GDC Working Papers

Noel Gaston and Doug Nelson

The Employment and Wage Effects of Immigration:
An Overview of Theory, Method and Results

No. 6, August 2007

Prepared for The World Bank.
Parts of this paper have appeared as Gaston and Nelson (2000a and 2002), and Hijzen and Nelson (2006).
We may not be living in the age of mass migration, but we are surely living in an age of mass migration. From 1965 through 2005 a fairly constant 2.2% of the world population have been migrants. However, this has involved an increasing rate of change to keep pace with the growing world population: the stock of migrants grew at 1.2% from 1965-1975; 2.2% from 1975-1985; and 2.6% from 1985-1990. More importantly, for the purposes of this paper, relative to regional population, the share of migrants in the US and Canada rose from 6% in 1965 to 8.6% in 1990 and 13% in 2005; while the share in Western Europe rose 3.6% to 6.1% and then 12% over the same period. This period has also seen a substantial shift toward developing countries as source countries for this migration: in the United States this share rose from 42% in 1960-1964 to over 80% in the 1980s through 2005; in Canada this share rose from 12% to over 80%; while this share in Australia rose from 7% to over 90%. In the 1990s, Germany and Austria experienced very large flows from Eastern Europe as well.

As is well known, this period of rising immigration of unskilled workers coincides with a period of strong deterioration of the relative (and possibly the real) return to native unskilled labor in nearly all industrial countries (Levy and Murnane, 1992; Davis, 1992; Blackburn and Bloom, 1995). While much of the research on the causes of this phenomenon has focused on demand-side factors, with special emphasis on international trade and skill-biased technical change, unskilled immigration has received a considerable amount of attention as a possibly relevant supply shock. However, unlike the case of the relationship between international trade and labor market outcomes, where there is considerable disagreement on the facts, the overwhelming majority of empirical studies agree that there is essentially no statistically significant effect of immigration on labor market outcomes, with

---

1 Carter and Sutch (1998), Hatton and Williamson (1998), and Sassen (1999) emphasize that large scale migration is not a new phenomenon, and was arguably quantitatively more significant in earlier periods. However, as Sassen (1999) points out, the development of democracy, nationalism, and welfare states have made immigration a politically more difficult, and potentially more explosive, issue in contemporary times than in earlier times.

2 The statistics in this paragraph are drawn from Zlotnik (1999) as extended to 2005 by UN (2006).
the possible exception of the least skilled domestic workers (i.e. that small share of the work force that are high school dropouts).

The apparent occurrence of a large-ish supply shock with minimal economic effect has produced a sizable literature, primarily among labor economists, attempting to either account for the measured smallness or to generate larger numbers. Because the overwhelming majority of empirical research on the labor market effects of immigration has been done by labor economists in the context of a relatively common framework, we will develop this framework and then survey the main results from this literature. While there is fairly widespread agreement on the smallness of measurable effects of immigration, the interpretation of this fact is a matter of some considerable dispute. We will first consider accounts that attempt to retain the main structure of the labor theoretic framework involving primarily issues of internal migratory response to international migration or the labor market microstructure issues like the possible presence of segmented labor markets. Where the inability to consistently identify significant effects from immigration was seen as something of a crisis for labor economists (something like the Leontief paradox for trade economists), trade economists have been quick to suggest that this finding is, at least prima facie, consistent with standard trade theoretic models. We will discuss the small body of research that seeks to empirically implement a trade theoretic approach to migration, but we will also emphasize that the essential difference has to do with the interpretation of a generally agreed phenomenon–small measured effects from a sizable shock.

Our analysis will be presented as follows: the next section presents a brief sketch of the fundamental differences between the labor and trade theoretic approaches to the analysis of immigration; section II reviews research based on the standard labor theoretic model; section III turns to attempts by economists using a trade theoretic approach to rationalize or extend the empirical findings discussed in section II; Section IV looks at the means by which
market economies respond to immigration shocks; and section V concludes with a discussion of the implications for policy evaluation.

I. Evaluating Labor Market Effects of Immigration: Two Simple Frameworks

Like most economists, when confronted with a macro phenomenon, like the immigration-wage or immigration-unemployment nexus, labor and trade economists are likely to reach first for fairly highly aggregated, perfectly competitive models. Thus, in this section and for most of the rest of the paper, the theoretical frameworks we will be considering are characterized by complete and perfectly competitive markets, and constant returns to scale production functions. Before being a bit more explicit, we comment briefly on three dimensions that discussions often seem to suggest divide the labor and trade economists: the absence or presence of commodity trade; the exogeneity or endogeneity of international labor flows; and the dimensionality of the model. We will ultimately conclude that, from the point of view of motivating or evaluating empirical work, only the last is of genuine importance.

Let us start with dimensionality. In either case, we characterize production via a standard neoclassical production function:

\[ y_j = f^j \left( z_j \right), \]  

where \( j \) denotes a sector, and we drop it in the one sector case, \( z_j \) is a vector of inputs, and \( f^j(A) \) is a linear homogeneous, strictly quasi-concave function.\(^3\) A convenient representation in either case is the unit-value isoquant—the locus of all input combinations that yield $1 worth of output (i.e., letting price be \( P_j \), this is the \( 1/P_j \) isoquant).

\(^3\) Where we need a general representation we will denote the set of all factors as \( I \) and its dimensionality as \( m \), while the set of all goods is \( J \) with dimensionality \( n \), \( i \) will index members of \( I \) and \( j \) will index members of \( J \).
In figure 1 we suppose that $z' = \{S, L\}$, skilled and unskilled labor, denotes the economy’s endowment, and the slope of the ray from the origin through $z'$ identifies $s = S/L$ the equilibrium input ratio. From cost minimization and competitive markets we know that, in equilibrium, the slope of the isoquant will be equal to $\omega = -w_L/w_S$. Thus, an increase in the relative endowment of unskilled labor (from $z'$ to $z''$), a fall in $s$, straightforwardly leads to a fall in $\omega$.\(^4\) Furthermore, if we suppose that the price of the final good is fixed, this translates to a real increase in the wage of $S$ and a real decrease in the wage of $L$.\(^5\) The entire adjustment has occurred through a change in relative factor prices. This is the basis of the standard labor theoretic approach to determining the effect of immigration on a host economy.\(^6\) As we shall see in the next section, this setup provides a set of identifying assumptions that permits a very straightforward econometric analysis of the price (or, *mutatis mutandis*, employment) effects of increased immigration.

Now suppose that we make only one change in the model, we add one more good and assume that good 1 is always $S$-intensive relative to good 2. Figure 2 labels the good from figure 1 “good 1” and the new good “good 2”. Since both of the isoquants are unit-value isoquants, they must be tangent to a common $\$1$ isocost line. As with the one good case, the tangent gives $\omega$, common to both industries as a result of free inter-sectoral factor mobility,

\(^4\) An alternative representation of this is that the value marginal product curve for unskilled labor is a downward sloping function of $1/s$. In the two factor case, with $S$ and $P$ (the price of final output) fixed, this is just the demand curve for unskilled labor.

\(^5\) This follows from the standard weighted-average property of price changes (Jones, 1965):

$$\hat{P} = \hat{\theta}_S \hat{w}_S + \hat{\theta}_L \hat{w}_L,$$

where the $\hat{\theta}$’s are distributive shares and the hats denote proportional changes. Thus, in Figure 1, $\hat{w}_S > \hat{P} = 0 > \hat{w}_L$. It is also straightforward to show that the gain to domestic skilled labor exceeds the loss to domestic unskilled labor. With appropriate redistributive policy, citizens must gain. However, without such a policy it is easy to see that households deriving most of their income from unskilled labor would lose while skilled labor owning households would gain.

