PLANT- AND FIRM-LEVEL EVIDENCE ON “NEW” TRADE THEORIES*

James R. Tybout

Pennsylvania State University and NBER

July 16, 2001

JEL classifications: F1, L1, L6

*This paper was written for the forthcoming Handbook of International Economics, Basil-Blackwell, edited by James Harrigan. I am grateful to James Harrigan and participants in workshops at the NBER, the University of Toronto and the University of Texas-Austin for many useful comments.
Abstract

By relaxing the assumption of perfect competition, the “new” trade theory has generated a rich body of predictions concerning the effects of commercial policy on price-cost mark-ups, firm sizes, exports, productivity and profitability among domestic producers. This paper critically assesses the plant- and firm-level evidence on these linkages.

Several robust findings are identified. First, mark-ups generally fall with import competition, but it is not clear whether this phenomenon reflect the elimination of market power or the creation of negative economic profits. Second, import-competing firms cut back their production levels when foreign competition intensifies, at least in the short run. This suggests that sunk entry or exit costs are important in most sectors. Third, trade rationalizes production in the sense that markets for the most efficient plants are expanded, but large import-competing firms tend to simultaneously contract. Fourth, exposure to foreign competition often improves intra-plant efficiency. Fifth, firms that engage in international activities tend to be larger, more productive, and supply higher quality products. However the literature is mixed on whether international activities cause these characteristics or vice versa. Finally, the short-run and long-run effects of commercial policy on exports and market structure can be quite different. Both types of response depend upon initial conditions, sunk entry costs, and the extent of firm heterogeneity.

James Tybout
Department of Economics
Pennsylvania State University
University Park, PA 16820
jtyboutz@psu.edu
Two decades ago, in an effort to become more relevant, trade economists began
developing models with imperfectly competitive product markets. The result was a richer
body of theory that describes how commercial policy might affect price-cost mark-ups,
firm sizes, productivity, exports, and profitability among domestic producers. The
literature also yielded formal representations of the channels through which commercial
policy might influence growth. This chapter selectively surveys and interprets the firm-
and plant-level evidence that has emerged on these theories.

Section 1 focuses on three static predictions of the “new” trade theory that have
attracted attention from empiricists. First, protection can change firms’ pricing behavior,
thereby affecting the allocative efficiency of the economy and the distribution of real
income. Second, when trade policies affect prices, they generally also change the set of
active producers and/or their output levels. These adjustments induce productivity
changes through scale effects and market share reallocations. Finally, changes in the
intensity of foreign competition and/or in firms’ opportunities to export can affect their
technical efficiency.¹

Section 2 continues to discuss firm-level responses to policy reforms in terms of
pricing decisions, output levels, exports and productivity. However, rather than focus on
comparative statics, the models and evidence in this section are explicitly dynamic. They
allow for sunk entry costs, firm heterogeneity, and uncertainty. Thus they highlight the
relation between responses, expectations and initial conditions. Finally, Section 3 briefly
recaps what is known and what we would like to know, then mentions some directions
for future research.
1. **Static results: Mark-ups, Scale and Productivity**

   **A. Pricing**

   1. **Theory.**

   Except when collusive equilibria are considered, trade models with imperfect competition treat firms’ pricing decisions as determined by static profit maximization. Accordingly, the ratio of output prices \( p \) to marginal costs \( c \) is typically a decreasing function of the elasticity of demand \( \eta \) that firms face:

   \[
   \frac{p}{c} = \left( \frac{\eta}{\eta - 1} \right).
   \]  

   (1)

   It follows that when trade liberalization increases \( \eta \), mark-ups should fall.

   This kind of elasticity effect has been generated by a variety of modeling devices. For example, under the “Armington assumption” that foreign and domestic goods are imperfect substitutes, the demand elasticity for domestic goods rises as the relative price of foreign goods falls (e.g., Devarajan and Rodrik, 1991). Or, when protection takes the form of non-tariff barriers (NTBs), the removal of a quota can create heightened competitive pressures (Bhagwati, 1978). Finally, when liberalization makes more product varieties available (Krugman, 1979) and/or reduces the market share of domestic firms (Helpman and Krugman, 1985, pp. 85-112), these producers may perceive their demand elasticities to rise.

   When collusive equilibria are modeled, trade liberalization can change the pay-off to defecting, change firms’ ability to punish defectors, or make it more difficult to detect them (Prusa, 1992; Staiger and Wolak, 1989). It is possible that cooperative behavior will become unsustainable and mark-ups will fall. Or, some have argued that collusive
firms are likely to use the (exogenous) tariff-distorted price of imports as a reference price. By construction, models that begin from this latter pricing rule predict that trade liberalization will depress the price of import-competing goods.

2. Evidence.

Several simple methodologies have been used to link mark-ups to import competition. Prices and marginal costs are rarely observable, so each technique infers mark-ups indirectly. The most common approach is to use the price-cost margin (PCM)—that is, sales net of expenditures on labor and materials over sales. If one assumes that unit labor and material costs are flat with respect to output, and we interpret $c$ as short-run marginal costs, this statistic is a monotonic transformation of the mark-up in equation (1):

$$PCM_{it} = \frac{p_i q_i - c_i q_i}{p_i q_i} = \frac{p_i - c_i}{p_i},$$

where $q_{it}$ is the physical output of the $i^{th}$ firm in period $t$. The PCM is also current economic profits ($\pi_{it}$) over sales plus the competitive return on capital over revenues:

$$PCM_{it} = \frac{\pi_{it}}{p_i q_i} + \frac{(r_t + \delta)k_{it}}{p_i q_i},$$

where $k_{it}$ is the capital stock, $r$ is the market return on capital, and $\delta$ is the depreciation rate. By this logic, after controlling for the ratio of capital stocks to sales, variables that measure the intensity of foreign competition should contribute nothing to the explanation of price cost margins in industries where free entry drives profits to zero. On the other hand, if economic profits are present ($\pi_{it} > 0$), these variables should correlate negatively with the $PCM$ whenever trade liberalization increases demand elasticities or destroys collusive equilibria.
Most analyses of mark-ups based on the PCM begin from a simple regression like:

\[ PCM_{it} = \beta_0 + \beta_1 (k_{it} / p_{it}q_{it}) + \beta_2 I_{it} + \cdots + \varepsilon_{it} \]  \hspace{1cm} (2)

where \( i \) may index either firms or industries, and \( I_{it} \) is a proxy for the intensity of import competition—either the import penetration rate, the effective protection rate, or a license coverage ratio. (Import competition can only be observed at the industry level, so when firm-level data are used, \( I_{it} \) takes the same value for all firms within each industry.)

When industry-level, cross-sectional data are used, the typical finding is that “the ratio of imports to domestic consumption tends to be negatively correlated with the profitability of domestic sellers, especially when domestic concentration is high” (Schmalensee, 1989, p. 976).\(^5\)

A handful of studies have implemented equation (2) with plant-level panel data, controlling for permanent cross-industry differences in technology with industry dummies, and controlling for efficiency-related variation in mark-ups by including plant-level market shares.\(^6\) Results for Mexico (1985-90), Colombia (1977-85), Chile (1979-86), and Morocco (1984-89) all reveal the same basic pattern: “In every country studied, relatively high industry-wide exposure to foreign competition is associated with lower [price-cost] margins, and the effect is concentrated in larger plants” (Roberts and Tybout, 1996, p. 196, their italics). This pattern seems robust with respect to measures of import competition. In the case of Mexico, where it was possible to explore alternative measures of protection, the pattern appears whether one uses import penetration rates, effective protection rates, or license coverage ratios (Grether, 1996).
The standard interpretation for these PCM findings is that large firms and/or concentrated industries enjoy the most market power, hence their prices are the most responsive to heightened foreign competition. But other explanations are also plausible. For example, it might be that “relatively efficient industries are more profitable, and thus better able to compete against potential imports (low import penetration)” (Roberts and Tybout, 1996, p. 195). Or, concentration might reflect large sunk entry costs instead of market power (e.g., Hopenhayn, 1992). Then, rather than squeezing monopoly profits, unanticipated foreign competition cuts into the revenues that firms had expected would cover their entry costs and makes them sorry, ex post, that they entered (e.g., Albuquerque and Rebelo, 2000). If this latter interpretation is correct, one should observe output contractions in high-sunk-cost industries and exit in the others when trade is liberalized. There is some evidence that this happens, as I shall argue shortly.

