture that the lagging supplier, Firm B , sells to all buyers in the period-2 cohort whose design complexity y falls below a threshold, $\hat{y} \in (0, 1)$, and confirm this allocation in equilibrium. The period-2 cohort of buyers remains in period 3, so we refer to the measure \hat{y} of buyers as Firm B 's "customer base" at the start of period 3, and $1 - \hat{y}$ as Firm A 's customer base. Note that buyers in the period-2 cohort would have to pay s in period 3 to switch suppliers. Taking \hat{y} as given, we solve the period-3 problem, recovering equilibrium pricing schedules as functions of \hat{y} . We then solve the period-2 problem to recover \hat{y} , which is the value of y for which a period-2 cohort buyer is just indifferent between starting production with Firm A and Firm B .

The discrete switching cost, s , introduces non-concavities into the payoff functions that are difficult to handle analytically, so we solve the model numerically.⁴⁸ We calibrate the model's four structural parameters, c^A , c^B , τ , and s , to target the long-run price levels and market share and to minimize supplier switching, in light of testimony by industry analysts that this is rare. Details are presented in Appendix D.⁴⁹

nation, Firm A is likely to charge a *higher* period-2 price to its period-1 locked-in cohort if discrimination is allowed. At the same time, it is likely to post a *lower* price for period-2 entrants, relative to its strategy in a model with no discrimination. Hence, we conjecture that a reasonable calibration of a model with price discrimination will still yield a relatively high *average* Firm A price in period 2 and that dispersion will diminish over the product's life.

⁴⁷While the model reflects a kind of hold-up mechanism in which startup costs sunk by buyers give the incumbent supplier market power, one consequence of restricting price discrimination is to limit a supplier's ability to capitalize on that market power on a buyer-by-buyer basis.

⁴⁸A switching cost makes it possible for a firm to increase its price without reducing sales. In this range of (high) prices, it concedes all new entrants to its competitor, but it does not lose any of its incumbents because $s > 0$. Thus, profits may decline (from a local peak) as prices rise, but then begin to increase again in this range of relatively high prices. The literature has introduced various ways to (sufficiently) concavify the problem. Klemperer (1987) assumes a (potentially large) share of incumbents randomly "die" in each period. Others, such as Shin and Sudhir (2009), assume buyers' preferences over suppliers (roughly, y in our application) are re-drawn each period. Each of these approaches reduces the extent of lock-in and smooths the profit schedule. However, neither approach seems suitable to the market under study in this paper.

⁴⁹Nishimura and Friedman (1981) provide conditions under which a pure strategy equilibrium may exist if payoffs are not quasi-concave, but the conditions can only be confirmed *after* a calibration is chosen. This is why the model

Table 5 presents the equilibrium prices and market shares given our calibration. We wish to highlight a few results in particular. First, the model implies that the price gap between the leading and lagging supplier falls over the product’s life cycle. In the period of the follower’s entry, there is a price gap of \$258, and the gap falls to \$149 in the following period. The intuition for this closing price gap is the following. Since Firm *A* is a monopolist in period 1, it captures the entire period-1 cohort of buyers. Then in period 2, Firm *A* has an incentive to charge a high price, to exploit the fact that it has a large measure of customers from the period-1 cohort who have already paid the sunk startup cost for Firm *A* and hence are partially locked-in. Firm *B*, in contrast, has just entered and has an incentive to charge a low price to win customers from the period-2 cohort. In period 3, the incentives reverse. Firm *A* has a smaller customer base and has an incentive to charge a lower price in period 3 to attract new entrants, while Firm *B* has a large base to exploit by charging a relatively high price. Hence, the price gap is large in period 2 and closes substantially in period 3.⁵⁰

Though these dynamics are consistent with the data, note that the period-2 price differential is somewhat smaller than what we observe in the data. This occurs because models of this sort tend to overstate the degree of competition in period 2 relative to 3. Suppliers know they will increase period-3 sales if they attract period-2 buyers, which drives down overall price levels in period 2 (Farrell and Klemperer 2007). This reduces the absolute price difference in period 2.⁵¹ Although the model is too simple in this regard to fully replicate the magnitude of price dispersion, we are encouraged that it rationalizes both the presence of substantial price gaps across suppliers and the price dynamics we observed in the data, and that it does so within a quantitative environment that captures the salient characteristics of the contract semiconductor production market.⁵²

Table 5 also shows that the terminal-period price gap is extremely close to the price gap that emerges from an otherwise identical model with no switching friction ($s = 0$). We discuss this finding in the next subsection and describe its implications for our measurement strategy.

4.3 Accounting for Unobserved Heterogeneity

In a model with supplier switching costs, $s > 0$, price dispersion occurs for two reasons. First, buyers from Firm *A* are willing to pay a price in excess of Firm *B*’s posted price for the same

must be solved numerically.

⁵⁰If there were a period 4, we conjecture that Firm *B* would still compete relatively aggressively, as its customer base would be smaller than Firm *A*’s. However, the price differential would remain small relative to that in period 2, simply because Firm *B* would still have a substantial customer base it would like to exploit; it will not price so aggressively as to drive the price differential to period-2 levels.

⁵¹However, note that the period-2 *log* price differential is not too different from the observed differential, in spite of the fact that we do *not* target this in calibrating the structural parameters.

⁵²We suspect there is another reason why the price differential narrows more in the data than in the model. Taiwan’s current product competes both with that of China and with Taiwan’s next product generation. This will amplify competition late in the product’s life cycle, relative to period-3 competition in the model. A fuller model that accounts both for frictional price dispersion and endogenous product introduction seems worthy of future study.

product because they do not incur the monitoring cost, τy . This feature captures heterogeneity in the quality of services provided by suppliers that an econometrician would not observe. In the frictionless equilibrium, this constant difference in quality drives the observed price dispersion. Second, the switching cost partly locks in buyers, enabling suppliers to charge a higher price for any given product complexity, y . For instance, in period 2, Firm *A* retains low- y buyers from the period-1 cohort, even as new entrants with the same design complexity (y) obtain a lower price from Firm *B*. This feature captures the possibility of pure price dispersion between suppliers for identical products. Table 5 suggests that the influence of the switching cost on dispersion abates by the terminal period, with the price differential converging to that in the frictionless ($s = 0$) equilibrium.

This intuitive result can be seen more formally by inspecting the period-3 problem. Having solved the model numerically, we have confirmed that (under a plausible calibration) each supplier retains its customer base in period 3. Accordingly, we can impose that restriction and solve the period-3 problem analytically. Under these conditions, Appendix D shows that the difference in period-3 prices is given by⁵³

$$\Delta_3 = \Delta^* + (2(1 - \hat{y}) - 1) \frac{\tau}{3}, \quad (1)$$

where $\Delta_t \equiv p_t^A - p_t^B$ is the equilibrium difference in market prices in period t , and Δ^* is the static price difference that would emerge if the switching cost were suspended. Therefore, Δ^* exclusively reflects the influence of differences in the quality ($\tau > 0$) of the production service on price dispersion.

Equation (1) indicates that if \hat{y} is in the neighborhood of $1/2$, the terminal-period price gap approximates the frictionless dispersion. Intuitively, each supplier charges a higher price *level* than in the frictionless model, but the two suppliers' incentive to exploit their customer bases is roughly the same if $\hat{y} \approx 1/2$. Hence, the price *difference* reflects almost entirely the difference in the quality of the manufacturing service. For our calibration, it turns out that $\hat{y} \approx 1/2$. This is not surprising: an interior solution to the period-3 problem exists only if neither firm has such a large customer base that it wants to “gouge” its own buyers and neglect new entrants. Thus, for a terminal period equilibrium to be consistent with observable behavior, it will involve a value of \hat{y} that sufficiently far from both 0 and 1. In this sense, an estimate of \hat{y} in the neighborhood of one half is consistent with our expectations.

This finding is noteworthy because it suggests that, under certain conditions, there is a straightforward way to adjust the observed price differences for unobserved heterogeneity. Assuming the data are generated according to a model of costly switching and $\hat{y} \approx 1/2$, we can use the *observed* price dispersion near the end of the product life cycle to proxy for a portion of the price differential, Δ^* , that can be attributed to unobserved heterogeneity. One can then uncover estimates of pure

⁵³This also assumes that each supplier acquires a positive share of new entrants; i.e. the solution does not lie at a corner. This is also consistent with the calibrated equilibrium.

constant-quality price dispersion by subtracting the end-of-product-life price difference, Δ_3 , from the estimated price differentials earlier in the life cycle, Δ_2 .

We now apply this simple strategy to recover estimates of pure price dispersion, netting out the effects of unobserved heterogeneity. We take the “long run” to be roughly five years following Chinese entry, based on the length of typical semiconductor fabrication contracts. Therefore, drawing from the quadratic fit in Figure 4, our estimate of the frictionless price difference, Δ^* , is \$147.2. We then subtract this difference from those in earlier points in the product life cycle. For example, the average price gap 10 quarters following Chinese entry is \$373.5, so we estimate that 60.6 percent ($\frac{373.5-147.2}{373.5}$) of this observed price gap reflects pure price dispersion rather than unobserved heterogeneity in the quality of the product or transaction. These calculations make clear that both unobserved heterogeneity and pure price dispersion are important drivers of the observed price differences between suppliers of otherwise identical semiconductor wafer manufacturing services. Of course, one could consider more elaborate implementations of this approach, perhaps in which Δ^* varies. Our aim here is to show that there is a simple, and still reasonable, correction that statistical agencies can undertake.

There are two reasons to expect that we understate the amount amount of pure price dispersion (and hence overstate the amount of unobserved heterogeneity). First, the model tends to understate the degree to which the leader maintains its advantage (as discussed in Appendix D), so the true value of \hat{y} is likely smaller than 1/2. As a result, equation (1) implies that Δ_3 overestimates its frictionless counterpart, Δ^* , and thus overstates the extent of unobserved heterogeneity. It follows that we obtain a lower bound on the degree of pure price dispersion.

Second, we have taken Δ^* , the frictionless price differential, as our measure of price dispersion resulting from unobserved heterogeneity in service quality. This approach assumes that the extent of quality-related price dispersion is independent of the amount of frictions, given by s . This is unlikely to be the case, since the size of the friction influences buyers’ choice of supplier, and hence influences the amount of quality-related price dispersion across suppliers. Consider a case where the typical buyer in China’s customer base has a higher value of y in the presence of frictions, as compared to the frictionless case. This buyer then experiences a lower quality of service than it would in the frictionless case. This would, in turn, suggest that a larger quality adjustment is needed when $s > 0$. However, this is not what we find in our calibrated example, and seems unlikely in general. In period 2, the switching cost enables the leading supplier to sell to buyers with relatively low values of y in its own customer base. Thus, the switching cost drives *down* the typical value of y among Chinese buyers, implying that they pay relatively little to transact with China. If anything we overstate the quality adjustment and hence generate a conservative estimate of the remaining amount of pure price dispersion.

5 Implications for Price Measurement

The preceding sections document the presence of price dispersion across suppliers of otherwise identical semiconductors and suggest that these differences arise because of large costs of switching suppliers. In this section, we investigate the implications of this price dispersion for the measurement of constant-quality price indexes. We use the dynamics of price dispersion documented in Section 3.3 and the theoretical findings in Section 4.3 to calculate a price index adjusted for unobserved quality differences across suppliers. We then show that this quality-adjusted index is bounded above by a standard matched-model index and below by an average price index following Reinsdorf (1993). The latter two indexes are feasibly calculated using the data already collected by statistical agencies, and provide bounds for the true price index irrespective of the particular market structure driving potential price dispersion across suppliers.