\(^6\) We develop this framework in more detail in the next section.
and identifies $s_j$ (the technology in use in each sector). By the small country assumption, the relative commodity price $(p = P_2/P_1)$ is fixed, which fixes the unit-value isoquants, and thus fixes the common isocost, whose slope gives $\omega$. The cone defined by the rays $s_1$ and $s_2$ is called the cone of diversification because any endowment in the interior of the cone involves production of both final goods at the given price, with the equilibrium technology in use. Thus, two economies, sharing the same technology sets and facing the same final good prices, but endowed with different proportions of $S$ and $L$, will choose the same technologies (i.e. $s_1$ and $s_2$) and have the same $\omega$. This is the Lerner (1952)-Samuelson (1948) factor-price equalization theorem. If we focus on a single country, this is easily seen as a very simple comparative static representation of immigration, with $z'$ the initial endowment and $z''$ the endowment after an immigration shock consisting purely of unskilled labor. It is this version of the theorem that Leamer (1995) calls the factor-price insensitivity theorem. The mechanism that brings this factor-price insensitivity about is the subject of the Rybczynski (1955) theorem. That is, with two goods, if commodity prices (and technology) are unchanged the location of the unit value isoquants cannot change, the equilibrium isocost cannot change, which means that the $\omega$ ratio cannot change unless the economy specializes. Thus, the only way this economy can respond to a change in endowment, from $z'$ to $z''$ (an increase in $L$ with $S$ fixed), is to change output mix, increasing output in the sector using $L$ intensively (by proportionally more than the increase in $L$) and decreasing output in the other sector, as illustrated by the arrows. The essential point here is not that factor-price insensitivity actually obtains, but that, in a world with more than one output, some of the adjustment to an endowment shock will occur via a change in the output mix, reducing the actual, and measured, costs to the competing factor (i.e. domestic unskilled labor). In the

---

7 Our assumption of no factor-intensity reversals guarantees that $s_1 > s_2$ for all $\omega$.

8 Interestingly, Samuelson concludes his original article of FPE with a discussion of its implications for immigration policy—though the policy in question was that of encouraging emigration from England to Australia.
Heckscher-Ohlin-Samuelson (HOS) model illustrated here, as long as both goods are produced, the only way to generate a change in relative factor-prices is to change the relative commodity prices. As we shall see, in section III, increasing the dimensionality further leads to a variety of complexities, but the potential for adjustment via the output mix will remain, and so, generally, will the expectation that forcing all adjustment through the wage will produce overestimates of the long-run labor market effects of immigration.

One might expect that, and some discussions seem to suggest that, the fundamental difference between the labor theoretic and trade theoretic approaches to framing empirical research relates to the explicit incorporation of international trade flows. This, however, is not the case. As we have just seen, both the labor and trade theoretic approaches tend to hold the prices of final commodities exogenously fixed. As Altonji and Card (1991) point out, one way to motivate this in the one good case is to suppose the domestically produced good is consumed and exported in exchange for an international good which is consumed, but not produced locally. Furthermore, the standard labor theoretic approach is to adopt a small country assumption that fixes the relative price of the exportable and the importable goods. Trade economists are fond of the small, open economy model for the same reason: analysis of the supply side of the model can be abstracted from demand considerations.

When labor economists say that their model is a “closed economy” model, what they mean is that it is closed to immigration. That is, immigration will occur as a comparative static change in the endowment. While a substantial trade theoretic literature has treated

---

9 Altonji and Card, however, adopt a version of the large country assumption in their own framework.
10 It should be noted that a sizable literature in labor economics is explicitly concerned with formally and econometrically modeling the migration decision, on the whole this literature is not particularly concerned with aggregate equilibria. Borjas (1994) and Lalonde and Topel (1997) survey much of this literature. For a survey that covers literature on migration decision-making in fields well beyond economics, as well as those in economics, see Massey, et al. (1998).
factor flows endogenously, there is no shortage of comparative static analysis.\textsuperscript{11} Our illustration in figure 2 does precisely that, and one might reasonably argue that a small country, comparative static framework is the natural framework for empirical analysis on this question.\textsuperscript{12} In any event, endogeneity of factor flows certainly does not distinguish between the labor and trade theoretic approaches.

In this section we have argued that the fundamental difference between the way in which labor and trade economists have approached the issue of the labor market effects of immigration lies in the dimensionality of the theoretical frameworks in common use. Labor economists tend to use a one sector model that straightforwardly yields a convenient framework for econometric applications, while trade economists tend to use a multi- (usually 2) good framework that yields much less clear empirical application, but offers a very different perspective on the labor market effects of immigration. We now turn to a more systematic development of the labor theoretic framework and the results generated within that framework.

\textbf{II. How Labor Economists Have Evaluated the Effects of Immigration}\textsuperscript{13}

In this section of the paper we discuss some of the major findings about immigration and labor markets that have been uncovered in recent research by labor economists. Our primary focus here is on the contribution of immigration to the growing inequality experienced in many OECD countries during the 1980s, and the implications of that

\textsuperscript{11} Ruffin (1984) provides a very clear presentation of the trade theoretic literature on international factor mobility.

\textsuperscript{12} As we shall see in section III, once we depart from the 1 sector labor theoretic framework or the \(2 \times 2\) framework of the HOS model, trade and immigration may be related in a variety of ways which need to be considered in evaluating empirical results.

\textsuperscript{13} Other, more detailed presentations of the material presented in this section can be found in Altonji and Card (1991), Borjas (1999a), Chiswick, Chiswick and Karas (1989) and Johnson (1980, 1998). With reference to general labor economics, Bronfenbrenner (1971, chapter 6) presents the underlying theory in a relatively elementary way, while Hamermesh (1993) provides a state of the art overview of theoretical and empirical work.
experience for future policy. In this section we consider in some detail empirical research by labor economists on the link between immigration and labor market outcomes (primarily wages). Contemporary empirical research on the labor market effects of immigration has grown quite large since its development in the early 1980s. We will divide this research into 2 broad categories: production function based studies; and cross-sectional wage/unemployment. As we noted in the introduction, the most striking result from that research is how small the measurable effects are of what is a fairly sizable labor market shock.

As we note in the preceding section, whether or not the implementation takes a structural form, the structure that drives both the econometric specification and the intuition for interpreting that analysis is generally a one-sector, perfectly competitive model. The labor economists’ standard approach to wage inequality and income distribution is firmly rooted to an analysis of “SDI” or “supply, demand and institutions” (Freeman, 1993, pp. 44-49). To evaluate the labor market effects of immigration, identifying how the immigration of workers with differing skills affects the relative supply of labor can be viewed as necessary first step. In turn, the skill group characteristics of new immigrants are affected by the returns to skill as well as the distribution of earnings in both the source and host countries. Finally, labor market institutions are important because they affect the degree of wage inequality, the structure of wages and the labor market response to shocks. We place some of the latter considerations on the back burner for now and start by outlining a template competitive labor market model. In this section we sketch this framework in a bit more detail.

Before proceeding with this discussion we comment briefly on what may be the best known gross distinction used to characterize this literature: area studies versus factor content studies (Borjas, Freeman, and Katz, 1997). The problem is that the label is misleading. We
have already noted, in our discussion of figure 1, that virtually all labor theoretic frameworks apply a factor content based approach—i.e. it is change in relative supply that generates the change in labor market outcomes. The issue is actually about level of analysis. That is: how large must the geographic unit (i.e. area) be such that observations on supplies and prices of various classes of labor are independent? As we shall see, there are good reasons for believing that geographic units like standard metropolitan statistical areas (SMSAs) or states are linked in ways that are inconsistent with cross-sectional observations being independent draws from some distribution, but it is not at all clear that the statistically optimal level of analysis is the nation. There is considerable evidence that national borders have economic effects, but, by the same token, there is also considerable evidence that quite local labor markets take significant periods of time to fully adjust to macro shocks.14 On balance, it is not clear to us that there is a clear reason to prefer one level of analysis to another. Level of analysis is always an important research decision, but this does not strike us as an essential distinguishing aspect in this body of research.