An alternative methodology for linking foreign competition and pricing begins from the standard Tornqvist growth decomposition. Suppose the $i$th firm produces output according to $q_i = A_i h(v_i)$, where $v_i = (v_{i1}, v_{i2} \ldots v_{ij})$ is the vector of $J$ factor inputs it uses and $A_i$ measures its productivity level at time $t$. Then, suppressing time subscripts, output growth can be decomposed into a weighted-average of growth rates in the factor inputs and a residual productivity growth term:

$$
\sum_{j=1}^{J} \frac{\partial \ln(h)}{\partial \ln(v_{ij})} d \ln(v_{ij}) + d \ln(A_i) .
$$

Hall (1988) notes that when product markets are imperfect, this expression can be combined with equation (1) and the cost-minimization conditions $c_i = \frac{w_j}{\partial q_i / \partial v_{ij}}$, $\forall j$, to link output growth, input growth, productivity growth and mark-ups:
Further, he argues that a regression of output growth on the share-weighted rate of input growth, treating $d \ln(A_i)$ as the mean productivity growth rate plus noise, should reveal the price-cost mark-up as the slope coefficient.

By allowing $\eta$ to vary through time with trade reforms, one can test whether import competition affects mark-ups. Similarly, one can look for trade-related shifts in the mean rate of productivity growth. Several analysts have performed these exercises by fitting generalized versions of equation (4) to plant-level panel data. Studying Turkey and Cote d’Ivoire, respectively, Levinsohn (1993) and Harrison (1994) conclude that certain protected sectors had significant mark-ups during the sample period, and that these mark-ups fell with trade liberalization or exchange rate appreciation. Krishna and Mitra (1998) repeat the exercise using a panel of Indian firms and report “strong evidence of an increase in competition (as reflected in price-marginal cost mark-ups)” after the 1991 trade liberalization (p. 447). Thus studies based on Hall’s approach are consistent with studies based on the PCM—both methodologies suggest that heightened foreign competition forces down mark-ups among domestic firms.

However, Hall’s approach is subject to several criticisms. First, profit-maximizing firms should adjust their factor demands in response to productivity shocks. Hence consistent estimators of the slope coefficient in (4) require instruments that are correlated with factor stock growth but not with transitory productivity growth. It is difficult to argue that any available instruments satisfy this criterion, so mark-up estimates are
probably biased upward, and may exhibit spurious correlation with the trade regime (Abbott, Griliches, and Hausman, 1989).

Second, the framework presumes that firms face no adjustment costs. If some or all of the factors are subject to such costs they will be paid less than their marginal revenue product during upswings (when factor inputs are growing rapidly) and more during downswings (when factor inputs are growing slowly or shrinking). This measurement error is counter-cyclic and productivity growth tends to be pro-cyclic, so the estimated mark-up may be understated. Further, if import competition depresses demand for domestic goods, it may appear to eliminate monopoly power when it merely creates under-utilization of capacity.

Third, inputs and outputs are typically poorly measured and year-to-year fluctuations in these variables are particularly noisy. For example, due to gestation lags and changes in capacity utilization, growth in capital stocks is quite different from growth in capital services. Perhaps more importantly, growth rates in physical output are not really observed; what we observe is growth in nominal revenue deflated by a broad price index. If firms that expand rapidly also tend to drive their output prices down relatively rapidly, as one would expect in a differentiated product market, than true output growth is understated when input growth is rapid, and the mark-up estimate should be biased downward. I will discuss these measurement problems further in section C below.

B. The firm size distribution and its effects on productivity

1. Theory.

The output changes that accompany price adjustments depend upon whether markets are segmented and whether entry or exit barriers inhibit adjustments in the number of
producers. Head and Ries (1999) provide a useful synopsis of some alternative theories. In the absence of collusive behavior, unilateral trade liberalization either reduces firm size (when there are entry/exit barriers or markets are segmented) or leaves it unchanged (when entry and exit are free). Alternatively, when firms collude to slightly undercut the tariff-inclusive price of imports, trade liberalization cum free entry and scale economies forces import-competing firms that remain in the market to operate on a larger scale.

As Head and Ries (1999) acknowledge, the invariance of firm size under free-entry and no collusion is an artifact of the Dixit-Stiglitz demand system that is used in the models they consider. More generally, free entry is consistent with firm size adjustments whenever trade liberalization induces changes in the demand elasticities (\( \eta \)) that domestic firms perceive. In particular, when demand elasticities rise with liberalization, price-cost mark-ups are squeezed according to equation (1), and this should induce exit until the remaining firms can make up on volume what they lost on margin.

Business and labor groups care about policy-induced output adjustments because they are generally accompanied by job creation or destruction and by capital gains or losses. But trade economists have focused mainly on the ways the changes in the size distribution affect productivity. To summarize these effects, I shall adopt Tybout and Westbrook’s (1995) decomposition of industry-wide productivity growth. As before, let output at the \( i^{th} \) firm in year \( t \) be given by \( q_{it} = A_i h(v_i) \), but now write \( h(v_i) = \gamma(g(v_i)) \) where \( g(v_i) \) is a constant-returns homothetic function of the input vector, \( v_i \), and \( \gamma(\cdot) \) captures any scale economies. Also, let \( S_{it} = g(v_i)/\sum_{i=1}^{n} g(v_i) \) be this firm’s market share.
in terms of its input use and let $B_{it} = q_{it} / g(v_{it})$ be its productivity level. Then the rate of growth in industry-wide average productivity, $B_t = \sum_{it} B_{it} S_{it}$, can be decomposed as:

$$\frac{dB_t}{B_t} = \sum_{it} \left( \frac{d\mu_{it}}{g_{it}} \right) \left( \frac{q_{it}}{q_t} - 1 \right) + \sum_{it} dS_{it} \left( \frac{B_{it}}{B_t} \right) + \sum_{it} \left( \frac{dA_{it}}{A_{it}} \right) \left( \frac{q_{it}}{q_t} \right)$$

(5)

where $\mu_{it} = d \ln(q_{it})/d \ln(g_{it})$ measures returns to scale at the $i^{th}$ plant in year $t$. The first right-hand-side term above quantifies efficiency gains due scale economies at the margin, the second term quantifies gains due to market share reallocations toward relatively efficient producers, and the last term picks up residual intra-firm average efficiency changes that are unrelated to internal scale economies. I shall hereafter refer to these three quantities as scale effects, market share effects and technical efficiency effects.