5.1 Matched-model Index Construction

Standard index construction procedures track price changes for a consistent set of items over time in an effort to omit price variation due to changes in product characteristics (Bureau of Labor Statistics 1997). Most statistical agencies use a matched-model framework, which calculates price growth for each product, or “model” m , in each period t as p_m^t/p_m^{t-1} , and then averages the price growth across models using an index number formula such as Laspeyres, Fisher, or Törnqvist. The agency must then decide how to incorporate newly entering models for which price growth cannot be calculated, a process known as “linking.” This is particularly difficult when there is both unobserved heterogeneity and pure price dispersion between continuing and newly entering models. If the price level difference between the new and continuing models solely reflects quality differences, then it should be omitted from the index. If the price gap reflects pure price dispersion for identical goods, then it should be included in the index. For intermediate cases, only a portion of the observed price gap should be included.

Statistical agencies such as the U.S. Bureau of Labor Statistics (BLS) typically assume that the law of one price holds across models within a narrow product classification, so any observed discount is assumed to reflect lower quality (Triplett 2006). In practice, this assumption involves omitting the new model’s price in the period of entry and including the model only in the subsequent period, when its price growth can be calculated.⁵⁴ Thus, the standard linking procedure omits price declines incurred when newly entering models appear.

Nakamura and Steinsson (2012) have illustrated how this linking procedure works in the BLS International Price Program (IPP). Their analysis of micro data reveals that the IPP indexes effectively define each model as a “contract between a particular buyer and seller.” Mapping this approach to our context defines a model based on technology and supplying country, so otherwise

⁵⁴This is called the “link-to-show-no-price-change” approach in the price index literature (Triplett 2006).

identical products from different suppliers are treated as different models. We refer to this as the “technology-country” index. When Chinese suppliers enter the market for a particular semiconductor technology with lower prices than the existing Taiwanese products, the Chinese product is considered a separate model, and the standard linking procedure completely omits the associated price decline. Given our empirical evidence showing substantial pure price dispersion across suppliers, this standard linking approach is likely to be problematic.

Reinsdorf (1993) proposes an alternate approach that makes the opposite assumption regarding unobserved heterogeneity: for goods with identical observable characteristics, all price differences across suppliers are assumed to reflect pure price dispersion.⁵⁵ This approach assumes there is no unobserved heterogeneity across suppliers, so one first calculates average prices across suppliers and then measures growth in those average prices as \bar{p}^t/\bar{p}^{t-1} .⁵⁶ The growth in average prices is then averaged across groups of goods using an index formula. In our context, this procedure defines a model based on technology alone, so we refer to it as the “technology-only” index. By including the newly entering Chinese price in \bar{p}^t , the index includes the entire price decline associated with Chinese entry.

In order to visualize the difference between these two price indexes, Figure 5 shows an example for a particular technology (300mm wafer, 90nm line width) from the third quarter of 2007 to the end of 2008.⁵⁷ In this example, Taiwan is the sole producer until China enters in the second quarter of 2008, at a substantially lower price. The technology-only index, which is based on the *average price* across producing countries, falls below the Taiwanese price series immediately upon China’s entry and continues to do so as China’s market share increases over time. In contrast, the technology-country index averages *the rates of inflation* across countries and omits variation in the price level due to shifting sourcing patterns. Since the average Chinese price remains constant over time, while the Taiwanese price falls, the technology-country index lies above the Taiwanese price in spite of China’s lower price level.

We have already shown that the observed price differences between China and Taiwan reflect both pure price dispersion and unobserved heterogeneity. Thus, the assumptions for both indexes regarding unobserved heterogeneity are violated in this context. The technology-country index understates the true rate of price decline by failing to include the portion of the China discount reflecting pure price dispersion, and the technology-only index overstates the true rate of price decline by incorporating the entire China discount, including the portion that reflects lower unobserved quality. Thus, these two indexes bound the true constant-quality price index by imposing opposite extreme assumptions regarding the importance of unobserved heterogeneity. Note that this

⁵⁵Reinsdorf (1993) proposed this approach to study the effects on the Consumer Price Index of consumer substitution across retail outlets.

⁵⁶We calculate the average price across suppliers as the quantity-weighted arithmetic average.

⁵⁷This technology and time period were chosen because they yield a particularly clean comparison between the two index approaches. In this sense it is a special case, but the overall index results presented in Table 6 demonstrate that the relationship between the indexes shown in Figure 5 holds more generally.

bounding property applies irrespective of the sources or magnitudes of unobserved heterogeneity and pure price dispersion across suppliers.

5.2 Price Index Estimates

We demonstrate this bounding in two steps. First, we calculate matched model indexes reflecting the technology-country and technology-only approaches, defining technology as the combination of wafer size and line width. The technology-country index begins by measuring growth in the average price within technology \times source country cells, p_{ij}^t/p_{ij}^{t-1} , where i indexes technology and j indexes source country. The price growth is then averaged across cells using a Fisher index, though results using other index formulas are similar.⁵⁸ The technology-only index measures growth in the average price within technology cells, p_i^t/p_i^{t-1} , and then averages across technologies using a Fisher index. The results, shown in Table 6, confirm our prediction that the technology-only index falls more quickly than the technology-country index, with a difference of 1.2 percentage points per year. When restricting the sample to transactions produced by Taiwanese and Chinese foundries, each index falls a bit more quickly, but the difference in growth rate remains the same.

These two indexes bound the true quality-adjusted price index by making the two most extreme assumptions regarding unobserved quality differences across suppliers. While this bounding approach applies to any general market, we can be more precise in this case using the results of Section 4.3. There we showed that the long-run price gap between suppliers reveals the effect of unobserved heterogeneity on the price gap. We adjust all Chinese prices upward by the long-run average price gap of \$147.2, documented in Figure 4. This is our proxy for the price difference resulting from unobserved quality. Having implemented this quality adjustment, we then proceed as in the technology-only index, assuming that adjusted gaps now reflect only pure price dispersion. The resulting index is shown in the rightmost column of Table 6, and all three indexes applied to Taiwanese and Chinese products are shown in Figure 6. The technology-country and technology-only indexes do in fact bound the quality-adjusted index. It is worth noting that, for reasons discussed in Section 4.3, our adjustment for unobserved heterogeneity is likely an upper bound. Accordingly, the quality-adjusted index is, if anything, biased toward the technology-country index. If the true contribution of unobserved heterogeneity is smaller than \$147.2, then the quality-adjusted index should be closer to the technology-only index.

⁵⁸The Fisher index is calculated as follows. First calculate Laspeyres and Paasche indexes, respectively, as $P_L^t = \sum_m s_m^{t-1} \frac{p_m^t}{p_m^{t-1}}$ and $P_P^t = \left[\sum_m s_m^t \frac{p_m^t}{p_m^{t-1}} \right]^{-1}$, where m represents each model, t is time (quarter), p is the average price for a given model, and s_m is the share of total output in time t accounted for by model m . As the Laspeyres index overstates the welfare-theoretic “true” price change and the Paasche understates it, the superlative Fischer index is a geometric mean of the two: $P_F^t = \sqrt{P_L^t P_P^t}$.

5.3 Practical Implications for Price Measurement

This exercise has important implications for international price measurement. Ideally, we would compare our price indexes to the analog in official statistics, but the BLS does not publish a similarly disaggregate index.⁵⁹ Still, our analysis is informative regarding the *process* of index calculation utilized in official indexes. The standard approach to price measurement results in upward biased price indexes. The results in Table 6 imply that the technology-country index is biased upward by at least 0.4 percentage points per year, comparing to the quality-adjusted index. As discussed above, this is likely a lower bound on the bias.

To get a sense for the practical implications of price index bias of this scale, we consider the following thought experiment: How would U.S. multifactor productivity measures be affected if a similarly sized bias applied to all imported intermediates, i.e. if overall imported intermediate materials prices were biased upward by 0.4 percentage points per year? Intuitively, overstating input prices results in understating input quantities, which in turn implies upward biased productivity growth. Assuming that imports are split between intermediate and final use in the same proportion as domestic production (the “import comparability assumption”), we estimate the share of materials inputs accounted for by imported intermediates. We then adjust the materials input price index from the BEA National Income and Product Accounts to account for upward bias and recalculate multifactor productivity. Under these assumptions, U.S. productivity growth would be biased upward by 0.066 percentage points per year during 2004-2010, implying that the official multifactor productivity growth measure of 1.5 percent per year during the time period is biased upward by 4.6 percent. The price index bias may be quite different outside the semiconductor manufacturing industry, but this exercise demonstrates that bias of this scale can have an important effect on productivity measurement.

These findings are closely related to those of Houseman et al. (2011), who examine how substitution toward suppliers in developing countries affects imported intermediate price indexes and productivity measures in the U.S. manufacturing sector as a whole. They attempt to identify quality-adjusted price dispersion by, in part, comparing prices across supplying countries for narrowly defined products in the BLS International Price Program micro data.⁶⁰ Their productivity bias measures are somewhat larger than ours, finding upward bias of 0.1 to 0.2 percentage points

⁵⁹The closest comparison index is the Bureau of Labor Statistics’ International Pricing Program (IPP) index for Harmonized System Code (HSC) 8542, electronic integrated circuits and microassemblies. This product category differs from our indexes in three respects: i) our data shows prices for wafer fabrication services purchased by chip makers around the world, while the BLS samples only U.S. importers, ii) Our data include only arms-length transactions while the IPP also includes intra-company transfers, and iii) the IPP index is much more aggregated than our indexes, as the IPP includes finished semiconductor chips and microassemblies in addition to processed wafers and die.

⁶⁰In the working-paper version of their study, (Houseman, Kurz, Lengermann and Mandel 2010), the authors also use the quality measurement procedures developed in Mandel (2010), which relies on prices and market shares to infer quality differences.

per year. The difference in findings could occur for a few reasons. First, the within product category comparisons in Houseman et al. (2011) may partly be confounded by unobserved heterogeneity, either in product characteristics or in service quality. In the semiconductor context, failure to account for such differences would result in a substantial overstatement of the constant-quality price gaps between Taiwan and China, as seen in comparing columns (1) and (2) in Table 3, and hence would overstate the bias from shifting sourcing. Second, there may be heterogeneity in the effects of shifting sourcing across industries, so our estimate from semiconductor wafer fabrication may not generalize to other manufacturing industries. Third, substitution in the semiconductor industry typically involves switches between offshore suppliers, as initial offshoring out of the U.S. occurred decades earlier. Shifts from U.S. to foreign producers that Houseman et al. focus on may involve different price gaps and hence different biases. Testing these conjectures will require future research on other industries using similarly detailed data to those we examine in Section 3.

Our findings suggest two approaches to addressing the price measurement problems in offshore intermediate markets. First, we have confirmed the practical value of collecting information on product characteristics, even within very narrowly defined product classifications such as the 10-digit Harmonized System. Table 3 shows that controlling for a few observable product characteristics cuts the price gap between Taiwanese and Chinese producers by 37 percent. Our data also reveal substantial price differences between suppliers even for technically identical products. Using the dynamics of price adjustment in this market, we separate the influence of unobserved heterogeneity in products or services across suppliers from pure price dispersion. Both sources of price variation are quantitatively important, particularly early in the life of a product. Our procedure for adjusting prices to account for unobserved heterogeneity in calculating a quality-adjusted price index could be applied to other intermediate markets where there are nontrivial costs of switching suppliers, as long as one can observe supplier entry and product turnover.