In addition to the issue of the appropriate level of analysis, another essential research question is the manner in which the common theoretical framework structures the research. We make a broad distinction between structural (or “production function”) methods and regression based methods. We start with the former.

**B. Production Function Based Methods**

The most direct implementation of the framework considered in the previous section, and the first to be developed in the current wave of research on the labor market effects of immigration, involves selecting a specific functional form for the production function given

---

in equation (1), estimating that function on cross-sectional data, and testing hypotheses on the
degree of substitutability or complementarity between inputs. In addition, elasticities of
derived demand can then be used to carry out policy experiments. Recalling that our
aggregate production function is \( y = f(z) \), \( z = \{z_1, ..., z_m\} \), we seek to calculate the Hicksian
partial elasticities of complementarity between any two of the inputs \( i \) and \( k \) as:

\[
\varsigma_{ik} = \frac{f_i f_k}{f_i f_k}, \quad \forall i, k \in I,
\]

where we have used subscripts to denote partial derivatives. Following Hicks (1970; also see
Sato and Koizumi, 1973), \( i \) and \( k \) are called \( q \)-complements if \( \varsigma_{ik} > 0 \) and \( q \)-substitutes if \( \varsigma_{ik} < 0 \). Because it is easier to interpret the quantity elasticities of inverse input demand,

\[
\eta_{ik} = \frac{\partial \ln w_i}{\partial \ln z_k},
\]

these are usually calculated using the relationship:

\[
\eta_{ik} = \varsigma_{ik} \theta_k,
\]

where, again, \( \theta_k \) is the distributive share of input \( k \).

In carrying out work of this sort, investigators must select a functional form that does
not prejudice the conclusion from the start. In particular, we would like the data to determine
the values of the elasticities defined in (2) and (4). Thus, the commonly used Cobb-Douglas
and CES forms will be inappropriate for any input vector with more than two arguments. As
a result, investigators have generally used one or another of the flexible functional forms.
In addition to selecting a specific functional form, the other major choice in this body of research involves the definition of the input vector. Broadly speaking, there are two approaches here: one defines the input vector in terms of observable characteristics (e.g. gender, age, immigrant status, etc.); while the other seeks to identify production relevant characteristics (e.g. quantity of human capital).

In the first paper using this approach, Jean Baldwin Grossman (1982) used cross-sectional data for 1970 to estimate a translog function of native workers, first generation immigrants, second generation immigrants, and capital.\(^\text{18}\) She finds that both first and second generation immigrants substitute for native labor, but that second generation immigrants are much closer substitutes for natives, and that new immigrants are closer substitutes for second generation immigrants than for natives. In addition, Grossman finds that capital is complementary with each type of labor, but that this complementarity is strongest with first-generation immigrants and weakest with natives. Grossman’s analysis concludes with a policy simulation using the relationship in equation (4) to calculate own- and cross-elasticities to study the effect of a 10% increase in the number of legal immigrants in the

\(^\text{18}\) The translog function is:

\[ \ln y = \ln f(z) = \beta_0 + \sum_{i} \beta_i \ln z_i + \frac{1}{2} \sum_{k} \sum_{l} \beta_{ik} \ln z_i \ln z_k. \]

our assumptions on the technology yield restrictions: \( \beta_{ik} = \beta_{ki} \) (Young’s theorem); and \( \sum_{i} \beta_i = 1 \) and \( \sum_{i} \beta_{ik} = \sum_{k} \beta_{ik} = 0 \) (constant returns to scale). This yields a set of distributive share equations:

\[ \theta_i = \frac{\partial y}{\partial z_i} = \beta_i + \sum_{k} \beta_{ik} \ln z_k + v_i, \ \forall i \in I, \]

that can be estimated using Zellner’s seemingly unrelated regression technique, to take the correlation among the \( v_i \) into account, to generate values for the function’s parameters. From these, one can calculate the Hicksian elasticities of complementarity, equation (15), as:

\[ \zeta_{ik} = \frac{\left( \beta_{ik} + \theta_i^2 - \theta_k \right)}{\theta_i} \quad \text{and} \quad \zeta_{ii} = \frac{\left( \beta_{ii} + \theta_i \theta_i \right)}{\theta_i \theta_k}. \]
labor force on a short-run equilibrium in which native wages are fixed (and thus adjustment occurs on the employment margin) and a long-run in which all wages are flexible. In the short-run, native employment falls by 0.8%, second generation wages fall by 0.06%, first-generation wages fall by 2.2%, and the return to capital rises by 0.2%. In the long-run, wages are flexible, so all markets clear: native wages fall by 1%, second generation immigrant wages fall by 0.8%, first-generation immigrant wages fall by 2.3%, and the return to capital rises by 4.2%.\(^\text{19}\)

In an important series of papers, Borjas (1983, 1986a, 1986b, 1987) uses a number of data sets from the 1980s to study different disaggregations of labor in the context of a generalized Leontief production function.\(^\text{20}\) Depending on the particular breakdown of labor (e.g. by gender, race, and immigration status), immigrants tend to be complements to some native labor and substitutes to others, though in all cases these effects are small—except for the effects of immigrants on other immigrants of the same type, for whom the effects can be sizeable and negative. Given Borjas’ more recent position as a leading opponent of

\(^\text{19}\) In related studies, Bürgenmeier, Butare, and Favarger (1991) estimate a translog function of immigrant labor, native labor, and capital using Swiss time series data from 1950-1986, while Akbari and DeVoretz (1992) estimate a translog function on an industrial cross-section based on Canadian data for 1980. In addition to finding qualitatively similar results on the pattern of complementarity between factors, the Swiss study finds evidence of a positive relationship between immigration and capital accumulation. At the economy-wide level, the Canadian study finds no significant effect of immigrants—i.e. all Hicksian elasticities of complementarity between immigrants and natives are insignificantly different from zero. However, when the sample is restricted to labor intensive industries only, the Canadian study does find evidence of labor displacement as a result of immigration.

\(^\text{20}\) The generalized Leontief production function is defined as:

\[
y = f(z) = \sum_{i=1}^{n} \sum_{k=1}^{K} \gamma_{ik} (z_i z_k)^{\gamma_k},
\]

where, as with the translog function, Young’s theorem requires \(\gamma_{ik} = \gamma_{ki}\), while concavity requires \(\gamma_{ik} \geq 0\) for \(i \neq k\). As Borjas points out, the generalized Leontief production function leads to linear-in-parameters wage equations, rather than the linear share equations derived from the translog production function. Thus, Borjas estimates

\[
w_i = \sum_{k=1}^{K} \gamma_{ik} \left( \frac{z_i}{z_k} \right)^{\gamma_k},
\]

on individual level data, sorted by SMSA, usually with a variety of controls. Borjas estimates these labor demands using both OLS and a two-stage least squares procedure to control for endogeneity of labor supply, though the latter generally has little impact on the results.
immigration and search for large effects, it may be worthwhile to quote his own summary of this, and other, work circa 1990:

the methodological arsenal of modern econometrics cannot detect a single shred of evidence that immigrants have a sizable adverse impact on the earnings and employment opportunities of natives in the United States. (Borjas, 1990, pg. 81).

In particular, Borjas fairly consistently finds that, while immigrants may be substitutes for white native born men, and thus increased immigration may have had a small negative effect on their labor market outcomes, immigrants are found to be complements to black native born men who, thus, may have gained from increased immigration.

This approach is also used to examine the effects of legal Mexican immigration on labor market outcomes of Hispanic natives (King, Lowell, and Bean, 1986) and illegal Mexican immigration on a wide variety of labor groups (Bean, Lowell, and Taylor, 1988) with essentially the same results: the first study finds evidence of complementarity, suggesting that Mexican immigration may have a positive effect on the wages of native born Hispanics; and the second study finds effects of legal immigration like those in Borjas, and finds that illegal immigration may have a small negative effect on white, non-Hispanic workers, but essentially no effect on native Hispanic workers.