In most trade models, all firms within an industry are characterized by a common technology and face identical demand conditions, so they expand or contract together in response to liberalization. Productivity gains or losses, when they are present, thus come exclusively from scale effects. However, several models deal explicitly with intra-industry heterogeneity and show how size adjustments (including entry or exit) might affect productivity through the market share effects. For example, Bond (1986) shows how heterogeneous workers might endogenously allocate themselves between entrepreneurial positions and salaried employment. In his “normal” case, protection of the industrial sector increases firm heterogeneity and lowers average productivity by drawing low quality entrepreneurs into managerial roles.

Melitz (2000a) obtains a related set of results in a forward-looking model of steady state trade with firm heterogeneity and imperfect competition. Movement toward
freer trade increases a country’s imports and erodes each domestic firm’s domestic sales and profits. Firms at the lowest end of the productivity distribution contract or exit, while firms at the high end of the productivity distribution expand their exports more than they contract their domestic sales. Accordingly, aggregate productivity improves.

Still another version of the same basic idea can be found in Andrew Bernard et al. (2000), who use a static model to study the effects of liberalization on the size and productivity mix of producers. They show that when firms use Bertrand pricing rules to compete, trade liberalization expands the market shares of the most efficient firms by providing them with larger export markets, and it forces firms at the low-end of the productive efficiency spectrum to shut down as they face competition from abroad.

What do we know empirically about size distributions and trade? Many analysts have fit cross-sectional regressions that relate firm size measures to the intensity of import competition, controlling for a few other factors like domestic market size. Whether the competition proxy is the import penetration rate or a measure of the industry-wide rate of protection, this literature finds that import competition reduces the average plant size, if it has an effect at all. Further, studies that include export shares in the explanatory variable set find that average plant sizes are relatively large in the export-oriented industries.

One limitation of this literature is that domestic output appears in the denominator of import penetration rates, so there may be spurious negative correlation between output per firm and this foreign competition proxy. A second problem is that causality may run from size to protection. Concentrated industries that are dominated by a few large producers may have an easier time coordinating lobbying efforts because they face less of
a free-rider problem. Finally, most of these studies presume that firms in all industries will adjust to foreign competition in the same way. This runs contrary to theory, which tells us that industries with low entry barriers, like apparel, are likely to show relatively less size adjust and more adjustment in the number of active firms.

Several more recent studies handle the first two criticisms by measuring exposure to foreign competition with policy variables like tariff rates and license coverage ratios; and by focusing on intra-industry changes in average firm size rather than cross-industry differences. Comparing industrial census data before and after Chile’s trade liberalization, Tybout, de Melo and Corbo (1991) find that plants in “sectors with relatively large declines in protection have shown a greater tendency toward employment reductions.” (p. 236). Similarly, Tybout and Westbrook (1995) find that during Mexico’s unilateral trade liberalization of 1984-89, firms in the sectors that underwent relatively large reductions in license coverage ratios tended to grow relatively slowly, while firms grew quickly in sectors with rapid export growth.

A subset of studies that deal with the first two criticisms also deal with the third by allowing intra-industry changes in firm size to vary with entry costs (proxied by industry-specific plant turnover rates). Perhaps the best is Head and Ries’s (1999), which uses the Canada-US Free Trade Agreement as a natural experiment. Their regressions suggest that “Canadian tariff reductions lowered scale [in Canada] while U.S. tariff reductions increased scale” (p. 309). Further, they confirm that entry barriers affect the way that firms respond: industries with high turnover (low entry costs) show relative mild reductions in scale in the face of heightened import competition. Roberts and Tybout (1991) obtain similar findings by contrasting industry-specific size distributions in Chile.
and Colombia and relating them to cross-country industry-specific differences in effective protection.

3. The Evidence, Part II: Trade-Induced Size Adjustments and Scale Efficiency.

In sum, the finding that foreign competition is associated with smaller firms in import-competing industries seems robust. There is also some evidence that foreign liberalization increases the size of exporting firms. We might reasonably ask, then, how dramatically these trade-induced adjustments have affected scale efficiency?

Most of the studies that address this question are based on computable general equilibrium (CGE) models, and they suggest that the scale-based efficiency gains when trade is liberalized can range from 1 to 5 percent of GDP. However, these findings are suspect for two reasons. First, while CGE models often predict firm-size expansion in all traded goods industries, the econometric evidence clearly suggests that firms in import-competing sectors contract when import competition intensifies, at least in the short run. Second, even if exporter expansion were the dominant effect of liberalization, it is unlikely that the gains in scale efficiency would amount to much. Although CGE studies often presume returns to scale ranging from 1.10 to 1.25 at the margin, this is probably a gross overstatement of the extent of unexploited scale economies. Exporting plants tend to already be the largest in their industry (Bernard and Jensen, 1995 and 1997; Bernard and Wagner, 1997; Aw, Chen and Roberts, 1997; Das, Roberts and Tybout, 2000). Thus they are not likely to exhibit much potential for further scale economy exploitation. Similarly, since most of the production in any industry comes from large plants, scale efficiency losses due to contraction in import-competing sectors are also typically minor (Tybout and Westbrook, 1996).
As an alternative to CGE analysis, Tybout and Westbrook (1995) used panel data on Mexican firms to estimate returns to scale (μₙ) as a function of size. Then they combine these estimates with the firm-specific growth rates observed during Mexico’s unilateral trade liberalization of 1984-1990 to implement equation (5). Although the cumulative weighted-average growth rate in output was 53 percent for the manufacturing sector, they find that the associated productivity growth rate due to scale efficiency effects was only one-half of one percentage point. This reflected the fact that large plants were operating in the flat portions of their average cost schedule, and these plants accounted for the bulk of the output adjustments.

4. The Evidence, Part 3: Market shares and productivity effects.

Of course, scale effects are not necessary to link size adjustments and productivity growth. Trade-induced market share reallocations can affect industry-wide performance so long as firms are heterogeneous in terms of Aₙ. What do we know empirically about these effects?

A simple way to address this question is to view firms’ sizes as reflecting their productivity. Then, if liberalization causes large firms to expand while small firms contract or exit, the associated market share reallocations should improve efficiency. From this perspective, the very robust finding that larger firms are more likely to export suggests that access to foreign markets allows the most efficient firms to become larger, thus pulling up industry-wide productivity levels.

However, studies that associate changes in trade protection with changes in the intra-industry size distribution deliver mixed evidence. Head and Ries (1999) find that large Canadian firms grew the most dramatically with U.S. tariff reductions, and they
shrank the most dramatically in response to Canadian tariff reductions. Similarly, Roberts and Tybout (1991) find that shrinkage in response to import competition—proxied by import penetration rates or effective protection rates—was relatively dramatic among the large firms in Chile (1979-85) and Colombia (1977-87). But Dutz (1996) finds that as Morocco dismantled NTBs during the 1980s, small plants shrank relatively dramatically and their exit probabilities increased relative to others'. Also, Tybout, de Melo and Corbo (1991) find that in Chile, reductions in effective protection between 1967 and 1979 were associated with balanced percentage reductions in employment across the entire size distribution.