Second, the bounding procedure we have described can be implemented with much less detailed information on product characteristics and does not rely on the price dynamics we used to generate the quality-adjusted index. Statistical agencies such as the BLS can implement the bounding procedure using the data they already collect. For example, the BLS groups closely related items into “detailed product categories” (Bureau of Labor Statistics 1997) that serve the same role as the “models” in the bounding analysis above. Thus, using existing BLS micro data, one can calculate average prices within detailed product category and use those average prices to generate price relatives, as in Reinsdorf (1993) and our technology-only index. This index, in conjunction with a standard matched model index, provides bounds on the true quality-adjusted price index.

6 Conclusion

This paper uses a novel database of contract semiconductor fabrication transactions to study price and quality differences across suppliers located in different countries. We document large price differences across suppliers of observably identical products, and show that the price gaps converge over the life of a given product generation. We provide strong evidence that these price differences reflect elements of both pure price dispersion and unobserved quality differences. These features present serious challenges when attempting to construct constant-quality price indexes, as the analyst must determine what fraction of an observed price difference reflects unobserved quality difference. We use a novel approach to measuring unobserved quality differences across suppliers that relies on price dynamics across leading and following suppliers. This allows us to construct a quality-adjusted price index that we use to show that the standard index is biased upward. We additionally propose a simple approach to bounding the true constant-quality index that uses the same data as the standard index and applies irrespective of market structure.

In the semiconductor context we argue that large costs of switching suppliers drive much of the observed price dispersion, consistent with our model of switching costs and sequential entry. While buyers in the semiconductor industry face particularly large technological switching costs, we suspect that these frictions are not unique to this industry. The costs of initiating new overseas supplier relationships are likely substantial in other contract manufacturing industries as well. Building management relationships, initiating communication channels, and customizing production processes to a buyer's design may consume nontrivial amounts of time and resources. Thus, we expect other contract manufacturing industries to exhibit price dispersion and similar dynamics to those we have documented here.

It is likely that international sourcing will become more commonplace over time as trade barriers and transportation costs fall and communication technologies improve, facilitating further fragmentation of production (Fort 2013). Moreover, existing offshore contract manufacturing services already being purchased by domestic factoryless manufacturers such as fabless semiconductor firms will be treated as service imports following the 2017 NAICS revision (Office of Management and Budget 2011, Doherty 2013), and will need to be priced accurately. These developments all suggest that the measurement challenges we address will grow in scale and importance over time. Hence, it is essential that statistical agencies' price measurement procedures are robust to the presence of product and service heterogeneity and pure price dispersion across suppliers.

References

- Abbott, Thomas A., “Price Dispersion in U.S. Manufacturing: Implications for the Aggregation of Products and Firms,” *CES Discussion Papers*, 1992, (CES 92-3).
- Aizcorbe, Ana, “Price Measures for Semiconductor Devices,” *Board of Governors of the Federal Reserve System (U.S.), Finance and Economics Discussion Series*, March 2002, 2002-13.
- , Kenneth Flamm, and Anjum Khurshid, “The role of semiconductor inputs in IT hardware price decline: Computers vs. Communications,” *Board of Governors of the Federal Reserve System (U.S.), Finance and Economics Discussion Series*, 2002, 2002-37.
- Allan, Gabriel, “A Marriage of Unequals,” *Electronics Design Chain*, Summer 2002, 1, 17–20.
- Baldwin, Richard E. and Paul R. Krugman, “Market Access and International Competition: A Simulation Study of Random Access Memories,” in Robert C Feenstra, ed., *Empirical Methods for International Trade*, MIT Press, 1988, pp. 171–202.
- Bayard, Kimberly, David Byrne, and Dominic Smith, “The Scope of U.S. Factoryless Manufacturing,” *mimeo*, 2013.
- Beggs, Alan and Paul Klemperer, “Multi-period Competition with Switching Costs,” *Econometrica*, May 1992, 60 (3), 651–666.
- Bernard, Andrew B. and Teresa C. Fort, “Factoryless Goods Producers in the US,” *NBER Working Paper*, August 2013, (19396).
- Berndt, Ernst R. and Neal J. Rappaport, “Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview,” *American Economic Review*, 2001, 91, 268–273.
- Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov, “The Economics of Predation: What Drives Pricing when there is Learning-by-Doing?,” *mimeo*, 2011.
- , —, —, and Mark Satterthwaite, “Learning-by-Doing, Organizational Forgetting, and Industry Dynamics,” *Econometrica*, 2010, 78, 453–508.
- Boskin, Michael J., Ellen R. Dullberger, and Robert J. Gordon, “Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee from the Advisory Committee to study the Consumer Price Index,” in Dean Baker, ed., *Getting Prices Right: The Debate over the Consumer Price Index*, Sharpe, 1998.

- Brown, Claire and Greg Linden, "Offshoring in the Semiconductor Industry: Historical Perspectives," in "Brookings Trade Forum: 2005," Washington, DC: Brookings Institution Press, 2006.
- Burdett, Kenneth and Kenneth Judd, "Equilibrium Price Dispersion," *Econometrica*, 1983, 51, 955–69.
- Bureau of Labor Statistics, *BLS Handbook of Methods*, Chapter 15
- Byrne, David and Carol Corrado, "Prices for Communications Equipment: Rewriting a 46-Year Record," *mimeo*, 2012.
- Carlton, Dennis W, "Equilibrium Fluctuations when Price and Delivery Lag Clear the Market," *Bell Journal of Economics*, 1983, 14 (2), 562–572.
- Carlton, Dennis W., "The Rigidity of Prices," *American Economic Review*, 1986, 76 (4), 637–658.
- Doherty, Maureen, "Reflecting Factoryless Goods Production in the U.S. Statistical System," *mimeo*, 2013.
- Doms, Mark, "Communication Equipment: What has Happened to Prices?," in "Measuring Capital in the New Economy," Vol. 65 of *NBER Studies in Income and Wealth*, University of Chicago Press, 2005, pp. 323–362.
- , Ana Aizcorbe, and Carol Corrado, "When Do Matched-Model and Hedonic Techniques Yield Similar Measures?," *Federal Reserve Bank of San Francisco, Working Papers in Applied Economic Theory*, June 2003, 2003-14.
- Drozd, Lukasz A. and Jaromir B. Nosal, "Understanding International Prices: Consumers as Capital," *American Economic Review*, 2012, 102 (1), 364–95.
- Dulberger, Ellen R., "Sources of Price Decline in Computer Processors: Selected Electronic Components," in "Price Measurements and their Uses," Vol. 57 of *National Bureau of Economic Research Studies in Income and Wealth*, University of Chicago Press, 1993, pp. 103–124.
- Eaton, Jonathan, Marcela Eslava, David Jinkins, C.J. Krizan, and James Tybout, "A Search and Learning Model of Export Dynamics," *mimeo*, 2013.
- EE Times, "U.S., China at Odds on Fab-Gear Export," *EE Times*, April 1998.
- Farrell, Joseph and Paul Klemperer, "Coordination and Lock-In: Competition with Switching Costs and Network Effects," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization*, Vol. 3, North-Holland, 2007.

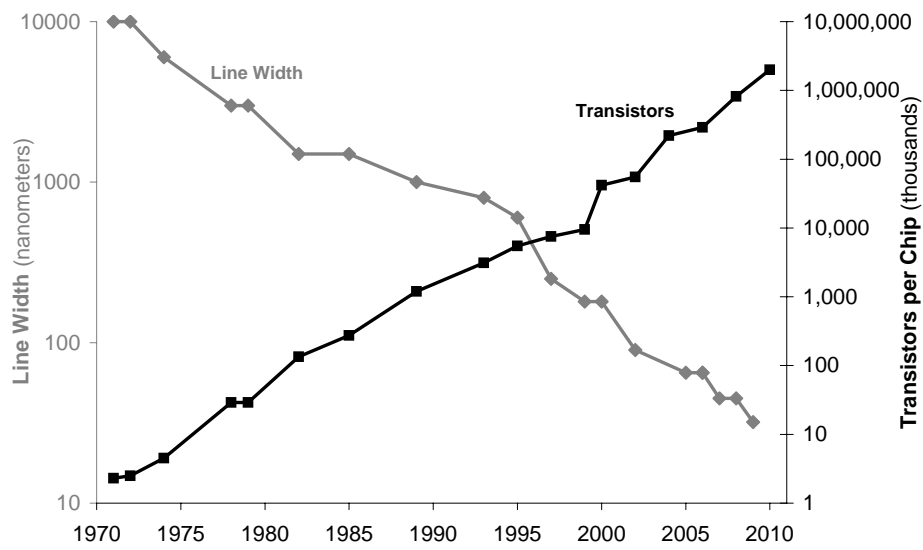
- Feenstra, Robert C. and John Romalis, “International Prices and Endogenous Quality,” *mimeo.*, 2013.
- , Benjamin R. Mandel, Marshall Reinsdorf, and Matthew Slaughter, “Effects of Terms of Trade Gains and Tariff Changes on the Measurement of U.S. Productivity Growth,” *American Economic Journal: Economic Policy*, 2013, 5 (1), 1–38.
- Flamm, Kenneth, “Measurement of DRAM Prices: Technology and Market Structure,” in “Price Measurements and their Uses,” Vol. 57 of *National Bureau of Economic Research Studies in Income and Wealth*, University of Chicago Press, 1993, pp. 157–197.
- , *Mismanaged Trade: Strategic Policy and the Semiconductor Industry*, Brookings, 1996.
- , “The Microeconomics of Microprocessor Innovation,” *mimeo*, 2007.
- Fort, Teresa C., “Breaking Up Is Hard To Do: Why Firms Fragment Production Across Locations,” *mimeo*, 2013.
- Foster, Lucia, John Haltiwanger, and Chad Syverson, “The Slow Growth of New Plants: Learning about Demand?,” *mimeo*, 2012.
- Fudenberg, Drew and Jean Tirole, “Learning-by-Doing and Market Performance,” *Bell Journal of Economics*, 1983, 14 (2), 522–530.
- Global Foundries, “Fact sheet: Fab 2 module 1,” March 2009.
- Greenlees, John S. and Robert McClelland, “New Evidence on Outlet Substitution Effects in Consumer Price Index Data,” *BLS Working Papers*, 2008, (421).
- Grimm, Bruce T., “Price Indexes for Selected Semiconductors, 1974-96,” *Survey of Current Business*, 1998, 78 (2), 8–24.
- Hallak, Juan Carlos and Peter Schott, “Estimating Cross-Country Differences in Product Quality,” *Quarterly Journal of Economics*, 2011, 126(1), 417–474.
- Hausman, Jerry and Ephraim Leibtag, “CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?,” in Erwin W. Diewert, John S. Greenlees, and Charles R. Hulten, eds., *Price Index Concepts and Measurement*, National Bureau of Economic Research, 2009.
- Holdway, Michael, “Quality-Adjusting Computer Prices in the Producer Price Index: An Overview,” 2001. Accessed January 26, 2011.
- Houseman, Susan, Christopher Kurz, Paul Lengemann, and Benjamin Mandel, “Offshoring Bias in U.S. Manufacturing,” *Journal of Economic Perspectives*, 2011, 25(2), 111–132.