The production function approach receives its most sophisticated treatment to date in a series of papers by Michael Greenwood and Gary Hunt with a variety of colleagues. In Greenwood and Hunt (1995), the authors are interested in examining a variety of adjustment channels beyond change in wage. For input demands, they estimate a translog cost function on SMSA level data for 1970, and find immigrant labor to be a substitute for domestic labor.\(^{21}\) In addition, they estimate labor supply functions and aggregated output demand

\(^{21}\) The translog cost function is:

\[
\ln C(w, y) = \sum_{i=1}^{n} \beta_{0i} \ln w_i + \sum_{i=1}^{n} \sum_{k=1}^{n} \beta_{ik} \ln w_i \ln w_k + \ln y, 
\]
functions for the local markets. With these results they construct a large number of simulations permitting adjustment via flexibility in native labor supply (via both variable participation rates and internal migration) and changes in demand for final output, as well as adjustment along a given isoquant as in the previous studies. As with the previous studies, the wage, and now labor force participation, effects of immigration are uniformly small and, perhaps not surprisingly the magnitude of effects generally fall with the opening of additional channels of adjustment. The final output demand channel in particular seems to have a consistent effect of reducing the wage effects of immigration (or even making the effects on natives positive). These results can be seen to be closely related to our claim that, with multiple sectors the existence of adjustment at the output mix margin will generally lead to smaller effects.

By the mid- and late-1980s, researchers working in applied production analysis had begun to recognize that standard flexible functional forms (including both the translog and generalized Leontief forms) could fail to satisfy concavity, but that flexibility may be destroyed if concavity is imposed globally (Diewert and Wales, 1987). Greenwood, Hunt, and Kohli (1996) begin their analysis by pointing out that virtually all of the studies we have reviewed to this point present results indicating the presence of failures of concavity, in addition they estimate CES, translog, and generalized Leontief cost functions on a common data set to illustrate violations. As a result, they conduct their analysis using the symmetric normalized quadratic form, developed by Diewert and Wales (1987), that permits curvature conditions to be imposed globally without endangering flexibility.\(^2\) The authors calculate

\[ \beta_{ik} = \beta_{ki}, \]  
and the share equations are, from Shephard’s lemma,  
\[ \theta_i = \beta_{ik} + \sum_{k=1}^{I} \beta_{ik} \ln w_k + v_i, \, i \in I. \] 

The elements of the \( z \) vector are domestic labor, immigrant labor, and capital.  
\(^2\) The symmetric normalized quadratic functional form of the unit cost function is:
the Hicksian elasticities of complementarity and find that native labor and immigrants are $q$-substitutes, while all other input pairs are $q$-complements. Thus, an increase in immigrants would lower the wage of native workers, and raise the wage of non-recent immigrants and capital, but these effects are quite small. For example, a 10% increase in the supply of recent immigrants would reduce the wage of native-born labor by 0.96%. The effect of this change on other recent immigrants, however, is quite large.

The research that we have considered to this point focused on immigration status (as well as race or country of origin, immigrant cohort, and sometimes gender) as a key production-relevant fact. That is, natives and immigrants are treated as internally undifferentiated categories of labor. By contrast, both theory and a sizable body of econometric work focused on the native labor market suggest that factors such as education and labor market experience (usually proxied by age cohort) are significant predictors of labor market performance. Rivera-Batiz, Sechzer, and Gang (Rivera-Batiz and Sechzer, 1991; Gang and Rivera-Batiz, 1994), however, argue that there is no particular reason to believe that immigrant status, or race or gender, is directly production relevant. They prefer, instead, to assume that individuals with identical bundles of production relevant traits will receive the same wage. As a result their strategy involves estimating a translog production function of education, experience, and unskilled labor to derive the relevant Hicksian elasticities, and then using data on the skill composition of immigrants versus natives to derive distributional effects. Like Borjas and Bean et al., they use individual data sorted into local market areas to

$$c(w) = \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{K} \beta_{ik} w_i w_k + \sum_{i=1}^{n} k_i w_i,$$

where $\beta_{ik} = \beta_{ki}$, and $\sum_{i=1}^{n} \sum_{k=1}^{K} \beta_{ik} = 0$, $\beta_i \geq 0$ and $\sum_{i=1}^{n} \beta_i = 1$. Diewert and Wales (1987) provide a method for imposing global concavity and show that it does not undermine flexibility. The elements of the $z$ vector in Greenwood, Hunt, and Kohli (1996) are native labor, recent immigrant labor, non-recent immigrant labor, and capital.
estimate, like Grossman, a translog production function, and then use equation (2) to get the Hicksian elasticities of complementarity, and (4) to get the relevant factor demand elasticities. In the first stage they find, for both US and European data, that own supply elasticities are negative, as expected, and that the cross-elasticities imply that unskilled labor, education, and skill are all complements for one another (i.e. \( \zeta_{ik} > 0 \) for \( i \neq k \)). In addition, own elasticities are all estimated to be considerably larger than cross-elasticities. The authors then construct skill inventories of immigrant and native groups and use those, along with the estimated elasticities, to compute composite elasticities of complementarity that summarize this information. As with other work that we have reported, there are a variety of sign patterns, but “the impact of all the immigrant groups on all the native-born groups are small in absolute magnitude” (Rivera-Batiz and Sechzer, pg. 106). The largest effect is that of Mexican immigrants on Mexican-Americans, where an increase in Mexican immigration of 10% will result in slightly less than an 1% fall in wages of Mexican-Americans (with a similar effect on native black labor). Similarly small results are found for the European case in Gang and Rivera-Batiz.

Finally, Greenwood, Hunt, and Kohli (1997) mix the approaches of Grossman and Borjas with that of Rivera-Batiz by disaggregating native and immigrant labor into four skill categories each (based on earnings), as well as capital, and estimating a symmetric normalized quadratic cost function on a cross-section of SMSAs. The authors determine skill category by estimating earnings functions for native and foreign born workers and, based on the national distribution of expected earnings for native workers, native and

---

23 In a study of the impact of low-skilled migration from Mexico, Davies, Greenwood, Hunt, Kohli, and Tienda (1998) estimate a symmetric normalized quadratic production function in which the arguments are: low-skilled natives divided by gender and ethnicity (Mexican, non-Mexican); native high-skilled males and females (one category); foreign born, low-skilled Mexicans; foreign born, low-skilled non-Mexicans; and capital. As in the previous studies, the authors find that in both 1980 and 1990 immigrants has negative effects on the native born, but that these effects were small. The effects on other immigrants were found to be large. Furthermore, whatever might be the effects of trade and factor mobility within the US, the effects are larger in areas of high immigrant concentration.
immigrant workers were classified into four earnings categories that were then taken to identify skill categories. Not surprisingly, given the number of factors, there is quite a variety of $q$-substitutability and -complementarity, but unskilled immigrants appear to be strong $q$-substitutes for low- and medium-skilled native labor, and $q$-complements for unskilled native labor. The hypothesis that natives and immigrants are perfect substitutes, within skill categories, is easily rejected. Once again, however, the authors are unable to find any evidence that unskilled immigration leads to large changes in the income distribution or in employment opportunities, with the exception of the effect on other unskilled immigrants.

All of the papers we have considered to this point exploit variance across urban areas, assumed to be independent labor markets, to estimate the effect of immigration shocks to existing resident workers. Such variance may be unavailable or, even if available, inappropriate for the analysis. An example of the former would be Hong Kong, which has long relied on large flows of immigrant labor, but is physically small enough that any assumption of geographic segmentation would be doubtful. More generally, internal labor flows even in a large market might be sufficiently large as to undermine estimation based on intra-national, cross-sectional data. This latter has long been the basis of Borjas (1994, 199b, 2006) critique of what he calls “area studies”. While the empirical foundations of this critique are, at least, controversial (e.g. Card, 2001), it is sufficiently plausible that an alternative approach would be a valuable complement to the empirical comparison of local labor markets.