These mixed findings could mean that the selection effects emphasized by Melitz (2000a) are not robust, or they could mean that size is a poor proxy for productivity, or both. To get at the latter issue, several studies measure share effects directly by constructing firm- or plant-specific $B_u$ trajectories. Tybout (1991) simply uses revenue per worker as his productivity measure and measures share-based gains for Chile (1979-1985), Colombia (1977-1987) and Morocco (1984-1987). He finds that market share reallocations contribute to productivity growth among tradeable goods, but his data span periods of major macro shocks rather than major trade liberalization episodes so it is difficult to argue that the gains are trade-induced. Using the same Chilean data set Pavcnik (2000) measures total factor productivity much more carefully and also finds that the shifting of market shares toward more efficient plants was an important source of efficiency gain during the sample period. However, she does not investigate the link between market share reallocations and foreign competition. Similarly, Tybout (1991), Liu (1993), Liu and Tybout (1996), and Pavcnik (2000) all find that exiting plants were
substantially less productive than surviving plants in Chile (and elsewhere), but none of these studies links this gap to import competition or exporting opportunities.¹⁹

Tybout and Westbrook (1995) have a better basis for inference in the unilateral Mexican liberalization of 1984-1989. Using equation (5), as well as a similar decomposition based on cost functions, they find that this liberalization was associated with efficiency gains, and that some of these gains were due to market share reallocations. However, they do not find strong evidence that rationalization effects were concentrated in the tradeable goods industries. Similarly, studying the Canada-U.S. FTA, Trefler (2001) finds little evidence that turnover-based productivity gains were concentrated in the industries subjected to the largest tariff reductions.²⁰

In sum, market share reallocations (including entry and exit) do matter, but it is difficult to find empirical studies that convincingly link these processes to the trade regime.²¹ This is not surprising, given that the effects of import competition on industrial evolution are inherently dynamic, and poorly captured by contemporaneous, reduced-form correlations. I will return briefly to this issue when I discuss transition dynamics in Section 3.

C. Other intra-firm productivity gains

Leaving aside productivity effects due to adjustments in the firm size distribution, there are many other linkages between commercial policy and efficiency gains. These are bundled together in the third right-hand-side (technical efficiency) term in equation (5). Some have to do with changes in the incentives to innovate or eliminate waste. For example, foreign competition or access to foreign markets may change the effort that a firm’s managers put forth and/or the rate at which they improve their products and
processes. However, a diverse body of theory suggests that the direction of change in efficiency hinges critically upon model specifics (Corden, 1974; Goh, 2000; Hart, 1983; Miyagawa and Ohno, 1995; Rodrik, 1992; Scharfstein, 1988; Voustden and Campbell, 1994.

Other effects on intra-firm productivity are more robust. As Ethier (1982) noted, intra-firm productivity gains may accompany trade liberalization if it expands the menu of intermediate inputs available to domestic firms. This allows each producer to match his or her input mix more precisely to the desired technology or product characteristics. Similar comments apply concerning access to capital goods, as Bradford de Long and Summers (1991) have stressed.

Trade may also act as a conduit for disembodied technology diffusion if firms learn about products by observing imported varieties, or by exporting to knowledgeable buyers who provide them with blueprints and give them technical assistance (e.g., Grossman and Helpman, 1991). Similar knowledge transfers may occur when domestic firms enter into joint ventures or sell equity to foreign multinationals, although these activities are less directly related to commercial policy.

Finally, domestic knowledge spillovers further confound the picture. If learning externalities are generated by experience producing a good, then changes in a country’s product mix induced by commercial policy can change the rate at which domestic efficiency grows (e.g., Krugman, 1987; Young, 1991). Whether trade liberalization helps or hurts in this respect depends upon which productive processes generate the most positive externalities, and whether they expand or contract as protection is dismantled.

Very little firm-level empirical work has been done on the popular notion that increases in the menu of available inputs improve productivity. This lack of micro evidence reflects practical difficulties with identifying a firm’s desired input mix, observing the actual input mix, and relating discrepancies between the two to measures of firm performance. It may also reflect a presumption that diversification of input bundles makes input use more heterogeneous at the industry level but not at the firm level.

Feenstra, Markusen and Zeile (1992) provide the only exception I am aware of. They argue that Korean conglomerates (chaebols) are vertically integrated, and thus when new intermediate producers join a conglomerate they effectively diversify the input menu for its final goods producers. Regressions confirm that, over a four year period, total factor productivity growth among final goods producers in 45 chaebols was positively correlated with the fraction of input expenditure going to new intra-chaebol intermediate goods suppliers.

This innovative study provides tantalizing evidence that input diversification contributes to productivity gains. However, data limitations prevent the authors from observing the connection between input variety and productivity as directly as one would like. Simultaneity bias is also an issue, since chaebols with high productivity growth are probably inclined to expand and incorporate new firms regardless of whether input diversification occurs.

2. The Evidence, Part 2: Import discipline effects.

It is much more common to relate firm-level productivity measures to proxies for the vigor of import competition. Most micro empirical studies that do so are based on first- or
second-order approximations to the production function \( q_{it} = A_{it} h(v_{it}) \), expressing the log of productivity, \( \ln(A_{it}) \), as a function of import competition proxies, \( I_{it} \), and noise. In the first-order (Cobb-Douglas) case, this amounts to estimating:

\[
\ln(q_{it}) = \sum_{j=1}^{J} \beta_j \ln(v_{ij}) + \gamma I_{it} + \varepsilon_{it} \quad (6)
\]

Alternatively, the log of productivity can be thought of as a draw from a one-sided productivity distribution (e.g., \( \alpha_{it} < 0 \)) plus an orthogonal transitory shock beyond the control of managers: \( \ln(A_{it}) = \alpha_{it} + \varepsilon_{it} \):

\[
\ln(q_{it}) = \sum_{j=1}^{J} \beta_j \ln(v_{ij}) + \alpha_{it} + \varepsilon_{it} \quad (7)
\]

Then, treating the distribution of \( \alpha_{it} \) as dependent upon import competition, one can investigate whether mean productivity levels and/or productivity dispersion respond to trade liberalization.

Regardless of whether one uses equation (6) or equation (7), one cannot measure import competition at the firm level. Thus its effect is identified by cross-industry or temporal variation in \( I_{it} \). The former type of identification is problematic because cross-industry regressions describe long run equilibria, and all industry characteristics—including import penetration rates, protection rates and concentration—are endogenous in the long run (Schmalensee, 1989). Nonetheless, Caves and Barton (1990) use equation (7) to characterize the \( \alpha_{it} \) distribution for each U.S. manufacturing industry, and they use cross-industry variation in \( I_{it} \) to infer that “import competition (measured by imports’ share of total supply) increases efficiency in industries whose domestic producers are concentrated” (p. 111).
Other studies use temporal variation in $I_u$ to link import competition and productivity via equation (6) or (7). As I mention in Tybout (2000, p. 34), these studies “tend to find that trade liberalization is associated with rising average efficiency levels” (Harrison, 1996; Nishimizu and Page, 1982; Pavcnik, 2000; Trefler, 2001; Tybout, de Melo and Corbo, 1991; Tybout and Westbrook, 1995). Similarly, liberalization drives down measured productivity dispersion relatively more in import-competing industries (Haddad and Harrison, 1993; Pavcnik, 2000; Tybout, de Melo and Corbo, 1991). Both sets of findings are consistent with the import discipline hypothesis, but they also could reflect the kind of selection effects described by Bond (1986), Melitz (2000a), and Bernard et al. (2000).

The implications of these studies are further clouded by methodological problems. Excepting Pavcnik (2000), they do not deal with the simultaneity bias that results from the dependence of factor inputs on productivity levels. Also, all of the studies use industry-wide price deflators to convert plant-specific revenues to plant-specific measures of physical output. But since products within each industry are heterogeneous, this procedure attributes relative price fluctuation to physical output fluctuation, and it thus confounds efficiency with monopoly power. Trade-induced reductions in measured “productivity” dispersion may be no more than the reductions in mark-ups among firms with market power that I discussed in section A above.