- , — , Paul Lengermann, and Benjamin Mandel, “Offshoring Bias in U.S. Manufacturing: Implications for Productivity and Value Added,” *Board of Governors of the Federal Reserve System, International Finance Discussion Papers*, 2010, (1007).
- Houseman, Susan N. and F. Ryder Kenneth, *Measurement Issues Arising from the Growth of Globalization: Conference Summary*, W.E. Upjohn Institute for Employment Research, August 2010.
- Hummels, David and Peter J. Klenow, “The Variety and Quality of a Nation’s Exports,” *American Economic Review*, 2005, *95(3)*, 704–723.
- , Jun Ishii, and Kei-Mu Yi, “The nature and growth of vertical specialization in world trade,” *Journal of International Economics*, 2001, *54*, 75–96.
- Hurtarte, Jorge S., Evert A. Wolsheimer, and Lisa M. Tafoya, *Understanding Fabless IC Technology* 2007.
- IC Knowledge, “Can the semiconductor industry afford the cost of new fabs?,” March 2001.
- Igami, Mitsuru, “Who Offshores, Where, and How? - Evidence from the Global HDD Industry,” *mimeo*, 2010.
- , “Offshoring as Process innovation,” *mimeo*, 2013.
- Irwin, Douglas A. and Peter J. Klenow, “Learning-by-Doing Spillovers in the Semiconductor Industry,” *Journal of Political Economy*, 1994, *102* (6), 1200–1227.
- Kaplan, Greg and Guido Menzio, “The Morphology of Price Dispersion,” *NBER Working Paper*, 2014, (19877).
- Khandelwal, Amit, “The Long and Short (of) Quality Ladders,” *Review of Economic Studies*, 2010, *77(4)*, 1450–1476.
- Klemperer, Paul, “The Competitiveness of Markets with Switching Costs,” *RAND Journal of Economics*, 1987, *18* (1), 138–150.
- , “Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade,” *Review of Economic Studies*, 1995, *62*, 515–539.
- Kugler, Maurice and Eric Verhoogen, “Prices, Plant Size, and Product Quality,” *Review of Economic Studies*, 2011, *79(1)*, 307–339.

- Kumar, Rakesh, “The Business of Scaling,” *IEEE Solid-State Circuit Society News*, 2007, *12*(1), 22–27.
- , *Fabless Semiconductor Implementation*, McGraw-Hill Professional, 2008.
- Mandel, Benjamin, “Heterogeneous Firms and Import Quality: Evidence from Transaction-Level Prices,” *Board of Governors of the Federal Reserve System, International Finance Discussion Papers*, 2010, (991).
- McGregor, Jim, “The common platform technology: A new model for semiconductor manufacturing,” *In-Stat*, 2007.
- Monarch, Ryan, “It’s Not You, It’s Me: Breakups in U.S.-China Trade Relationships,” *mimeo*, 2013.
- Moore, Gordon E., “Cramming More Components onto Integrated Circuits,” *Electronics*, 1965.
- Nakamura, Emi and Jon Steinsson, “Lost in Transit: Product Replacement Bias and Pricing to Market,” *American Economic Review*, 2012, *102* (7), 3277–3316.
- Nishimura, Kazuo and James Friedman, “Existence of Nash Equilibrium in n Person Games without Quasi-Concavity,” *International Economic Review*, 1981, *22* (3), 637–648.
- Office of Management and Budget, “North American Industry Classification System; Revision for 2012,” *Federal Register*, 2011, *76* (159), 51239–51243.
- Reinsdorf, Marshall, “The Effect of Outlet Price Differentials on the U.S. Consumer Price Index,” in “Price Measurements and their Uses,” Vol. 57 of *National Bureau of Economic Research Studies in Income and Wealth*, University of Chicago Press, 1993, pp. 227–254.
- Rosen, Sherwin, “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,” *Journal of Political Economy*, 1974, *82* (1), 34–55.
- Salop, Steven and Joseph Stiglitz, “Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion,” *Review of Economic Studies*, 1977, *44* (3), 493–510.
- Schott, Peter K., “The relative sophistication of Chinese exports,” *Economic Policy*, 2008, *23*, 5–49.
- Scott, A.J. and D.P. Angel, “The Global Assembly-Operations of U.S. Semiconductor Firms: A Geographical Analysis,” *Environment and Planning A*, 1988, *20*(8), 1047–67.
- Shin, Jiwoong and K. Sudhir, “Switching Costs and Market Competitiveness: Deconstructing the Relationship,” *Journal of Marketing Research*, 2009, *46* (4), 446–449.

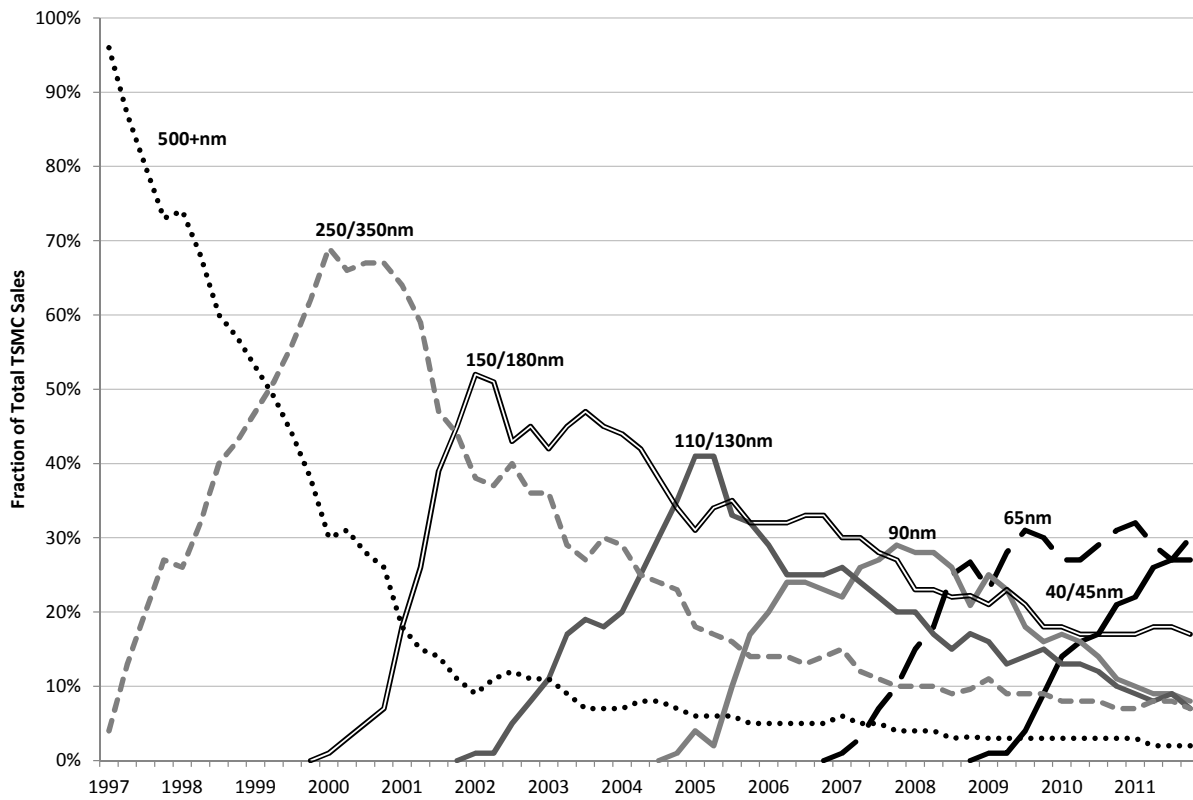
- Sorensen, Alan T., “Equilibrium Price Dispersion in Retail Markets for Prescription Drugs,” *Journal of Political Economy*, 2000, 108(4).
- Stigler, George J., “The Economics of Information,” *Journal of Political Economy*, 1961, 69, 213–25.
- and James K. Kindahl, *The Behavior of Industrial Prices*, NBER, 1970.
- Syverson, Chad, “Market Structure and Productivity: A Concrete Example,” *Journal of Political Economy*, 2004, 112 (6), 1181–1222.
- To, Theodore, “Multi-Period Competition with Switching Costs: An Overlapping Generations Formulation,” *Journal of Industrial Economics*, 1996, 44, 81–87.
- Triplett, Jack, *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, OECD, 2006.
- Turley, James L., *The Essential Guide to Semiconductors*, Prentice Hall PTR, 2003.

Figure 1: Moore's Law - Intel Processors



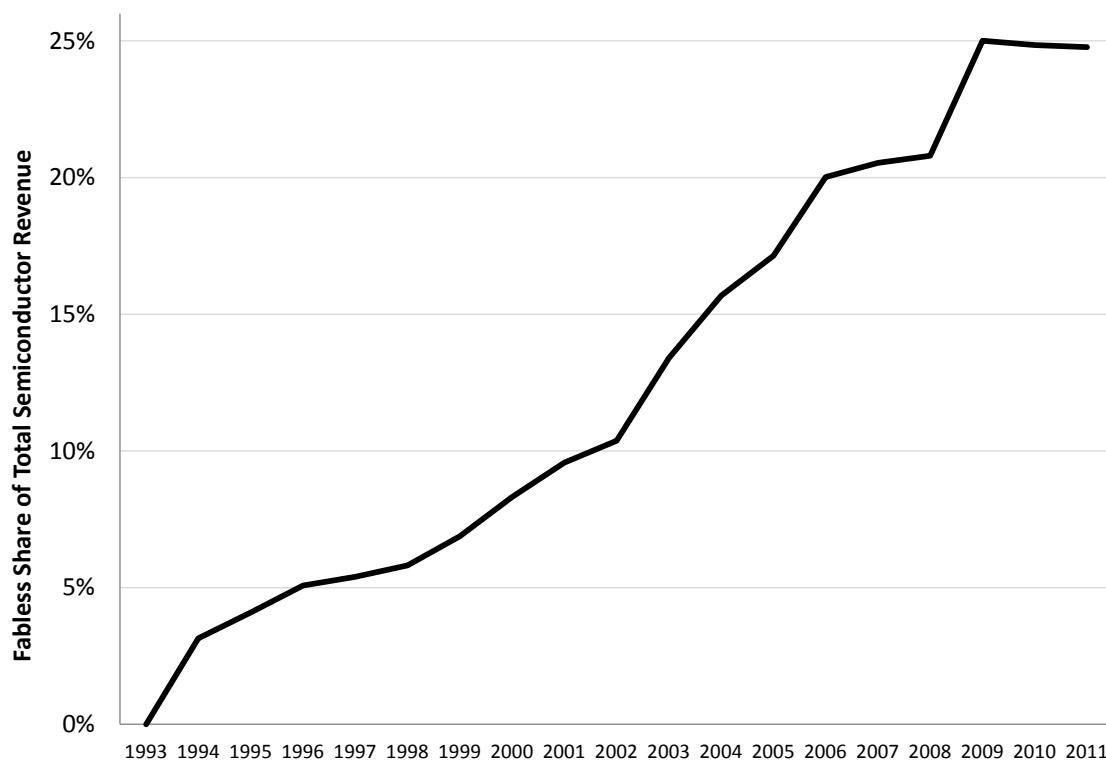
Data sources: <http://www.intel.com/technology/timeline.pdf> and <http://www.intel.com/pressroom/kits/quickreffam.htm>