One alternative that has been applied in other branches of empirical labor economics is to exploit variance across similar, but imperfectly substitutable, cohorts of workers in the context of a well-specified production technology. That is, differential immigration shocks to various skill categories can be used to identify the wage effects of immigration. With a large number of factors, this requires a technology, like the constant elasticity of substitution (CES)
production function, that is relatively parsimonious in parameters that need estimating.\footnote{Standard references here are Freeman (1976, 1979), Katz and Murphy (1992) Welch (1979). Of particular relevance here is the work of Card and Lemieux (2001). The basic approach applied in this work is to assume a nested CES structure (Sato, 1967; Bowles, 1970; Chung, 1994, chapter 11). This involves a top-level aggregator that gives output at time $t$ as:

$$y_t = f(z_t) = \left[ \sum_{i \in S_i} \zeta_{it} z_{it}^\rho \right]^\frac{1}{\rho},$$

where the $\zeta_{it} > 0$ are technical efficiency parameters; $\rho$ is a parameter related to the elasticity of substitution between aggregate factors, $\sigma$, as $\rho = (1 - \sigma)/\sigma$ and $-\infty < \rho < 1$. However, unlike the standard CES function, the nested CES function nests some lower level aggregates:

$$z_{it} = \left[ \sum_{k} \lambda_{ik} z_{ikt}^\theta \right]^\frac{1}{\theta},$$

where $S_i$ is the set of types of input $i$, and $\theta$ is related to the constant elasticity of substitution between these types, $\delta$, as above. For example, in labor applications $i$ might denote an education category (e.g. high school dropouts) and the elements of $S_i$ might be labor market experience (proxied by age category). There is no need to stop at two levels. A general treatment of multiply nested CES functions, with particular reference to computational applications, is given in Perroni and Rutherford (1995).}

This parsimony does come at some cost. If we refer to substitution between factors within a given aggregate as “inter-class” substitution and that between factors in different aggregates at the same level as “intra-class” substitution (Chung, 1994): we find that the inter-class partial elasticity of substitution is constant and identical for all input pairs in different branches. The intra-class elasticities are, of course, not so constrained.

In the first application of this approach, Suen (2000) assumes various age cohorts are imperfect substitutes and creates aggregate labor as a CES aggregate of those cohorts,

$$L = \left( \sum_{j} M_j^\rho \right)^\frac{1}{\rho}$$

where the $M_j$ are the 10 age cohorts, but then constructs the age cohorts as CES aggregates of various native and immigrant groups in each cohort,

$$M_j = \left( \sum_{k} N_{jk}^\rho \right)^\frac{1}{\rho}$$

where the $N_{jk}$ are the 4 groups in each cohort $j$ (1 native and 3 immigrant).

Estimating the econometric model implied by this framework on 1991 data for Hong Kong and then calculating the Hicksian elasticities, Suen estimates statistically significant, but very small magnitudes. Consistent with other work, his results suggest that the largest negative effect would fall on existing recent immigrants. His simulation based on those results
indicates that a 40% increase in the number of recent immigrants would result in less than 1% fall in the wages of recent immigrants, with smaller effects on other groups.

Borjas (2003, 2006) and Borjas and Katz (2005) adopt a structure similar to that in Suen (2000), but where the latter aggregates on age cohort and immigrant source, Borjas aggregates on age cohort (“experience”) and education. That is, like Rivera-Batiz and his co-authors, but contrary to the results of Greenwood, Hunt and Kohli (1997) and Suen (2000), as well as the earlier work with highly aggregated native and immigrant factors, Borjas assumes that immigrants and natives with the same experience-education traits are perfect substitutes. Using this framework, and data for the period 1980-2000, Borjas is successful at identifying large wage effects of immigration: the average native worker experiences a fall in wage of 3.2%, while the wage of high-school dropouts falls by 8.9 percent.

A recent paper by Ottaviano and Peri (2006) follows Borjas in adopting the national market as the level of analysis and the nested CES specification of technology. However, these authors make two distinctive assumptions: first, they permit native and immigrant labor within the same experience-education cell to be imperfect substitutes; and, second, they model the adjustment of capital to the immigration shock explicitly. The warrant for the first should be clear from other work reviewed in this section; the careful attention to the issue of capital adjustment is a major contribution of this paper. Borjas considers two periods: a short run in which capital is completely fixed; and a long-run in which capital adjusts completely. Given that data are constructed in 10 year periods, the short-run assumption is particularly unsatisfactory.25 The basic idea is that, just as immigration occurs gradually over a given decade, capital also adjusts continuously over time to changes in its economic environment. With these two changes, Ottaviano and Peri produce results which differ substantially from

25 In fact, Ottaviano and Peri find that, at least with respect to capital adjustment, 5 years is sufficient time for most of the adjustment to occur.
those of Borjas (see table 1), but are more consistent with those found in the other research reported in this section.

-Table 1 about here-

Overall, econometric research which explicitly exploits production theoretic structure, tends to find strong substitutability between immigrants and other immigrants of the same vintage and national origin and, otherwise, widely varying patterns of complementarity and substitutability between immigrants and natives (as measured by the Hicksian elasticity of complementarity). More importantly, the elasticities between immigrant and native labor are consistently small, and are smaller yet when other channels of adjustment than the wage are explicitly permitted in the analysis.

C. The Regression Approach to Estimating the Wage Effects of Immigration

While the production-theoretic framework directly implements the theory that forms the basis for much of the labor-theoretic research on the labor market effects of immigration, its requirements are demanding. To be set against the advantage of directly estimating cross elasticities of substitution is the reliance on functional form assumptions to identify the parameters of interest. As mentioned in the previous section, structural estimation of this sort invariably needs to trade-off the requirements of functional form flexibility, or ease of estimation, and strict adherence to the restrictions implied by the theory. In addition, while human capital variables can relatively easily be accommodated in the production-theoretic framework, the incorporation of a wide range of standard control variables does not fit easily within this framework. Thus, as a result of the relative ease of application, greater similarity to existing techniques in labor econometrics and the desirability of including a richer set of

---

26 There is a parallel literature applying regression analysis to unemployment. We focus on the wage results primarily because of the close link to the theory. We simply note here that the primary conclusions of this section—i.e. small to no effects, except on migrants of similar origin and vintage, and the least skilled native workers—holds as well for unemployment.
controls, the majority of the research on the labor market effects of immigration has taken place within a regression framework. The latter approach generally involves reduced form regression analysis of the use of natural experiments to examine empirical regularities and determine “causal” relationships.

As above, we assume two types of labor, skilled \((S)\) and unskilled \((L)\), for simplicity. \(^{27}\) Within each skill class, domestic and immigrant labor is assumed to be equally productive. Other factors of production, such as physical capital, land and so on are left in the background by assuming a fixed price and applying a separability assumption. Following Altonji and Card (1991), we suppose that the economy produces one good according the production function \(Y = f(S, L)\). \(^{28}\) This good is consumed and is exchanged at fixed world prices for another consumption good. \(^{29}\) We will find it convenient to work with the total cost function:

\[
C\left(w_s, w_L; Y\right) = Yc\left(w_s, w_L\right),
\]

where \(c(\cdot)\) is the unit cost function.

We suppose that there are \(L\) unskilled workers and \(S\) skilled workers in the economy and define total labor force as \(\Lambda \equiv L + S\). We define the per capita labor supply functions of skilled and unskilled workers, respectively, as: \(l_s(w_s)\) and \(l_L(w_L)\). Because, by Shephard’s lemma, the unit labor demand functions are \(c_i(w) = \frac{\partial c(\mathbf{w})}{\partial w_i}\), we can write the conditions for a local labor market equilibrium as:

\[
SL_s(w_s) = Yc_s(w) \quad \text{and} \quad LL_L(w_L) = Yc_L(w).
\]

\(^{27}\) It should, however, be noted that this is not without loss of generality (Hamermesh, 1993, pp. 33-42).

\(^{28}\) As above, we take \(f(A)\) to be twice differentiable, linear homogeneous, and strictly quasi-concave.