Finally, a general problem with this literature is that it tends to equate measured efficiency gains with welfare improvements. Thus when these gains are associated with trade liberalization, they are touted as a beneficial effect of foreign competition. But the costs of productivity gains are often embodied in overhead, license fees, training and
other items that do not get measured in the input vector. Further, the benefits these
expenditures generate are not fully reaped in the same periods in which their are incurred.
I know of no study that attempts to measure the present value of firms’ productivity-
enhancing expenditures and compare them to the present value of the resulting
productivity gains.


Does trade serve as a conduit for technology diffusion? Many studies have established
that exporters tend to be bigger, more skill-intensive, and more productive than their
domestically oriented counterparts (Aw and Hwang, 1995; Aw, Chen and Roberts, 1997;
Bernard and Jensen, 1995 and 1997; Bernard and Wagner, 1997; Handoussa, Nishimizu
and Page, 1986; Chen and Tang, 1987). Further, the case study literature on exporters
documents instances in which technologically sophisticated buyers transmit blueprints
and proprietary knowledge to the exporting firms. However, there is some doubt as to
whether the cross-sectional correlation between performance and exporting mainly
reflects causality from the latter to the former. Firms may self-select into export markets
and/or be sought out by foreign buyers because they are high quality.

Several authors have attempted to resolve this issue by studying temporal changes
in firms’ performance and their relation to export market participation. These studies
amount to Granger causality tests based on variants of the autoregressive specification:

\[
\ln(A_{it}) = \beta_0 + \sum_{j=1}^{J} \beta_j \ln(A_{ir-j}) + \sum_{j=1}^{J} \gamma_j y_{ir-j} + \cdots + \varepsilon_{it},
\]

where \( y_{it} \) is a dummy variable that indicates whether the \( i \)th firm exports in period \( t \).
Causality tests in this context establish whether exporting experience in the past helps
explain productivity in the present, once other determinants of current productivity
(including previous productivity) are controlled for. Given that $y_{it}$ responds to productivity shocks, the distributed lag $\sum_{j=1}^{J} y_{it-j}$ will be orthogonal to $\epsilon_{it}$ only when $\epsilon_{it}$ is serially uncorrelated, so it key to use a generous lag length ($J$) for the term $$\sum_{j=1}^{J} \beta_j \ln(A_{it-j}).$$

Fitting a version of equation (8) to plant-level panel data from Colombia, Mexico and Morocco, Clerides, Lach, and Tybout (1998) find very little evidence that past exporting experience improves performance. Bernard and Jensen (1999) obtain similar results using U.S.: “Exporting does not Granger-cause productivity, but does Granger-cause employment, shipments and wages.” (p. 14) On the other hand, Kraay (1997) finds that lagged $y_{it}$ values help explain current productivity among Chinese firms; Bigsten et al. (1999) find evidence of that exporting Granger-causes productivity among African firms, and Aw, Chen and Roberts (1997) obtain similar findings using census data from Taiwan and Korea.  

There are at least four problems with this literature. First, the contact between an exporting firm and its foreign client may occur well before export flows are actually observed in the data. Second, as with the import discipline literature, there is a strong tendency to interpret productivity gains as good, but no effort to quantify the costs of these gains. Third, the measures of performance are quite crude, as discussed in connection with the import discipline literature. Fourth, almost all of these studies focus on single conduits for technology transfer. But international activities like exporting, importing intermediates, importing capital goods, and selling equity abroad are often complementary, so firms pursue them in bundles (Kraay, Soloaga, and Tybout 2001).
Studies that focus on one at a time may generate misleading conclusions regarding channels of international technology diffusion.

Kraay, Soloaga and Tybout (2001) tackle the third and fourth methodological problems using the same data sets that Clerides et al. (1998) used to study learning by exporting. First, they document that international activities indeed come in bundles—exporting, importing intermediate goods, importing capital goods, and sales of equity to multinationals are clearly not independent activities. \(^{28}\) Next, by using a nested logit representation of demand for the differentiated products, and by exploiting information on the market share of each product, they are able separately measure product and process innovations at each firm. \(^{29}\) Finally, they relate quality trajectories and average cost trajectories to firms’ international activities, using generalized versions of equation (8). They find that activity histories don’t usually help to predict future product quality or reduce average production costs, once the histories of these performance variables are controlled for. Nonetheless, Colombian firms that engage in at least some international activities—especially those that import their intermediate goods—tend to have higher product quality. \(^{30}\) This finding suggests the kind of static efficiency effect that Ethier (1982) envisioned.

**D. Summary**

Measurement and methodological problems plague the literature I have reviewed in this section, but some findings seem robust. First, the evidence suggests that mark-ups fall with import competition. The most likely interpretation is that foreign competition increases the elasticity of demand that domestic firms face. However, it is not clear
whether these trade-induced reductions in mark-ups reflect the elimination of market power or the creation of negative economic profits.

Second, contrary to the predictions of many simulation models, import-competing firms cut back their production levels when foreign competition intensifies. This is not consistent with the Helpman and Krugman (1985) monopolistic competition model, under which some domestic plants would exit and the remaining plants would either remain the same size (if their demand elasticities do not change) or expand (if their demand elasticities rise). Instead, it suggests that sunk entry or exit costs are important in most sectors.

Third, trade does seem to rationalize production in the sense that markets for the most efficient plants are expanded. Further, if we discount the methodological problems with measuring productivity, most studies suggest that exposure to foreign competition improves intra-plant efficiency. (At what cost, we don’t know.) Finally, while firms that engage in international activities tend to be larger and more productive, it is not obvious whether the activities caused these characteristics or vice versa.

2. Transition Dynamics

The theories I have mentioned thus far describe static or steady state equilibria, and the regressions that give them empirical content deal mostly with patterns of contemporaneous correlation. But some important issues are inherently dynamic. For example, when a developing country dismantles its trade barriers and devalues its currency, as the World Bank often recommends, the effect of the new regime on the Central Bank’s foreign currency reserves will depend upon the resulting changes in the export trajectory. Further, the political support for a given reform package will depend
upon the associated changes in firms’ market values and employment trajectories that business representatives and workers anticipate. All of the literature that I have reviewed thus far is silent on these high profile issues.

The dynamic effects of policy reforms are difficult to characterize because they reflect complex decisions on the part of firms. Faced with an uncertain future, some managers find themselves weighing the earnings effects of shutting down plants and/or firing workers against the associated severance costs and the option value of retaining plants or workers for possibly better days. Others must weigh the sunk costs of breaking into foreign markets, building new plants, and/or hiring workers against the net revenue streams that these activities might generate. Their decisions are further complicated by the need to anticipate the decisions of other managers producing competing products. Below I discuss a nascent literature that tackles the relation between commercial policy reforms and industrial responses in settings with these features.