Figure 2: Technology Cycle - TSMC Sales by line width



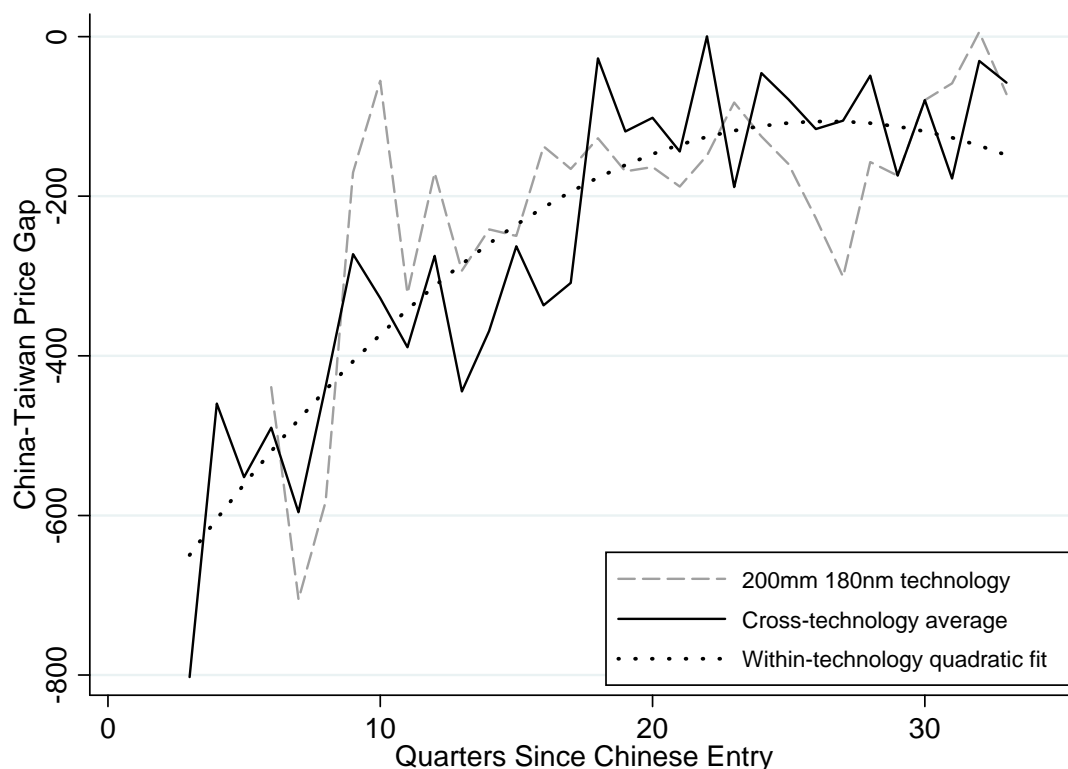
Authors' calculations based on quarterly reports from the largest semiconductor contract manufacturing firm, Taiwan Semiconductor Manufacturing Corporation (TSMC).

Figure 3: Growth of the Fabless Business Model



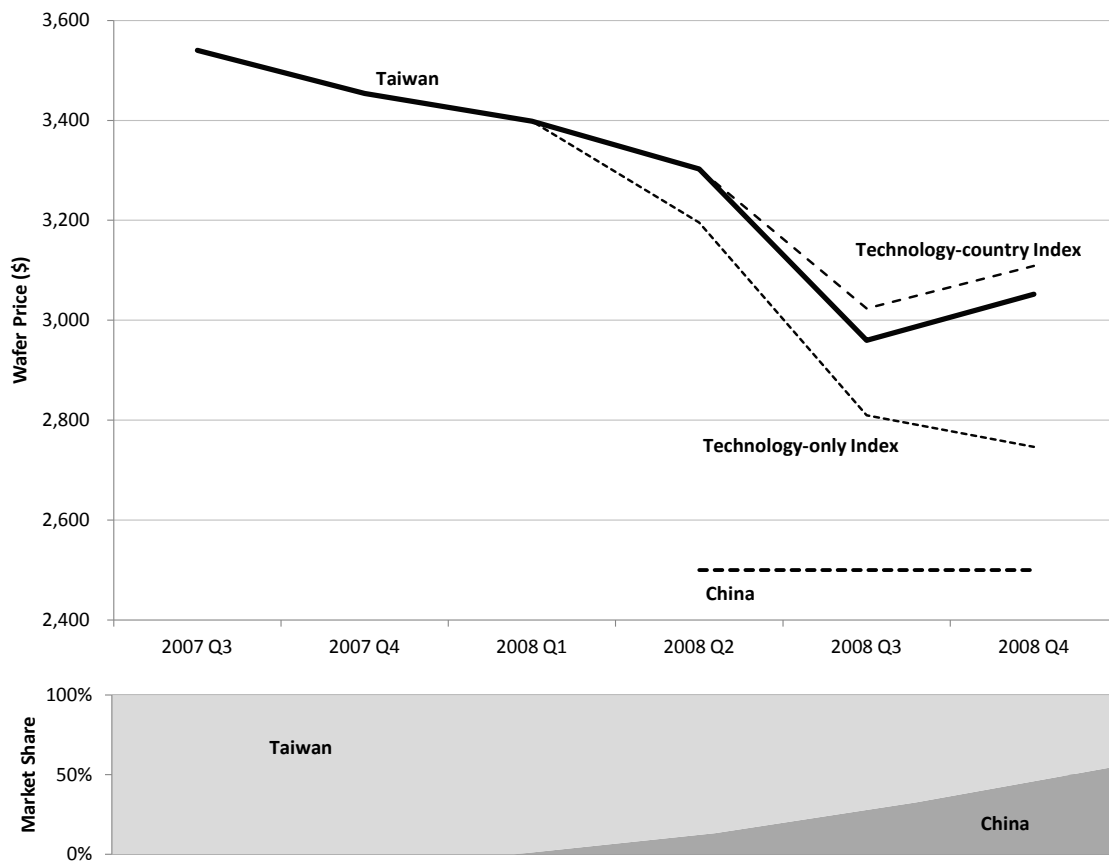
Authors' calculations based on data from the Global Semiconductor Alliance (GSA) and Semiconductor Industry Association (SIA)

Figure 4: Closing China-Taiwan Price Gap Following Chinese Entry



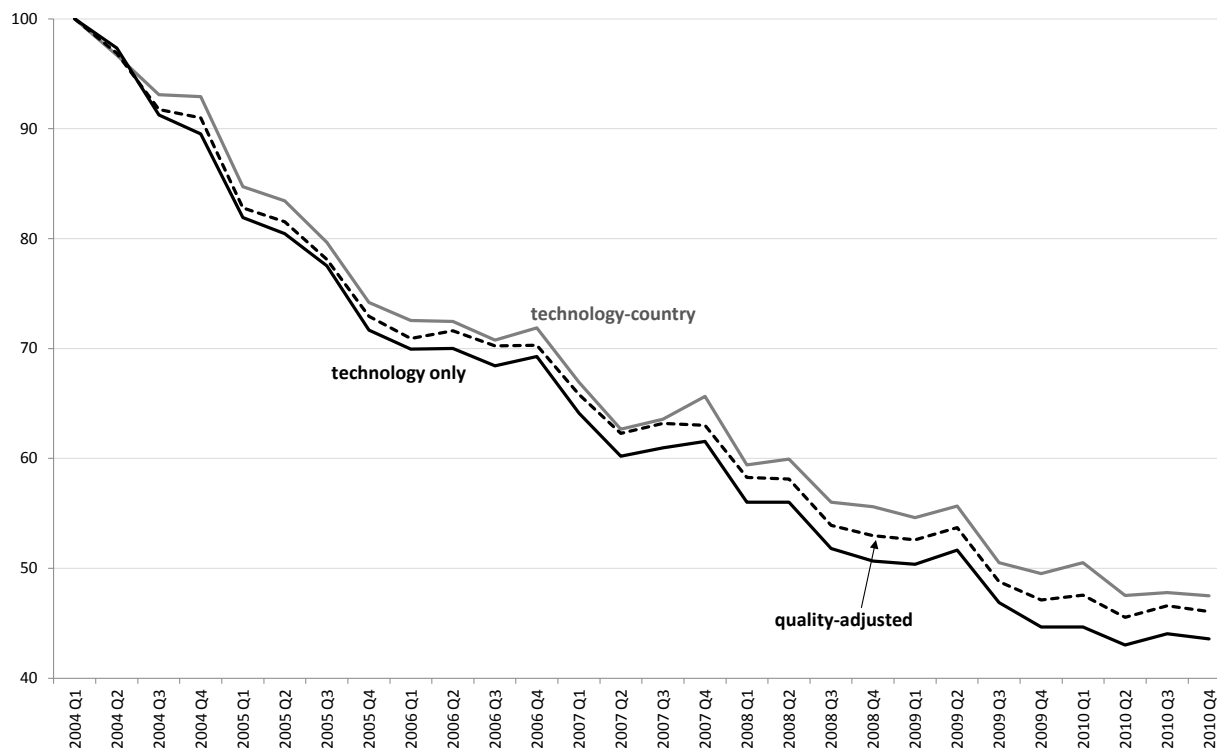
The x-axis measures the number of quarters since Chinese suppliers began producing the relevant technology. The y-axis measures the gap in average price for Chinese vs. Taiwanese producers. The gray dashed line shows the raw quarterly price gap data for the 200mm 180nm technology, which had the largest market share during our sample period, showing the closing gap over time within that technology. The solid black line shows the evolution of the price gaps averaged across technologies, which confirms the closing price gap but may be confounded by changing composition of technology over time. The dotted line shows predicted values from a quadratic trend estimated within-technology using technology fixed-effects. See text and Table 4 for details.

Figure 5: Index Calculation Example



Authors' calculations using GSA and IHS iSuppli data. See discussion in Section 5.1 for details.

Figure 6: Price Indexes with Various Quality Dispersion Assumptions



Authors' calculations using GSA wafer price data and IHS iSuppli quantity data. Sample restricted to wafers produced in Taiwan and China. The figure presents Fisher matched-model indexes with various assumptions about quality dispersion across suppliers. The technology-country index assumes all cross-country price level differences reflect quality differences, and hence defines a model based on both technology and country. The technology only index assumes that all cross-country price level differences reflect pure price dispersion, and hence defines a model based on technology only. The quality-adjusted index implements an intermediate quality assumption informed by the price dynamics across Chinese and Taiwanese suppliers, representing our best estimate of the true quality-adjusted price index. See text for details.

Table 1: Foundry Capacity by Country

	Global Capacity									
	Foundry	Total Industry	Taiwan	China	Singapore	Europe	USA	Japan	S. Korea	Malaysia
2000	875	9,462	63.2%	7.1%	7.5%	6.8%	4.4%	7.8%	3.2%	0.0%
2001	972	8,286	60.5%	8.9%	8.4%	6.3%	4.2%	7.1%	2.9%	1.7%
2002	1,011	8,646	56.0%	10.7%	10.0%	6.2%	5.3%	7.0%	2.9%	1.8%
2003	1,150	9,018	51.1%	15.5%	10.1%	6.5%	5.1%	6.4%	3.2%	2.2%
2004	1,429	10,000	50.3%	19.0%	9.6%	5.6%	4.1%	5.3%	3.1%	3.0%
2005	1,739	11,073	48.5%	23.2%	9.2%	4.9%	3.6%	4.5%	3.0%	3.1%
2006	1,951	12,320	48.0%	24.1%	9.9%	4.4%	3.3%	4.1%	3.2%	3.0%
2007	2,157	13,588	48.9%	23.4%	10.0%	4.6%	3.1%	3.7%	3.5%	3.0%
2008	2,401	14,297	49.9%	22.1%	11.2%	5.1%	2.9%	2.9%	3.3%	2.7%
2009	2,546	14,058	49.2%	22.0%	10.9%	6.5%	2.8%	2.7%	3.5%	2.5%
2010	2,812	14,230	49.4%	21.5%	11.3%	7.4%	2.6%	2.5%	3.3%	2.1%
2011	3,177	14,923	50.2%	21.8%	10.7%	7.9%	2.4%	2.3%	3.0%	1.8%

Authors' calculations from IHS iSuppli data. Sample includes contract manufacturers ("pure-play" foundries) only. Capacity measured in thousand 8-inch equivalent wafers per month.