\(^{29}\) As a result of the small country assumption, the foreign good will not enter our analysis explicitly, so we do not introduce any notation for it. The foreign good will be our numeraire.
To focus on the impact of unskilled immigration, we will represent migration as an influx of unskilled workers into the economy. Let $\lambda_L := L/\Lambda$, the proportion of unskilled workers in the labor force. Totally differentiating (6), suppressing arguments, and using obvious notation, we have:

$$
(1-\lambda_L) I_{SS} \; dw_s - I_S \; d\lambda_L = \frac{Y}{\Lambda} (c_{SS} \; dw_s + c_{S} \; dw_L)
$$

(7)

$$
\lambda_L I_{LL} \; dw_L + I_L \; d\lambda_L = \frac{Y}{\Lambda} (c_{LS} \; dw_s + c_{L} \; dw_L).
$$

Defining $\epsilon_i$ as the elasticity of the labor supply and $\eta_{ik}$ as the elasticity of labor demand for skill group $i$ with respect to wage $k$, the equations (7) may be rewritten as:

$$
(1-\lambda_L) \epsilon_S \frac{l_s}{w_s} \; dw_s - l_S \; d\lambda_L = \frac{Y}{\Lambda} \left( \eta_{SS} \frac{c_S}{w_s} \; dw_s + \eta_{S} \frac{c_S}{w_L} \; dw_L \right)
$$

(8)

$$
\lambda_L \epsilon_L \frac{l_L}{w_L} \; dw_L + l_L \; d\lambda_L = \frac{Y}{\Lambda} \left( \eta_{LS} \frac{c_L}{w_s} \; dw_s + \eta_{L} \frac{c_L}{w_L} \; dw_L \right).
$$

Letting hats (i.e. $\hat{x} = dx / x$) denote proportional changes, and using equation (6) to simplify, we have:

$$
(1-\lambda_L) \epsilon_S \hat{w}_{S} - d\lambda_L = \eta_{SS} \hat{w}_{S} + \eta_{S} \hat{w}_{L}
$$

$$
\lambda_L \epsilon_L \hat{w}_{L} + d\lambda_L = \eta_{LS} \hat{w}_{S} + \eta_{L} \hat{w}_{L}
$$

(9)

The labor demand elasticities are governed by the usual Hicksian formula, i.e., for given wage changes

$$
\eta_{ik} = \theta_k (\sigma_{ik} - \xi),
$$

(10)

where $\theta_k$ is the distributive share of labor type $k$, $\sigma_{ik}$ is the partial elasticity of substitution between labor types $i$ and $k$, and $\xi$ is the elasticity of final output demand with respect to its relative price $p$.30

---

30 Recall that we have taken the importable as our numeraire, so $p = P/P^*$, where the star denotes a foreign magnitude.
Comparative statics on equations (9) are straightforward. For instance, when the demand for skilled labor is independent of the wage paid to unskilled labor (i.e. $\eta_{SL} = 0$), we have

$$\hat{w}_S = \frac{1}{(1-\lambda_L)\epsilon_S - \eta_{SS}} d\lambda_L > 0$$  \hspace{1cm} (11)$$

Similarly, when the demand for unskilled labor is independent of the wage paid to skilled labor (i.e. $\eta_{LS} = 0$), we have

$$\hat{w}_L = -\frac{1}{\lambda_L \epsilon_L - \eta_{SS}} d\lambda_L < 0.$$  \hspace{1cm} (12)$$

Equations (11) and (12) are immediately instructive. As suggested in the extensive literature on the LeChatelier-Samuelson principle, the long-run impact on wages is likely to be far smaller than the short-run impacts. Similarly, it is easy to see that if factor-price equalization causes these elasticities to approach infinity, there will be no wage effects.\(^{31}\)

Equations (11) and (12) can be used directly to derive implications for the correlation between wages and shares of immigrants in a local labor market (see, e.g., Altonji and Card, 1991). An alternative approach to evaluating the empirical relationship between skilled and unskilled wages is to assume an explicit functional form for the technology and use equation (3) to develop an estimating equation. For illustrative purposes, suppose that $Y$ is produced from skilled and unskilled labor according the following CES function:

$$Y = A \left[ \tau_S z^p_S + \tau_L z^p_L \right]^{\rho},$$  \hspace{1cm} (13)$$

where $\sigma_{SL} = (1 - \rho)^{\frac{1}{\rho}} \geq 0$. The parameter $A$ indexes factor-neutral total factor productivity; and the $\tau_i$ index biased technical change that increases the “effective” quantity of the relevant input. The cost function associated with this production function is (Varian, 1992, pg. 56):

\(^{31}\) This conclusion is insensitive to the assumption that $\eta_{SL} = \eta_{LS} = 0$.  

22
\[ C(w_s, w_L; Y) = \frac{Y}{A} \left[ \frac{w_s}{\tau s} \right]^\rho \frac{\rho^{-1}}{\rho} + \left( \frac{w_L}{\tau L} \right)^\rho \frac{\rho^{-1}}{\rho} . \] (14)

Using Shephard’s lemma and taking the ratio of the unit labor demand functions, we obtain the relative demand for labor within a sector or location as:

\[ \frac{c_L(w_s, w_L)}{c_s(w_s, w_L)} = \left( \frac{\tau_s}{\tau_L} \right)^\rho \frac{(w_L)}{(w_s)} \frac{1}{\rho^{-1}} . \] (15)

Local labor market equilibrium requires that

\[ \frac{\lambda_L L(w_s, w_L)}{(1 - \lambda_L)L_s(w_s, w_L)} = \left( \frac{\tau_L}{\tau_s} \right)^\rho \frac{(w_L)}{(w_s)} \frac{1}{\rho^{-1}} , \] (16)

where the left hand side is the ratio of the supplies of unskilled to skilled labor. Solving for relative wages and taking natural logarithms gives a regression specification:

\[ \ln \left( \frac{w_L}{w_s} \right) = \rho \ln \left( \frac{\tau_L}{\tau_s} \right) + (\rho - 1) \ln \frac{\lambda_L L(w_s, w_L)}{(1 - \lambda_L)L_s(w_s, w_L)} . \] (17)

Equation (17) provides a general framework for understanding the determinants of relative wages. It should be kept in mind that the relative wage changes for different skill groups are for within local or regional labor markets. The link to outside markets comes from specifying how they respond to changing wages (i.e. the \( \epsilon_i \)). Equation (17) can then be used to examine the determinants of relative wages across different (and fully separate) labor markets. Borjas et al. (1997) constitutes a prominent example of this type of approach. What they term the ‘aggregate factor proportions approach’ involves regressing the ratio of skilled wages to unskilled wages in year \( t \), on the relative labor supply of the two types of labor. Borjas et al. (1997) find that immigration affected certain groups of workers more so than others. Specifically, immigration may have been responsible for the decline in the earnings of unskilled native workers that occurred during the 1980s. Their paper has contributed to
the view that, relative to the effects of growing international trade with less developed countries, immigration may have had a proportionately larger negative impact on the earnings of unskilled U.S. workers.

A qualitatively similar approach is used to derive estimating equations for regional unemployment or wages. For example, Altonji and Card (1991) and LaLonde and Topel (1991) estimate wage equations taking the form:

\[ w_j = \gamma_j \lambda_j + \beta_j X_j + \nu_j, \]  

(18)

where \( j \) indexes the local labor market, \( w \) is the logarithm of the wage for a particular skill group, \( X \) is a vector of control variables and, as above \( \lambda \) is the proportion of immigrants in the local labor market. In contrast to the above study by Borjas, et al. (1997), these studies find scant evidence that recent waves of immigration have disadvantaged U.S. workers.