A. Export dynamics

1. Theory.

In the past 15 years, several theoretical contributions to the trade literature have incorporated sunk costs and uncertainty in dynamic models. Among the first to do so were the papers by Dixit (1989), Baldwin (1988) and Baldwin and Krugman (1989) on the role of sunk costs and expectations in driving exporters’ behavior. Generalizing their specification in anticipation of discussion to follow, let us specify an export profit function for the \(i^{th}\) firm that depends on the exchange rate \(e_t\), marginal production costs \(c_a\), a foreign demand shifter \(x_t\), and serially uncorrelated noise \(\varepsilon_{at}\):

\[
\pi^f(e_t, c_a, x_t) + \varepsilon_{at}.
\]

Further, let us assume that firms without prior exporting
experience must establish distribution channels, repackage their products, and learn bureaucratic procedures. Call the sum of these entry cost for new exporters \( \Gamma \). Then, defining the indicator variable \( y_i \) to take a value of unity in periods when the \( i \)th firm exports and zero otherwise, the pay-off from being an exporter in year \( t \) is may be written as:

\[
u(e_i, x_{it}, \epsilon_{it}, y_{it}, y_{it-1}) = \begin{cases} 
\pi f (e_i, x_{it}) + \epsilon_{it} & \text{if } y_{it} = 1; y_{it-1} = 1 \\
\pi f (e_i, x_{it}) - \Gamma + \epsilon_{it} & \text{if } y_{it} = 1; y_{it-1} = 0 \\
0 & \text{if } y_{it} = 0; y_{it-1} = 0
\end{cases}
\]

Presuming that the vector \((e_i, c_{it}, x_i)\) follows a first-order Markov process, risk-neutral managers do best to choose a sequence of decision rules,

\[
y_{it} = g_j (e_i, c_{it}, x_{it}, \epsilon_{it}, y_{it-1}),
\]

that maximizes their expected profit stream from export market participation:

\[
E_{\text{t}} \sum_{j=1}^{\infty} u(e_j, c_{j}, x_{j}, \epsilon_{j}, y_{j}, y_{j-1}) \delta^{-j}.
\]

Equivalently, their patterns of export market participation should satisfy the following Bellman equation:

\[
V(e_i, x_{it}, c_{it}, \epsilon_{it}, y_{it-1}) = \max_{y_{it}} \left\{ \pi f (e_i, c_{it}, x_{it}) - (1 - y_{it-1}) \Gamma + \epsilon_{it} + \delta E_{\text{t}} V(e_{t+1}, c_{t+1}, x_{t+1}, \epsilon_{t+1}, y_{t+1}) \right\}
\]

Here expectations are taken conditioned on \((e_i, c_{it}, x_i)\) and the Markov process that govern this vector’s evolution.

This framework implies that seemingly identical policies and macro conditions can lead to different levels of exports, depending upon how many firms have a history of export market participation: When firms have no exporting experience, they weigh the sunk costs of entry against the expected profit stream. But when most firms are already
exporters, the aggregate response to export incentives reflects volume adjustments and has little to do with entry costs. Second, firms that begin exporting in response to a shock—say, a large devaluation—may not cease exporting when that shock is reversed. Third, expectations about future exchange rate trajectories and commercial policies may play a critical role in determining whether firms invest in becoming exporters today. Finally, export responsiveness to any shock or regime switch depends critically upon the amount of cross-firm heterogeneity in marginal costs and foreign demand, \( x_u \). Many firms may be poised on the verge of exporting, or just a scattered few.

2. Evidence.

Several studies have explored the empirical relevance of the sunk-cost export model sketched above. Roberts and Tybout (1997) begin from the implication of (9) that firms will find it optimal to export whenever:

\[
\pi^f(e_t, c_{it}, x_{it}) - (1 - y_{it-1}) \Gamma_S + \epsilon_{it} \\
+ \delta [E_t V(e_{it+1}, c_{it}, x_{it+1} + \epsilon_{it+1} | y_{it+1} = 1) - E_t V(e_{it+1}, c_{it}, x_{it+1} + \epsilon_{it+1} | y_{it+1} = 0)] > 0
\]

The second bracketed term describes the option value of being an exporter in period \( t \), that is, the expected current value of being able to export in period \( t+1 \) without having to pay sunk entry costs. Accordingly, its magnitude depends upon expectations about the future operating profits one might generate by exporting. Combining terms that depend upon current values of the state variables, the \( i^{th} \) firm will do best to export whenever:

\[
f(e_t, c_{it}, x_{it}) + y_{it-1} \Gamma_S + \epsilon_{it} > 0,
\]

where:

\[
f(e_t, c_{it}, x_{it}) = \pi^f(e_t, c_{it}, x_{it}) + \delta [E_t V(e_{it+1}, c_{it+1}, x_{it+1}, \epsilon_{it+1} | y_{it+1} = 1) - E_t V(e_{it+1}, c_{it+1}, x_{it+1}, \epsilon_{it+1} | y_{it+1} = 0)]
\]
Using a reduced form approximation to \( f(q) \), and assuming a particular distribution for the error term, \( \varepsilon_{it} \), this equation implies a dynamic discrete choice model of export market participation. Roberts and Tybout (1997), Bernard and Jensen (1999), Campas (1999) and Sullivan (1997) have fit this model as a dynamic Probit or logit and tested whether sunk entry costs affect export market participation. This simply amounts to testing whether lagged exporting status affects current status, once the other sources of persistence in behavior have been controlled for: \((x_{it}, e_t, \varepsilon_{it})\). Critically, if other sources of persistence are not completely controlled for, this approach to inference mis-attributes serial correlation in exporting status to sunk costs. So it is important to treat \( \varepsilon_{it} \) as a serially correlated disturbance when estimating the equation.

The universal finding of these studies is that sunk costs are important. Even after serial correlation in \( \varepsilon_{it} \) is treated, the probability that a firm will export, given \((x_{it}, e_t, \varepsilon_{it})\), can be up to 0.70 higher if it exported last period. From this, researchers have typically concluded that export aggregates are subject to important hysteresis effects and that sunk costs matter.

More recently, Das, Roberts and Tybout (2001) revisited the question of how sunk costs shape export responsiveness among Colombian chemical producers. Instead of using a reduced-form version of the decision rule, they fit a structural model that explicitly describes the profit function and the autoregressive processes that govern the vector \((x_{it}, c_{it}, e_t, \varepsilon_{it})\). Using their estimates, they then examine the option value of export market participation for each firm:

\[
\delta \left[ E_t V(e_{t+1}, x_{it+1}, \varepsilon_{it+1} \mid y_{it} = 1) - E_t V(e_{t+1}, x_{it+1}, \varepsilon_{it+1} \mid y_{it} = 0) \right].
\]

This expression measures...
the importance of expectations about the future in shaping exporting decisions. They find that it is quantitatively important for small scale exporters, whose foreign demand is relatively limited. However, the firms that supply the bulk of total exports earn operating profits that far exceed the option value term. Hence, hysetersis effects are important only fringe players in the export markets, and aggregate exports are relatively insensitive to history or expectations. Put differently, if one is interested only in the aggregates, sunk entry costs and the subtleties they introduce may be ignorable for many industries.\(^\text{32}\)

One robust finding concerning exporters is that they tend to sell very small fractions of their output abroad (Aw, Chen and Roberts, 1997; Sullivan, Roberts and Tybout, 1996; Campas, 1999). In principle this could mean that foreign demand for each firm’s product is very limited and inelastic, but this is not the way most people view foreign markets. A second explanation is that firms export just enough to exploit duty drawback schemes and purchase the imported intermediates or capital goods at duty-free prices. To my knowledge this hypothesis has not been pursued, although it would be easy to do so. A third hypothesis is that firms export partly to diversify their earnings stream, exploiting the imperfect correlation between foreign and domestic shocks. Small stable shares in foreign markets might be rational under these assumptions.

Maloney and Azevado (1995) develop a simple model of this diversification motive for exports and fit it to firm-level panel data from Mexico. They find, among other things, exchange rate volatility and the covariance between domestic and international demand shocks are significant determinants of export volumes. Hence, for example, when an over-valued exchange rate is allowed to float, the export response may be counter-intuitive.
In sum, the initiation of exports appears to invoke some sunk start-up costs. These costs matter a good deal for marginal exporters, but are unimportant relative to the operating profits that large exporters earn. Thus their effect on aggregate export responses to regime shifts or exchange rate shocks may not be large. Other determinants of export responsiveness that may be relevant include risk diversification considerations and domestic market demand shocks (when marginal costs aren’t flat). There is some evidence that the former matters; the latter remains largely unexplored.