Table 2: Wafer Price Descriptive Statistics

	Mean	Std. Dev	Yearly Means							
			2004	2005	2006	2007	2008	2009	2010	
Price Per Wafer (\$)	1,204.22	949.28	1,321.01	1,293.51	1,298.54	1,189.68	1,120.34	1,080.39	1,126.09	
Number of Wafers Contracted	11,865.00	22,302.74	8,344.61	14,912.17	10,544.97	16,767.34	9,260.24	9,692.42	13,533.24	
Layers										
Metal Layers	4.60	1.77	4.22	4.56	4.85	4.76	4.62	4.56	4.64	
Mask Layers	25.72	7.31	23.69	24.52	26.37	26.74	25.86	26.27	26.58	
Polysilicon Layers	1.23	0.46	1.36	1.21	1.20	1.19	1.30	1.14	1.20	
Wafer Size										
150 mm	0.16	0.37	0.18	0.17	0.15	0.10	0.17	0.16	0.17	
200 mm	0.73	0.44	0.79	0.76	0.76	0.81	0.68	0.68	0.63	
300 mm	0.11	0.32	0.03	0.07	0.09	0.09	0.15	0.16	0.20	
Line Width										
45 nm	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.03	
65 nm	0.03	0.16	0.00	0.00	0.00	0.01	0.04	0.07	0.08	
90 nm	0.04	0.21	0.00	0.02	0.04	0.05	0.08	0.06	0.06	
130 nm	0.15	0.36	0.12	0.16	0.18	0.17	0.14	0.14	0.15	
150 nm	0.06	0.23	0.07	0.08	0.10	0.09	0.03	0.01	0.01	
180 nm	0.26	0.44	0.25	0.23	0.24	0.28	0.26	0.31	0.26	
250 nm	0.12	0.33	0.17	0.17	0.13	0.15	0.10	0.07	0.08	
older vintage	0.33	0.47	0.40	0.35	0.30	0.25	0.34	0.35	0.33	

Authors' calculations based on GSA Wafer Fabrication and Back-End Pricing Survey and IHS iSuppli shipments data. Summary statistics based on sample of 6253 transaction-level observations.

Table 3: Hedonic Wafer Price Regressions

Variable	(1)		(2)		(3)		(4)	
	No Attribute Controls	Linear Attribute Controls	Flexible Attribute Controls	China and Taiwan Only	Coefficient	Std. Err.	Coefficient	Std. Err.
Foundry Location								
China	-0.297	(0.042)***	-0.186	(0.027)***	-0.188	(0.027)***	-0.196	(0.028)***
Malaysia	-0.274	(0.065)***	-0.278	(0.042)***	-0.286	(0.041)***		
Singapore	-0.046	(0.026)*	-0.061	(0.016)***	-0.068	(0.016)***		
United States	-0.093	(0.021)***	0.068	(0.030)**	0.064	(0.031)**		
Wafer Size								
150 mm			-0.467	(0.032)***				
300 mm			0.671	(0.021)***				
Line Width								
≥ 500 nm			-0.245	(0.053)***				
350 nm			-0.167	(0.033)***				
250 nm			-0.061	(0.026)**				
150 nm			0.169	(0.027)***				
130 nm			0.356	(0.018)***				
90 nm			0.479	(0.032)***				
65 nm			0.676	(0.030)***				
45 nm			0.962	(0.062)***				
Wafer Size x Line Width Indicators					X		X	
Number of Metal Layers			0.076	(0.007)***	0.076	(0.007)***	0.081	(0.007)***
Number of Polysilicon Layers			0.027	(0.024)	0.028	(0.024)	0.025	(0.024)
Number of Mask Layers			0.005	(0.002)***	0.005	(0.002)***	0.004	(0.002)**
Epitaxial Layer Indicator			0.064	(0.037)*	0.067	(0.037)*	0.066	(0.038)*
log Number of Wafers Contracted			-0.056	(0.004)***	-0.057	(0.004)***	-0.058	(0.005)***
Quarter Indicators	X	X	X	X	X	X	X	X
R-squared	0.046	0.909	0.913	0.922				
Observations	6253	6253	6253	5378				

Observations represent individual semiconductor wafer transactions from GSA data, weighted using IHS iSuppli shipment weights. Baseline case (omitted category) is a 200mm 180nm wafer produced in Taiwan. Standard errors clustered for 28 quarter clusters, * significant at 10% level, ** 5%, *** 1%.

Table 4: Closing China-Taiwan Price Gaps Within Technology

	<i>dependent variable: China-Taiwan price gap</i>			
	(1)	(2)	(3)	(4)
Time since Chinese entry	13.036 (2.819)***	14.283 (2.956)***	46.229 (13.179)***	52.287 (12.710)***
(Time since Chinese entry) ²			-0.863 (0.326)***	-0.989 (0.301)***
China-Taiwan difference in average:				
number of metal layers		46.108 (15.394)***		60.734 (14.814)***
number of polysilicon layers		80.616 (73.088)		72.578 (64.409)
number of mask layers		3.116 (5.713)		6.721 (5.167)
epitaxial layer		-108.458 (83.844)		-95.153 (71.247)
log number of wafers contracted		-31.746 (14.114)**		-21.874 (13.142)*
Technology (wafer size x line width) indicators	X	X	X	X
R-squared	0.508	0.647	0.547	0.692
Observations	91	91	91	91

Observations represent technology (wafer size and line width) and quarter combinations, and the dependent variable is the difference in average price between Chinese and Taiwanese producers in the associated cell. “Time since Chinese entry” reflects the elapsed number of quarters since China first began producing the relevant technology. All specifications include technology fixed effects, so trends reflect within-technology changes in the price gap as time elapsed following Chinese entry. Heteroskedasticity-robust standard errors in parentheses, * significant at 10% level, ** 5%, *** 1%.

Table 5: Model Equilibrium Results

	<u>Calibrated Model</u>		<u>Frictionless Model</u>	<u>Data</u>	
	Period 2	Period 3		<i>Time since Chinese entry</i> 10 quarters	20 quarters
Price gap	257.9	148.9	150.3	373.5	147.2
Share of t-1 cohort buyers switching in t	11.3%	0.0%	0.0%		
Firm A total market share	68.4%	55.2%	61%	77.8%	70.2%
Cutoff complexity in period 2 (\hat{y})	0.52		0.39		

“Calibrated Model” column presents equilibrium results for the numerical solution to the calibrated model. Calibration: $c^B = 334$, $c_A = 400$, $\tau = 395$, $s = 0.54\tau$. “Frictionless Model” column presents equilibrium results for the same parameters except $s = 0$. This is a static model, so outcomes do not vary with time. “Data” column presents data analogs to the model’s outcomes, calculated from the quadratic fit in Figure 4.

Table 6: Price Indexes with Various Quality Dispersion Assumptions

Sample: Index:	All countries		China and Taiwan		
	technology-country	technology only	technology-country	technology only	quality-adjusted
2004 Q1	100.0	100.0	100.0	100.0	100.0
2004 Q2	97.0	96.0	96.7	97.3	96.9
2004 Q3	92.8	90.5	93.1	91.3	91.8
2004 Q4	91.9	88.2	92.9	89.5	91.0
2005 Q1	83.7	81.2	84.7	81.9	82.8
2005 Q2	82.8	81.0	83.5	80.4	81.5
2005 Q3	79.6	78.6	79.7	77.5	78.1
2005 Q4	74.3	72.8	74.2	71.7	72.9
2006 Q1	72.9	71.2	72.6	69.9	70.9
2006 Q2	72.8	70.5	72.5	70.0	71.6
2006 Q3	71.0	68.3	70.8	68.4	70.2
2006 Q4	71.3	68.0	71.9	69.3	70.3
2007 Q1	66.9	63.1	66.9	64.1	65.8
2007 Q2	63.0	59.8	62.7	60.2	62.3
2007 Q3	63.6	60.5	63.6	61.0	63.2
2007 Q4	65.5	61.0	65.6	61.5	63.0
2008 Q1	59.9	56.1	59.4	56.0	58.3
2008 Q2	60.4	55.8	59.9	56.0	58.1
2008 Q3	56.6	51.5	56.0	51.8	53.9
2008 Q4	56.6	51.1	55.6	50.7	53.0
2009 Q1	56.0	50.9	54.6	50.4	52.6
2009 Q2	56.7	52.0	55.7	51.7	53.7
2009 Q3	51.7	47.4	50.5	46.9	48.8
2009 Q4	50.7	45.7	49.5	44.7	47.1
2010 Q1	52.0	46.7	50.5	44.7	47.6
2010 Q2	48.6	43.9	47.5	43.0	45.5
2010 Q3	48.8	44.6	47.8	44.0	46.6
2010 Q4	48.8	44.2	47.5	43.6	46.1
2004	95.4	93.7	95.7	94.5	94.9
2005	80.1	78.4	80.5	77.9	78.8
2006	72.0	69.5	71.9	69.4	70.8
2007	64.7	61.1	64.7	61.7	63.6
2008	58.4	53.6	57.7	53.6	55.8
2009	53.8	49.0	52.6	48.4	50.5
2010	49.5	44.8	48.3	43.8	46.4
Avg. Yearly Change '04-'10	-10.4%	-11.6%	-10.8%	-12.0%	-11.2%

Authors' calculations using GSA wafer price data and IHS iSuppli quantity data. Fisher matched-model indexes with various assumptions about quality dispersion across suppliers. The technology-country index assumes all cross-country price level differences reflect quality differences, and hence defines a model based on both technology and country. The technology only index assumes that all cross-country price level differences reflect pure price dispersion, and hence defines a model based on technology only. The quality-adjusted index implements an intermediate quality assumption informed by the price dynamics across Chinese and Taiwanese suppliers, representing our best estimate of the true quality-adjusted price index. See text for details.

A Data

In this appendix, we describe the data sources and steps taken in the construction of the dataset used in our analysis.

A.1 Wafer Prices

As discussed in Section 3.1, our wafer price data come from a proprietary database of semiconductor wafer purchases from foundries, collected by the Global Semiconductor Alliance (GSA). We implement a number of data cleaning procedures before the main analysis.

First, we implement a variety of sample restrictions. We drop observations corresponding to engineering runs that occur during the design stage prior to volume production. We omit a few observations reporting the obsolete 100mm wafer diameter. As discussed in Section 2.1, we keep only observations corresponding to CMOS process technology, which dominates the foundry market, and omit other processes that are quite distinct and serve niche markets for high-power, defense, aerospace applications. We also keep only observations for wafers produced by so-called “pure-play” foundries, and omit transactions involving integrated device manufacturers (IDMs) that do both design and fabrication.