To eliminate region-specific fixed effects, due to ethnic enclave effects, for example (see Bartel, 1989), the first-differenced version of equation (15) can also be estimated. More generally, if the immigrant share in market \( j \) is correlated with unobservable variables only through a time-invariant individual fixed effect, then estimating fixed effects regressions may be appropriate (e.g., see Altonji and Card, 1991; Topel, 1994a, 1994b). LaLonde and Topel (1991) estimate both equation (15) and its differenced version and find that the estimates of the effect of immigration produced by the two methods are nearly identical, i.e., the wage effects are negligible.

Unfortunately, fixed effects estimation is not a cure all for most sample selectivity and endogeneity problems. In the case of immigration and wages, the very nature of sorting on unobservable variables suggests that the migration decision of individuals may involve a process of learning about what is their correct state (i.e., industry, occupation, location, etc.). We discuss the endogeneity and sample selection further below in connection with the
instrumental variables and natural experimental approaches to the study of the impact of immigration.

Of course, the regression specifications based on either equation (14) or (15) are quite general. For instance, there have been many studies using the regression framework that have focused on the importance of the large increase in the relative supply of workers during the 1970s to the increasing wage inequality that occurred throughout the later 1980s and early 1990s. The increase in the U.S. workforce caused by the labor force entry of the baby boomers easily dwarfs the increase in the labor force caused by immigration. Welch (1979), Berger (1985), Murphy, Plant and Welch (1988) and Murphy and Welch (1991) are among the better-known U.S. studies here. A common finding of these studies is that changes in cohort size associated with the Baby Boom generation did not have a significant impact on cohort earnings. Overall, supply-side changes in the United States were very quickly discounted as a candidate explanation for the increased dispersion in the income distribution in the United States during the 1980s.

Notwithstanding, the preceding findings on the effects of domestic labor supply shocks do not necessarily imply that all supply-side “shocks” are unimportant. In the current context, some authors claim that immigration may have been responsible for the decline in the earnings of unskilled native workers that occurred during the 1980s. The immigration issue has been increasingly seen as one of “distribution” rather than “efficiency” (see LaLonde and Topel, 1997). Freeman (1998, p.110) argues that immigration may have had substantially larger effects on native unskilled workers than increased international trade with low-income countries, for instance. During the 1980s, a period during which wage inequality rapidly increased in the United States, immigration raised the supply of high school dropouts by approximately 25 percent, which far exceeds the increase in the “implicit labor supply” of such workers attributable to trade. Furthermore, Borjas et al. (1992, 1997) conclude that the
large increase in the number of unskilled immigrants explains about one third of the decline in the relative wage of high school dropouts during the 1980s. For the United States, wage inequality increased most in the West where the largest inflow of less-skilled immigrants was experienced (Topel, 1994a; 1994b).

In principle, changes in cohort quality can be analyzed in the same way as changes in cohort size. Borjas (1994) considers the declining cohort quality of recent waves of immigrants to the United States to have been the result of the shift in U.S. immigration policy, specifically the passage of the 1965 Immigration Act. However, his findings of decreasing cohort quality have recently been questioned by Butcher and DiNardo (2002) who focus on changes in the wage distribution through time. Using the methodology developed by DiNardo et al. (1996), they investigate the counterfactual of what the wage distribution would have looked like for new immigrants if they had faced the wage distributions from different eras. They find that earlier immigrants would have had wages much more similar to today’s new arrivals, if they had faced the present day prices for their skills.32 Race and ethnicity, and not the changing education levels of the new immigrants, explain much of the change in comparative economic fortunes of recent immigrants once wage structure changes have been held constant. The point, as also stressed by LaLonde and Topel (1991), is that recent cohorts of immigrants will look as if they do worse, even if they have the same set of characteristics as earlier cohorts of immigrants, if the distribution of wages has become more dispersed and if the new immigrants lie near the lower tail of the income distribution.

The use of regressions to uncover the wage effects of immigration by regressing immigrant shares and other controls on wages or relative wages poses many familiar problems. Among the more prominent concerns with multiple linear regression analysis is

---

32 With similar implications, albeit from a different perspective, Friedberg and Hunt (1995) note that “composition problems” make it difficult to ascertain the impact of immigration on wage inequality. For example, they argue that including the newly arrived waves of less-skilled migrants in inequality calculations is likely to bias the conclusion towards finding greater inequality in the United States.
the omission of important right-hand side variables. Biased estimates result if relevant characteristics or controls are not included in the regression equation. Similarly, how do various characteristics that are included in a model specification interact with one another? More generally, empirical work usually forces researchers to assume an appropriate functional form in order to reduce the problem at hand to one of estimating the parameters of interest. For example, would a linear function involve a serious mis-specification loss? As the previous section revealed, there is a wide range of functional forms from which to choose and so the robustness of parameter estimates is invariably an issue that needs to be confronted.

Variable (mis-)measurement and interpretation also pose problems. For instance, when does a migrant finally assimilate and become a native? The latter problem is particularly obvious one in those countries that are essentially composed of older generations of immigrants (e.g., Australia and the United States). More formally, there is the issue of weak separability (see Berndt and Christensen, 1974) of the various types of labor – not just of skilled versus unskilled labor, but also of native workers versus immigrant workers as well as first generation migrants versus second and later generations of migrants.

One of the most important difficulties in the empirical immigration and labor market effects literature is the likely possibility that labor supply functions are not independent of wages. The problem is reminiscent of the difficulties faced by the labor economists who attempted to uncover the effects of trade liberalisation on relative wages (see Gaston and Nelson, 2000b). Economic commonsense suggests that the immigrant labor force share is endogenous. To make the endogeneity issue transparent, consider a simple 2-equation model:

---

33 Zimmermann (1995) reminds readers of the literature that the European research on immigration has more to do with the effects of possessing citizenship. Unlike the U.S. literature, which has tended to focus on the effects of newly-arrived immigrants on native workers as well as on earlier generations of immigrants, the European data do not distinguish individuals as foreign-born or not.
\[ w_j = \gamma_\lambda X_j + \beta_{\lambda} X_j + \nu_j \]
\[ \lambda_j = \gamma_{\omega} w_j + \beta_{\omega} R_j + \nu_j, \]  
(19)

where \( X \) and \( R \) are (exogenous) scalars and all variables are expressed in deviations from their means. As before, \( j \) indexes a local labor market. The sign of the OLS bias is given by:

\[
\text{plim} \gamma_{\omega}^{\text{OLS}} - \gamma_{\lambda} = \left[ \frac{\sigma_{\lambda}^2}{\sigma_{\lambda}^2 - \sigma_{\lambda,\omega}^2} \right] \left[ \frac{\gamma_{\omega}\sigma_{\omega}^2 + \sigma_{\omega,\nu}^2}{1 - \gamma_{\omega}\gamma_{\lambda}} \right].
\]  
(20)

It is not possible to argue a priori that the sign of the bias is either positive or negative. For illustration, suppose \( \sigma_{X,\lambda}^2 = \sigma_{\omega,\nu}^2 = 0 \) and that \( \gamma_{\omega} > 0 \) (i.e. higher relative wages are associated with higher relative supply). If the “true” effect of a higher migrant share of unskilled workers is to depress unskilled wages, i.e. \( \gamma_{\lambda} < 0 \), then the bias is positive. That is, a failure to account for endogeneity will bias upward (i.e. toward zero) estimates of the impact of immigrants on wages.

However, note that if we are estimating some variant of equation (17) that, strictly speaking, our focus is on wage inequality. Furthermore, in many of the early studies in this literature, \( \lambda \) is simply taken to be the share of migrant labor in market \( j \). Under this interpretation, it is no longer obvious that \( \gamma_{\omega} > 0 \). Models of immigrant worker self-selection, based on the pioneering work of Roy (1951), are extremely illuminating here.