B. Industrial Evolution

1. Theory.

Theoretical models of industrial evolution demonstrate how the combination of sunk entry costs with imperfect foresight and cross-firm heterogeneity can lead to continual flux in the population of active firms (Jovanovic, 1982; Hopenhayn, 1992; Ericson and Pakes, 1995). They also describe the implications of this flux in terms of job turnover patterns and productivity growth. However, very little theoretical work has been done on the effects of commercial policy in an economy with these features.

Two exceptions merit note. The first is Melitz (2000a), who focuses on the relation between openness and the steady state distribution of firm types (see Section 1). The other is Albuquerque and Rebelo (2000), who abstract from intra-industry heterogeneity to derive some analytical results about dynamic responses to trade liberalization. Only the latter paper deals with transition issues, so I shall focus on it here.

Albuquerque and Rebelo consider an open economy with homogeneous firms in each of two sectors. New firms must pay a sunk fee to initiate production, so incumbents may earn positive profits in steady state without inducing entry and multiple equilibria
are possible. Further, responses to policy shocks depend upon the pre-reform equilibrium. When profits net of entry costs are zero in the exportable goods, and when profits before entry costs are zero in the import-competing sector, small reductions in the rate of protection should generate entry in the former and exit in the latter. Unanticipated reforms also induce inter-sectoral reallocations of variable factors in the period before entry and exit occur. Pre-announcing eliminates this short-run adjustment period. On the other hand, if the economy begins from an interior steady state and reforms are too mild to trigger entry or exit, the effects of policy reforms are limited to variable factor movements and capital gains or losses for the owners of incumbent firms.

The dichotomy between responses beginning from zero-profit versus interior profits is an artifact of the assumption that firms within each sector are homogeneous. Intra-industry heterogeneity will generally mean that operating profits are close to zero for the marginal incumbent, and profits net of entry costs will be close to zero for the marginal entrant. Nonetheless, the results I mentioned above suggest how responses to reforms should depend on the density of incumbents and potential entrants near the zero-profit margin.

2. The Evidence, Part 1: Descriptive Studies. It is well established that, even within narrowly defined industries, plants are quite heterogeneous in terms of their size and measured productivity. (See, for example, the references in part I above.) Also, simultaneous plant entry and exit are the norm, as are market share reallocations and job creation/destruction among incumbent firms. (e.g., Dunne, Roberts and Samuelson, 1989; Davis, Haltiwanger and Schuh, 1996; Baldwin, Dunne and Haltiwanger, 1998, Roberts and Tybout 1996). These are the stylized facts
inspired the modern theory of industrial evolution and they are commonly cited as evidence of its relevance.

We know much less about the effects of commercial policy shocks on industrial evolution patterns, nor how these effects depend upon the initial population of firms. A number of studies document patterns of contemporaneous correlation between openness, firm size distributions, and entry/exit or market-share-based efficiency gains (see Section 1). There is also a small amount of evidence relating openness to patterns of job turnover (Levinsohn, 1999). However, these studies tell us little about the dynamic responses to reforms when threshold costs and uncertainty make firms’ adjustments forward-looking, gradual, and/or dependent upon initial conditions.

3. The Evidence, Part 2: A Structural Model.
Lu and Tybout (2000) attempt to go beyond patterns of contemporaneous correlation and quantify these dynamic relationships. Drawing heavily on Ericson and Pakes (1995) and Pakes and McGuire (1994), they develop an empirical model with sunk costs, heterogeneity and uncertainty. It portrays an import-competing industry populated by a finite number of potential entrepreneur/owners, including those already in the industry (incumbents) and those contemplating entry (potential entrants). Each incumbent is characterized by a unique product and a time-varying productivity index that summarizes both his product’s appeal and his unit production costs. Imports are represented by a single foreign variety whose price responds to exchange rate shocks and commercial policy reforms, but not to domestic producers’ behavior.

Entrepreneurs in this industry play a Markov-perfect dynamic game against one another. Each period, each entrepreneur attempts to maximize his discounted net profit
stream, given the available information set. Potential entrants choose whether to enter the market, given their privately observed entry costs. Incumbents decide whether to remain in the market or exit, given the privately observed scrap value of their firms. The incumbents who remain active engage in Bertrand-Nash product market competition with one another, given the current price of the import-competing good and a simple logit demand system.

At the beginning of each period, all entrepreneurs learn the productivity level of each incumbent firm (industry structure), as well as the current realizations on the number of consumers and the real effective exchange rate (market conditions). If an incumbent firm remains in the industry, its productivity evolves from period to period according to a common knowledge exogenous Markov process, as do the exchange rate and the number of consumers. Firms solve for their optimal strategies and make their exit or entry decisions simultaneously. From period to period, the industry structure evolves with entry, exit, and random shocks to each firm’s productivity.

Using Colombian panel data on the pulp and paper industry, Lu and Tybout (2000) estimate the demand parameters of their model. Combined with observed market shares these allow them to impute product quality trajectories for each producer, and to estimate the associated Markov processes. Finally, given these primitives, they calibrate the entry cost and exit cost distributions so that simulated plant turnover rates approximate the industry’s actual figures.

Lu and Tybout’s main computational experiment is to simulate responses to a change in the exchange rate process that gradually intensifies import competition. The impact effect of this regime switch is to squeeze price-cost mark-ups, just as the
econometric evidence suggests. However, the new exchange rate regime also discourages entry (but not exit), so over time, the number of domestic producers gradually shrinks. With the menu of varieties falling, elasticities of demand for each variety fall too, allowing the remaining incumbents to restore their mark-ups and cover their operating costs. This transition path suggests that the robust margin squeeze effects and output contraction effects identified by contemporaneous correlation patterns may not be permanent (see Sections 1.A and 1.B above).

Although consumers initially benefit from cheaper imported goods and cheaper domestic goods, they are ultimately left with fewer domestic varieties at prices close to pre-appreciation levels. Hence, in the scenario that Lu and Tybout analyze, the present value of consumer welfare actually falls with heightened import competition. Producers suffer capital losses, of course, so they are worse off too.

Extra costs are also imposed on workers, who endure higher job destruction rates during the transition period. Indeed, the job turnover effects predicted by this model are implausibly high, suggesting that it should be generalized to including severance costs and/or screening costs, as in Hopenhayn and Rogerson’s (1993) simulations. By the same token, the apparent importance of hiring and firing costs means that firms’ expectations are critical and suggests that static calculations of the employment effects of trade policies can be very inaccurate.

Finally, this framework provides a conceptually rigorous way to address the question of how changes in the intensity of import competition affect the market-share-based efficiency changes that are described by the second term in equation (5). Lu and Tybout (2000) find that this type of efficiency gain is small for two reasons. First, most of
the adjustment in varieties comes from less entry rather than more exit. Incumbent firms that are relatively inefficient don’t increase the rate at which they jump out of the market because their entry costs are already sunk, their scrap values are small, and they perceive a possibility that conditions will improve in the future. Second, the firms that do enter or exit account for a relatively small fraction of total production. This is consistent with what we actually observe in the data, as discussed in Section 1.B.