Table A.1 shows that 6,916 observations satisfy these sample restrictions. We drop an additional 663 observations for a variety of reasons. 576 observations do not report the location of production, 1 observation reports an implausibly large order (which distorts the price per wafer), and 86 observations report technologies (wafer-size, line-width combinations) for which other sources (see below) report no production in the reported country and quarter.

Finally, we also combine closely related line widths. Any line width greater than or equal to 500nm is combined into the 500nm code. 140 and 150nm widths are combined into the 150nm code. 80 and 90nm widths are combined into the 90nm code. 60 and 65nm widths are combined into the 65nm code. 40 and 45nm widths are combined into the 45nm code. In all of these cases, one of the combined widths is vastly dominant in the market, and it would be difficult to separately identify prices for the less prevalent line width with very few observations.

A.2 Semiconductor Wafer Shipments

Our price indexes use quantity information to weight observations across process technologies (though the qualitative nature of the results does not depend upon the choice of weighting scheme). The GSA data include quantity information for each shipment, but aggregating this information yields quite volatile aggregate quantity measures for each technology. Instead of using this quantity information from GSA, we employ data published by IHS iSuppli in their “Pure Play Foundry Market Tracker.” This report is a global census of semiconductor foundries, including 91 fabs belonging to 20 companies. Annual and quarterly frequency data begin in 2000 and 2002, respectively. Characteristics provided for each fab include company of ownership, location, wafers shipped per month at full capacity, and diameter of wafers shipped. This allows us to construct capacity by region and wafer size (Table 1).

In addition, the report provides information on wafer shipments for specific technologies (line widths) by *company*, but not by *plant*. For 11 of the companies covered, this information is sufficient to construct the needed wafer size-by-line width-by-country weights for our analysis without further assumptions. In two cases, only one fab is active in the period we cover, and in nine cases, all the company’s fabs employ the same wafer diameter and are located in the same country.

The remaining companies either have fabs in multiple regions, fabricate wafers of multiple diameters, or both. In these cases, we first estimate wafer shipments by technology for each fab, which allows us to divide production between countries for each technology. To do this, we employ three additional resources: the partial information on the timing of technology introduction by plant from the IHS iSuppli report; company

information provided in public statements; and extensive discussions with iSuppli.⁶¹ Specifically, the “Market Tracker” lists the technologies employed (without output shares) in each fab at the time of publication, which we have for previous vintages of the database beginning in Q1 2010. This information allows us to construct technology-by-country-by-quarter weights for an additional 2 companies. In these cases, each with two fabs in operation, we were able to identify one company fab exclusively employing a single category of legacy technology (500nm and above), leaving the remaining fab to account for the residual company production.

In the remaining 7 cases, to arrive at the weights, we need further information about the technology employed at specific plants. A search of company press releases and industry press reports yielded information on the timing of introduction of particular line widths at specific fabs in several cases, but information on the relative importance of each technology *by fab* is not available. To fill this gap we appeal to information on industry norms gathered from iSuppli and other reports and discussions with industry analysts. We assume that several rules hold in general: (1) fabs add new technologies progressively-adding a line width more advanced than all technologies used in previous periods; (2) fabs ramp up output of new technologies linearly over a four-quarter period; (3) companies introduced new technologies first at the company’s most advanced fab; (4) when not ramping a new technology, fabs split production evenly among the 2-4 technologies in production. Individual companies often required deviations from these rules based on information regarding specific fabs to match the overall company technology mix.

Table A.2 shows the specific assumptions we made for each foundry in generating the weights, and we have made available the resulting set of weights by country, wafer size, line width, and quarter at:

http://www.andrew.cmu.edu/user/bkovak/bkm_semicon_data/index.html

⁶¹We thank Len Jelinek of iSuppli for invaluable assistance in understanding and supplementing the information in the report.

Table A.1: Dropped Observations

Total observations	6,916
Used in analysis	6,253
Dropped	663
Missing foundry location	576
Implausibly large order reported*	1
Inconsistent	86

There may be multiple reasons to drop an observation

* Confirmed with GSA

Table A.2: Technology Assumptions by Country

Company	# of Fabs	# of Countries	# of Wafer Diameters	Notes on Assumptions
Altis Semiconductor	1	1	1	
ASMC	3	1	2	Small diameter fab produces legacy technology. Large diameter fab produces remaining, more advanced technology.
CRMC	4	1	2	All fabs are same diameter until 2009. No assumptions needed. 2009-2011: small diameter fabs produce all reported legacy technology and amount of 350nm indicated by historical pattern. Remaining shipments from large diameter fabs.
Dongbu Hi Tek	2	1	1	
Episil	4	1	1	
Globalfoundries/ Chartered	11	2	3	2004-2008: Single country Small diameter fab capacity split evenly between 350nm and legacy technology. Large diameter fabs produce all shipments using 130nm and more advanced technology. Medium diameter fabs produce remaining, more advanced technology. 2009-2011: Begin with 2008 mix from existing locations. Ramp up 65nm and 45nm with timing indicated in reports. Residual is production in single remaining location.
Grace	2	1	1	
He Jian	2	1	1	
HHNEC	3	1	1	
HuaLi	1	1	1	
Landshunt Silicon Foundry GmbH	2	1	1	
Phenittec	3	1	1	
Silterra	3	1	1	
SMIC	10	1	2	All 150nm and less advanced technology produced at medium-diameter fabs. All 65nm and 90nm is produced at large-diameter fabs. Assign 130nm production based on guidance from industry analysts.
SSMC	1	1	1	
TowerJazz	5	3	2	Small diameter fab produces 350nm and legacy technology. Remainder split among other locations proportional to capacity.
TSMC	13	4	3	Production at one location known with certainty. Second set of fabs known to be divided between 250nm and 350nm, assumed to be split evenly. Third set of fabs assumed to be split evenly among 5 technologies (130nm, 150nm, 250nm, 350nm, 500nm) Small diameter fab accounts for nearly all legacy technology and small share of 350nm & 250nm. Small amount of legacy production in second location indicated by data. Large diameter fabs account for all production using 90nm and more advanced technologies. Residual capacity at these fabs split evenly between 130nm, 150nm, and 180nm until drawing down to minimal to offset ramp-up of 65nm technology. Residual goes to medium diameter production at remaining location.
UMC	12	3	3	Production at one location known with certainty. Small diameter fab split evenly between 350nm and legacy technology. Remaining 350nm and legacy technology and all 150nm, 180nm, and 250nm production from medium-diameter fabs. Large diameter fabs account for all production using 65nm and more advanced technologies. assumptions required to split 90nm-130nm between 200mm and 300mm. 2008-2011: Remaining medium-diameter capacity split evenly between 90nm and 130nm. Residual production using these technologies at large-diameter fabs.
Vanguard	2	1	1	
X-Fab Semiconductor Foundries AG	6	4	2	Split legacy technology production between small-diameter fabs in two locations proportional to capacity. 350nm at known location, medium diameter. Residual is implied at third location.

Notes: All companies require an estimate of the share of production using CMOS process.

B Yield Analysis

From Q4 2004 to Q3 2008 the GSA survey included a question reporting yields in the following four ranges: 0-25%, 26-50%, 51-75%, and 76-100%. Table B.1 tabulates these yield ranges for each supplier country. The findings are consistent with our discussions with industry insiders that Chinese yields are comparable to those in Taiwan. No observations report a yield below 76% for a transaction produced by a Chinese supplier. Given the lack of variation in this yield measure across suppliers, including indicators for the yield range in hedonic specifications like those in Table 3 has essentially no effect on measured price dispersion across suppliers. However, because the reported ranges are quite wide, this yield range measure leaves room for non-trivial variation in yields across suppliers.

To address that issue, GSA added a continuous yield question to the Q3 2011 survey at our request. This question asked respondents to report the transaction's yield as a share between 0% and 100%. While this measure provided much more detail on yields, it also was omitted or implausibly reported as zero by a large fraction (60%) of respondents. While this high rate of non-response sharply reduces the power of any tests of yield differences across suppliers, the data nonetheless support the industry perception that Chinese yields are no lower than those in Taiwan. Column (1) of Table B.2 examines yield differences across suppliers using a specification similar to that in Table 3. Although the supplier location coefficients are imprecisely estimated, there is no evidence for lower yields at Chinese producers (note that the Malaysia coefficient cannot be estimated due to a lack of valid yield data). Moreover, as shown in Table B.2, if anything observations corresponding to Chinese transactions were more likely to have a valid value for the continuous yield measure, which is inconsistent with the notion that customers were simply less likely to report their lower yields from Chinese foundries.

While the power of these analyses is curtailed by a lack of detail in the yield measure or by substantial non-response, the results all point to similar yields across suppliers, consistent with our discussions with industry insiders.

Table B.1: Yield Ranges

Yield Range:	0-25%	26-50%	51-75%	76-100%
China	0.0	0.0	0.0	100.0
Malaysia	0.0	0.0	0.0	100.0
Singapore	0.0	0.0	6.3	93.7
Taiwan	0.0	1.8	4.6	93.7
United States	0.0	0.6	8.1	91.3

Share of transactions from each country reporting yields in the relevant range. Observations represent individual semiconductor wafer transactions from GSA data, weighted using IHS iSuppli shipment weights. Yield range measure available from Q4 2004 to Q3 2008. 3,208 nonmissing observations.

Table B.2: Continuous Yield Analysis, Q3 2011

Dependent Variable:	(1)		(2)	
	<u>yield (in percent, 0...100)</u> Coefficient	Std. Err.	<u>indicator for missing yield</u> Coefficient	Std. Err.
Foundry Location				
China	0.485	(1.419)	-0.308	(0.133)**
Malaysia	.	.	0.189	(0.132)
Singapore	-5.954	(6.471)	0.141	(0.109)
United States	-0.470	(1.945)	-0.002	(0.106)
Wafer Size x Line Width Indicators	X		X	
Number of Metal Layers	0.754	(0.467)	0.040	(0.033)
Number of Polysilicon Layers	-0.368	(1.637)	0.139	(0.069)**
Number of Mask Layers	-0.246	(0.122)**	-0.010	(0.008)
Epitaxial Layer Indicator	0.650	(1.313)	-0.199	(0.084)**
log Number of Wafers Contracted	0.352	(0.350)	0.004	(0.017)
R-squared	0.254		0.113	
Observations	106		265	

Column (1) examines cross country differences in yields, as reported in the continuous yield question in Q3 2011, while column (2) examines differences in the probability of reporting a valid (nonzero) yield, using a linear probability model. Observations represent individual semiconductor wafer transactions from GSA data. Sample restricted to Q3 2011 for availability of continuous yield measure. Baseline case (omitted category) is a 200mm 180nm wafer produced in Taiwan. Robust standard errors in parentheses, * significant at 10% level, ** 5%, *** 1%

C Chinese Foundry Entry Timing

The time of entry by technology for Chinese foundries was assessed using data from IHS iSuppli's Pure Play Foundry Market Tracker. This report provides historical information on the composition of production for all major Chinese foundries. The data is tabulated three ways, including annual (quarterly) company wafer shipments by feature size beginning in 2000 (2003), annual (quarterly) capacity by fab beginning in 2000 (2002), and annual (quarterly) company capacity by line width beginning in 2000 (2002). Because each fab has a unique wafer size, and the only important Chinese foundry at the time (Semiconductor Manufacturing International Corporation, SMIC) employed 200 mm wafers exclusively in the early 2000s, combining these data unambiguously determines the year of Chinese for each technology introduced from 2001 to 2003. Identifying the quarter of introduction within 2002 required an assumption that SMIC began producing technologies in decreasing order of line width during the year. The oldest technologies were omitted from the entry analysis, as these were present years before Chinese firms entered the foundry market. The resulting entry timing is shown in Table C.1.