These simulations are subject to several criticisms. Most fundamentally, they are partial equilibrium and thus do not document the capital gains and growing number of product varieties in sectors that benefit from exchange rate appreciation. Second, they do not permit the number of imported varieties to adjust. If foreign firms face sunk entry costs when breaking into the domestic market, there will probably be some new ones that are induced to enter by the change in the exchange rate regime. Third, the model is highly stylized in many respects, including the demand system, the productivity growth process (which is presumed exogenous) and the distributions for entry costs and scrap values. Nonetheless, at a minimum the model demonstrates that conclusions based on contemporaneous patterns of correlation can be very misleading, and it brings together in a unified framework the phenomena that firms, workers, and consumers care about.

3. An Agenda

I shall close with a few observations on directions for future trade research using firm- or plant-level data. First, as the previous section suggests, I am personally enthusiastic about the new insights that we might gain from dynamic structural models that link trade
regimes and industrial evolution. These models suffer from some serious limitations, but they integrate many pieces of the response story that were heretofore treated in disjoint literatures. They also provide a basis for counter-factual simulations in the presence of threshold costs, uncertainty and heterogeneous firms. As computers become more powerful and solution algorithms improve I am hopeful that econometrically estimated industrial evolution models can be made more realistic and used for applied policy work.

Second, despite the large volume of research on the link between trade and productivity, there are several senses in which this literature might be improved. One is to get away from pretending that firms in manufacturing industries produce homogeneous products, and to deal with pricing, output and productivity measurement in unified frameworks (e.g., Melitz, 2000b). Another is to tighten the link between theory and tests. Theory has emphasized the effects of enhanced input variety—including both capital and intermediate goods—and, more recently, efficiency gains due to geographic agglomeration. But we have very little direct micro evidence on the importance of either. These are relatively difficult topics to tackle, but creative empiricists should be able to make progress on both fronts.

Finally, although the relationship between trade and wages has attracted considerable attention, we have only limited evidence on the micro details of worker displacement, job searches processes and reemployment patterns that are triggered by changes in the trade regime. The census bureaus of several countries (including the United States) have recently devoted some resources to matching household survey data with establishment survey data, so the characteristics of plants and workers can now be analyzed together and workers can be tracked as they change jobs. These matched data
sets should provide a much better basis for inference on the employment effects of commercial policy reforms or changes in the exchange rate regime.
References


———, “Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?” *Journal of Economic Literature* 38 (2000):11-44.


Endnotes

1 I shall ignore the empirical literature on multinationals and foreign direct investment, which is treated in chapter __.

2 In these models protection takes the form of institutional arrangements for anti-dumping measures.

3 See Head and Ries (1999) for discussion and references.

4 This measure presumes that intermediate input use and labor use are proportional to output, and the proportions are fixed across plants. See Schmalensee (1989) for further discussion of the limitations of PCM as a performance measure.

5 See also Lee (1991) and Roberts and Tybout (1996, pp. 188-199) for surveys of the literature on developing countries.

6 “Efficient plants should be larger and have higher profits, so a positive correlation is generally expected bewteen market shares and price-cost margins, regardless of whether firms have market power …” (Roberts and Tybout, 1996, p. 196).

7 Generalizations have included allowing for non-constant returns to scale, and letting η and the mean productivity growth rate vary across firms. For example, see Harrison (1994).

8 Pakes and Griliches (1984) estimate that it may take several years for newly installed capital to reach full productivity.

9 Klette and Griliches (1996) and Melitz (2000b) discuss the consequences of this measurement problem for estimates of production function parameters.

10 The most common form of entry/exit barrier is sunk start-up costs. Firms will continue to operate so long as their expected earnings stream covers their expected future earnings.
expenditures, even if *ex post*, they discover they cannot also recoup the sunk costs that they paid to enter (e.g., Albuquerque and Rebelo, 2000). Uncertainty about future market conditions is likely to increase the option value of remaining in operation, effectively compounding persistence in status. Firms that enjoyed excess profits before import competition intensified will also fail to exit.

11 I will not treat external returns to scale because these are nearly impossible to measure.

12 The cases he analyzes are: autarky versus free trade, more versus fewer countries in a customs union, and high versus low non-tariff barriers (at home and abroad).

13 These studies span a wide range of countries. See Scherer et al. (1975), Muller and Owen (1985), Baldwin and Gorecki (1986, Table 7-1), Caves (1984), and Schwalbach (1988). Tybout (1993a, Table 2a) provides further details on these studies.

14 This pattern is less apparent when size is measured with output or value-added, suggesting that efficiency gains occurred in the import-competing industries.

15 On the other hand, they find no significant cross industry correlations between firm size and effective protection rates or import penetration rates.


17 The size-productivity linkage is common in models with heterogeneous firms. See, for example, Hopenhayn (1992), Melitz (2000a), and Bernard et al. (2000).

18 His decomposition does not distinguish intra-plant productivity gains due to scale efficiency from other sources of intra-plant gains. Bernard et al. (2000) show that revenue
per unit output is a monotonic function of true total factor productivity if firms compete
Bertrand.

19 In any case, as Liu and Tybout (1996) point out, the impact of this differential on
productivity growth was minor, given that they typically account for a very small fraction
of output.

20 Trefers’s (2001) intra-industry data are grouped by plant size, so he cannot rule out the
possibility that the FTA generated productivity gains through reallocations within size
classes or through entry and exit.

21 Bernard and Jensen (2000) link entry and exit patterns to trade indirectly by arguing
that, with output prices pinned down by international arbitrage, Rybczynski effects should
induce net entry in the sectors intense in the factors that are growing relatively rapidly.
They confirm this conjecture using data from the U.S., first with cross-industry
regressions at the national level, then with similar regressions at the regional level. They
find that where human capital and physical capital have grown relative to unskilled labor,
exit rates have been low among skill-intensive goods and high among low skill goods.

22 While not at the firm level, Feenstra, et al. (1999) do use detailed data on trade flows to
link sectoral productivity to the diversity of final good and upstream exports.

23 Detailed discussions of this approach to productivity analysis may be found in the
“stochastic frontier” literature (e.g., Greene, 1993).

24 There is also evidence that innovative activities are stimulated by import competition.
See Blundell, Griffith and van Reenen (1999).

25 Much of this literature focuses on East Asia. Pack (2000) and Westphal (2001) provides
recent surveys.
Both Kraay (1997) and Bigsten et al. (1999) are based on annual data with short lag lengths, $J$, and do not provide tests for serial correlation. Hence they may be picking up spurious correlation. Aw, Chen and Roberts (1997) compares censuses at five year intervals, so their study is likely to suffer from this problem.

For interesting discussions of the case study literature on pre-exporting contacts with buyers, see Pack (2000) and Westphal (2001).

Aw, Roberts and Winston (2001) document similar dynamic complementaries between worker training, exporting, R&D and exporting using multinomial probit models.

One unappealing feature of their approach is that must assume that the ratio of physical output to intermediate input use is constant across all producers in a given 4 digit industry and geographic region.

Given the way that Kraay, Soloaga and Tybout (2001) impute quality, this is almost a corollary to the finding that firms engaging in international activities have large domestic market shares.

Domestic product market conditions are kept out of the analysis by assuming flat marginal cost schedules with respect to output.

Using a reduced-form econometric model and descriptive statistics, Campas (1999) draws similar conclusions from Spanish data.

Levinsohn finds that job turnover patterns in Chile during the 1980s were not closely linked to commercial policy or exchange rate shocks He does argue, however, that turnover rates were higher among tradeable goods than among non-tradeables. Thus, liberalization in economies like Chile’s should reduce job security, and may meet resistance for the political economy reasons details by Fernandez and Rodrik (1991).
This result is partly an artifact of the demand system they use, which probably overstates the value consumers place on goods with small market shares.