Table C.1: Chinese Foundry Entry Timing

Wafer Diameter	Line Width	Entry Period
150 mm	500 nm	.
150 mm	350 nm	.
200 mm	500 nm	.
200 mm	350 nm	2002 Q1
200 mm	250 nm	2002 Q2
200 mm	180 nm	2002 Q3
200 mm	150 nm	2003 Q4
300 mm	130 nm	2004 Q2
300 mm	90 nm	2007 Q3
300 mm	65 nm	2009 Q1

Quarter in which Chinese foundries first began production for each process technology. Authors' calculations based on IHS iSuppli data.

D Model

D.1 Setup

The problem is solved by backward induction. To analyze the terminal-period problem, we first conjecture that, in the prior period, Firm B attracts all period-2 entrants with $y < \hat{y} < 1$: In other words, we assume the least efficient producer will attract buyers with the least complex designs. This conjecture will be confirmed in equilibrium. In what follows, since y is uniformly distributed, we refer to the mass of buyers \hat{y} as Firm B 's *customer base* as of the start of period 3. The mass of higher-quality buyers $1 - \hat{y}$ makes up Firm A 's customer base.

Firm A 's terminal-period problem can now be stated as follows. There are three groups of buyers to whom Firm A may sell: members of its own customer base, buyers in Firm B 's customer base, and buyers who enter in period 3, referred to as period-3 entrants. The demand schedules for each of these cohorts is given below. Throughout, we let σ_t^{jj} represent the share of Firm j 's customer base that it retains in period t ; σ_t^{ji} the share of Firm i 's customer base that switches to Firm j ; and σ_t^{j0} the share of period- t entrants that Firm j attracts. It follows that

$$\begin{aligned} \sigma_3^{AA}(p_3^A; p_3^B, \hat{y}) &= \Pr[p_3^A < p_3^B + \tau y + s \mid y > \hat{y}] \\ \sigma_3^{AB}(p_3^A; p_3^B, \hat{y}) &= \Pr[p_3^A + s < p_3^B + \tau y \mid y \leq \hat{y}] \\ \sigma_3^{A0}(p_3^A; p_3^B) &= \Pr[p_3^A < p_3^B + \tau y]. \end{aligned} \quad (D1)$$

Since $y \sim U[0, 1]$, each of the expressions in (D1) yield simple, linear demand schedules. It follows that total sales by Firm A in period 3 are given by

$$Y_3^A = \sigma_3^{AA}(p_3^A; p_3^B, \hat{y})(1 - \hat{y}) + \sigma_3^{AB}(p_3^A; p_3^B, \hat{y})\hat{y} + \sigma_3^{A0}(p_3^A; p_3^B). \quad (D2)$$

Firm A now selects p_3^A to maximize profit, $(p_3^A - c^A)Y_3^A$.

Firm B faces a symmetric problem. The intersection of the two firms' best response functions determines the terminal-period equilibrium, given \hat{y} . We denote equilibrium prices by $\mathcal{P}_3^A(\hat{y})$ and $\mathcal{P}_3^B(\hat{y})$. It remains to recover the period-2 market share \hat{y} .

A period-2 entrant with design y will purchase from Firm A only if the discounted sum of period-1 and 2 prices is less than those the entrant would face if purchasing from Firm B . This implies that the threshold, \hat{y} , is implicitly defined by the indifference relation,

$$\begin{aligned} p_2^A + \beta \min[\mathcal{P}_3^A(\hat{y}), \mathcal{P}_3^B(\hat{y}) + \tau\hat{y} + s] \\ = \mathcal{P}_3^B(\hat{y}) + \tau\hat{y} + \beta \min[\mathcal{P}_3^A(\hat{y}) + s, \mathcal{P}_3^B(\hat{y}) + \tau\hat{y}], \end{aligned} \quad (D3)$$

where $\beta < 1$ is the discount factor. Equation (D3) solves the threshold in terms of period-2 prices, $\hat{y} = \hat{y}(p_2^A; p_2^B)$. Note that $(1 - \hat{y})$ represents Firm A 's sales among period-2 entrants.

In addition, Firm A may have some its old buyers poached by Firm B . Specifically, Firm B attracts a buyer y in Firm A 's base if $p_2^B + \tau y + s < p_2^A$. Since Firm A captures all (mass one) of the period-1 entrants, this means that Firm B attracts a share of them equal to $\frac{p_2^A - p_2^B - s}{\tau}$. It follows that Firm A 's sales among members of its customer base are $1 - \frac{p_2^A - p_2^B - s}{\tau}$.

Putting these pieces together, Firm A 's optimization problem in period 2 can be written as

$$\max_{p_2^A} \left\{ (p_2^A - c^A) \cdot Y_2^A(p_2^A; p_2^B) + \beta (\mathcal{P}_3^A(\hat{y}(p_2^A; p_2^B)) - c^A) \cdot Y_3^A(\hat{y}(p_2^A; p_2^B)) \right\},$$

where $\hat{y} = \hat{y}(p_2^A; p_2^B)$ solves (D3); $Y_3^A(\hat{y}(p_2^A; p_2^B))$ is given by (D2) after substituting in equilibrium prices $\mathcal{P}_3^A(\hat{y})$ and $\mathcal{P}_3^B(\hat{y})$; and

$$Y_2^A(p_2^A; p_2^B) = 1 - \hat{y}(p_2^A; p_2^B) + \left(1 - \frac{p_2^A - p_2^B - s}{\tau}\right).$$

Firm B solves a symmetric problem.

The challenging aspect of the model is that best responses typically fail to be continuous and, as a result, there may exist no equilibrium. The reason can be traced to the fact that, given $s > 0$, no firm wishes to charge a price so as to acquire only a *marginal* share of new entrants. If Firm A does this, for instance, it makes the $y = 1$ entrant just indifferent between suppliers. But in that case, A 's incumbents will be strictly infra-marginal because $s > 0$. As a result, the firm can increase profit by discretely raising its price: it makes higher profit off incumbents while sacrificing an infinitesimal share of new entrants. Accordingly, Firm A instead follows a strategy such that, over a range of relatively low Firm B prices, it sells only to its incumbents, matching one-for-one each unit increase in p^B so as to keep incumbents indifferent. Then, when p^B is sufficiently high, Firm A can increase profit by reducing its price discontinuously and capturing a *discrete* share of new entrants, even while still charging a high price level to its incumbents. Figure D1 shows Firm A 's period-3 best response function, exhibiting the pattern just described.

Despite these discontinuities in the best responses, we identify a realistic calibration of the model under which there does in fact exist a Nash equilibrium in pure strategies in which both suppliers sell to new entrants in each period (a “no-sale” equilibrium, to borrow from Farrell and Klemperer (2007) language). We now discuss this calibration in greater detail.

There are 4 parameters to calibrate - c^A , c^B , τ , and s . The costs of production and the quality premium, τ , are chosen to target the long-run price levels and the market share. To be more precise, we seek to have the model's terminal-period outcomes match observed outcomes “late” in a product's life cycle. As for how “late” ought to be measured, the model suggests that we would like to observe market outcomes after the initial cohort of Taiwan's customers conclude their production runs. The evidence available from supplier agreements indicates that customers usually arrange for at least 3-year production runs, but with an option to renew.⁶² To be conservative, we focus on market outcomes after the first five years of a product's life. We have re-calibrated based on 3-4 year horizons, but the basic message in terms of price dynamics is unaffected.

Lastly, we select s . It is hard to obtain direct estimates of this, but the testimony of industry experts noted earlier suggests that switches are very rare. We also observe that customers remain in very long-term arrangements with suppliers. Fabless firms' annual reports to shareholders, for instance, show that most of the large fabless firms have a decades-long relationship with TSMC, which suggests that they will source new wafers to TSMC once production runs end on the current chip. Fabless firms that have initiated relationships with China's SMIC in the last decade have also tended to sustain those relationships for at least 4-5 years. Therefore, we choose to s to imply a “low” switch rate, which we take to be on the order 10 percent of Firm A 's period-2 customer base.⁶³ The calibrated parameter values are $c^B = 334$, $c_A = 400$, $\tau = 395$, and $s = 0.54\tau$.

As Table 5 shows, we replicate the price levels, but it is still somewhat difficult to replicate the full extent of Taiwan's sustained lead in the market. We show that the leader has 55.2 percent of the market in the terminal period but in the data, Taiwan claims 70.2 percent of aggregate sales.⁶⁴ The limitation of the model in this dimension is straightforward to understand: since price discrimination is prohibited, the lagging supplier is able, regardless of $c^A - c^B$, to carve out a share of the market, because the leader wants to set a high average markup and focus its sales on the high- y buyers.

D.2 The terminal-period solution

Here, we solve the period-3 problem by looking for an equilibrium in which each supplier retains its customer base and attracts a positive share of new entrants. The numerical analysis confirms that this equilibrium is the unique solution of this game under the proposed calibration.

⁶²For example, a contract between Quicklogic and TSMC states, “The term of this Agreement shall ... continue for a period of three (3) years, renewable annually as a rolling three (3) year Agreement.”

⁶³It is also difficult to achieve pure-strategy equilibrium if s is too high, as this widens the discontinuities in the best response functions.

⁶⁴This statement defines aggregate sales as the sum of Taiwanese and Chinese sales, which is consistent with the model treatment. Taiwan's share of the broader market is lower.

We begin with Firm A 's problem. Since the firm retains the mass $1 - \hat{y}$ in period 3 and attracts a share of new entrants equal to $\Pr [p_3^A < p_3^B + \tau y] = 1 - \frac{p_3^A - p_3^B}{\tau}$, its optimization problem is summarized by

$$\max_{p_3^A} (p_3^A - c^A) \left(2 - \hat{y} - \frac{p_3^A - p_3^B}{\tau} \right).$$

The first-order condition is

$$p_3^A = \frac{(2 - \hat{y})\tau + p_3^B + c^A}{2}.$$

Firm B solves a symmetric problem, setting a price

$$p_3^B = \frac{\tau\hat{y} + p_3^A + c^B}{2}.$$

Differencing these two, we obtain

$$p_3^A - p_3^B = \frac{2(1 - \hat{y}) \cdot \tau + c^A - c^B}{3}.$$

We compare this to the frictionless equilibrium that emerges if the switching cost is suspended in period 3. In the absence of a switching cost, buyers are not tied to any supplier, and firms compete anew for each buyer in each period. Accordingly, suppliers in period 3 face a measure 2 of buyers arrayed uniformly on the unit interval. Firm A attracts a share of buyers $1 - \frac{p_3^A - p_3^B}{\tau}$. It therefore sets its price to maximize $(p_3^A - c^A) \left(1 - \frac{p_3^A - p_3^B}{\tau} \right)$, yielding $p_3^A = \frac{\tau + p_3^B + c^A}{2}$. Firm B sets price $p_3^B = \frac{p_3^A + c^B}{2}$, and so the difference in prices is

$$p_3^A - p_3^B = \frac{\tau + c^A - c^B}{3}.$$

Comparing results, it follows that if \hat{y} is not too far from $1/2$, dispersion in the presence of a switching cost will mimic in the terminal period that which would emerge in the frictionless setting.

Figure D1: Firm A Best Response in Period 3