Workers with high earnings potential are likely to migrate from a country with an egalitarian wage structure (where they cannot easily make high earnings), while workers with low earnings potential are especially likely to migrate from a country with great wage inequality. In terms of source country characteristics, equality of the income distribution encourages what is termed “positive selection bias”.\(^{34}\) Negative selection bias results when source countries have unequal income distributions and therefore migrants are likely to be the

\(^{34}\) In fact, a point often overlooked is that host country labor market conditions are absolutely central to the migration decision. For example, Hanson and Spilimbergo (1999a) found that attempted illegal immigration from Mexico is extremely sensitive to changes in real wages in Mexico.
least skilled. Recent waves to the United States tend to have been increasingly drawn from the latter group (Borjas, 1994). Immigrants are mobile, but they have tended to cluster in cities where their fellow countrymen reside. The clustering effects tend to dominate such economic incentives as differences in unemployment rates or welfare benefits across areas (Bartel, 1989; Bartel and Koch, 1991). The effects of clustering are borne by the gateway cities, while the geographic concentration tends to reduce economic progress and the rate of assimilation. Of importance for the present discussion is that, given that the primary adverse wage impact of new immigrants is upon previous generations of migrants, the clustering effect may imply $\gamma_w < 0$. If the effect of clustering is sufficiently strong, then it is possible that OLS estimates are biased downwards, and not upwards. Clustering may be occupational as well as geographic. Friedberg’s (2001) findings are consistent with this line of argument. She studies the impact of Russian migration on occupational wages in Israel and finds that IV estimates are higher than OLS estimates. That is, rather than immigrants choosing occupations based on them offering higher wages, she finds evidence of occupational immobility (so that $\gamma_w < 0$). That is, immigrants, irrespective of their skill levels are confined, initially at least, to low-paying occupations. Hence, OLS estimates overstate the impact of immigrants on wages.

Handling the endogeneity problem is the motivation for the use of the instrumental variables (IV) approach (e.g., Altonji and Card, 1991 and Friedberg, 2001) and the quasi-experimental approach in the labor literature (e.g., Card, 1990; Hunt, 1992). Altonji and Card

---

35 Interestingly, increasingly negative self selection produces labor market outcomes in both the source and host countries similar to the picture of the effects of outsourcing on wage inequality painted by Feenstra and Hanson (1996a, 1996b). That is, if workers emigrating from Mexico are relatively high skilled from Mexico’s viewpoint and unskilled from the United States’ viewpoint, then wage inequality tends to rise in both countries. 

36 Though see Chiquiar and Hanson (2005) for a somewhat more nuanced analysis based on a non-linear immigration cost function yielding empirical results that are inconsistent with the negative selection hypothesis.

37 Friedberg and Hunt (1995) make a related criticism of Goldin’s (1994) findings. Using data for 1890 to 1923, Goldin found a significant negative correlation between the percentage of foreign-born residents and wages in U.S. cities. However, this may be a ‘composition’ effect, i.e., if immigrants earn lower wages than natives, then even if immigrants have no effect on native wages, they tend to be clustered into cities with lower average wages.
(1991) investigate the impact of immigrants on low-skilled native workers. They relate changes in the earnings and employment of low-skilled natives across cities to changes in the migrant population. As discussed, the problem is that the immigrant flows are likely to be correlated with current labor market conditions. Hence, Altonji and Card instrument the change in immigrants with the size of the immigrant enclave in an earlier period. They argue that the size of the immigrant enclave in the past is likely to affect immigrant flows but is not necessarily correlated with current demand shocks. In other words, the IV approach attempts to use only the variation in immigrant flows associated with variation in enclave “pull” and not that associated with current demand shocks. Interestingly, Altonji and Card’s estimate of γₖ is one of the most negative. Notwithstanding, they conclude that immigrants and natives face little competition from one another. They find that there is some industry displacement from low-wage immigrant intensive industries; but still, the implied elasticities are small.38 Despite these mobility effects, the effects on employment and unemployment rates are virtually zero.

Recent papers by Card (2001, 2005; Card and Lewis, 2006) seek to explicitly incorporate internal migration by natives in the local labor market framework. In order to account for immigrant heterogeneity Card stratifies the national labor market along occupational lines. Using the individual-specific occupational probabilities, which are estimated from a multinomial logit model for six occupational categories, he constructs a measure for the number of people that would be expected to work in a certain occupation and region in the absence of any distortions due to local demand and supply conditions. This method allows one to bypass classification problems that are likely to arise when assigning immigrants and natives to specific skill categories. The regression analysis then proceeds in

38 Friedberg and Hunt (1995) note that Altonji and Card’s “large and negative” estimates imply that a 10 percent increase in the percentage of foreign-born in a local labor market implies a minuscule 0.86 percent reduction in wages.
two steps. In a first step, Card estimates the impact of immigration inflows during the period 1985-1990 on population shares in 1990. We will discuss this in more detail in the next section. In the second step, he regresses population shares on labor market outcomes. OLS estimates suggest that a 10% rise in the population share of an occupational group depresses employment and wages in that group by about 0.5%. Instrumental variable estimation is used to net out the exogenous supply-push component of immigration. The IV estimates suggest that the impact of immigration is three to four times larger in the case of employment, but still not very large. He does not find any robust effects for wages. Card concludes that even after controlling for the role of local demand conditions as well as the possibility that immigrants displace natives, the results suggest that the impact of immigration on natives is relatively small, and limited to employment.

Dustmann, Fabbri and Preston (2005) apply the local labor market methodology to the case of the UK. Rather than adopting the two-step procedure as in Card (2001) they directly control for the size of the native population in their econometric model in order to control for changes in the labor supply due to movements of natives between regions. Furthermore, the use of panel data allows them to control for time-invariant differences in demand across regions and nation-wide differences in demand over time by including region and time fixed effects. The use of panel data further allows them to employ GMM in order to control for the potential correlation between immigration inflows and unobservables. They find a positive effect of immigration inflows on wages using OLS, but the sign becomes negative when they turn to GMM. With GMM, a 10% increase in immigration leads to a decrease of 0.7% in native employment. For unemployment, labor force participation and wages the estimated

---

39 As well as measurement problems due to small sample size of the Labour Force Survey.
effects of a 10% increase in immigration inflows are 0.07%, -0.04, and 0.91% respectively. However, none of these estimates are statistically significant.

Due to the substantial difficulties associated with choosing “good” instruments (e.g., see Nelson and Startz, 1990; Bound, Jaeger and Baker, 1995), considerably more weight in this branch of the literature has been attached to the results of the quasi- or natural experiments. Natural experiments occur when exogenous variation in independent (explanatory) variables (that determine “treatment assignment”) is created by abrupt or exogenous shocks to labor markets (Meyer 1995). For example, natural experiments can arise due to institutional peculiarities (e.g., Vietnam-era draft lotteries) or due to exogenous policy changes that affect some groups but not other groups (e.g., changes in policies in some states but not others).\(^\text{40}\) In the latter case, Hanson and Spilimbergo (2001) examine how enforcement of the U.S.-Mexico border is affected by changes in illegal immigration. They find that the equilibrium level of border enforcement varies inversely with relative demand shocks (and consequently, demand for undocumented labor). In other words, the authorities relax border enforcement when the demand for undocumented workers is high.

Natural experiments are most useful in situations in which econometric estimates are ordinarily biased because of endogenous variables due to omitted variables or to sample selection. The basic approach involves a comparison of changes for “treatment” and “control” groups (i.e., differences-in-differences). This can be accomplished in a components of variance scheme (time effects, location effects, treatment group effects, interaction terms and so on) or by using an IV approach in which one instruments for the treatment dummy variable with the natural experiment indicator variables. In this sense, the IV and natural experimental approaches are qualitatively equivalent. With IV, legitimate instruments

\(^{40}\) Hamermesh (2000) argues that, unlike “acts of God”, treating changes in the legal environment as exogenous is rarely convincing.
generate a natural experiment that assigns treatment in a manner independent of the unobserved covariates. The advantage is that the source of the identifying information is transparent.

To illustrate the basic approach, consider the following example based on Meyer (1995). Suppose that we have:

\[ w_{js} = \alpha + \beta D_s + \varepsilon_{js}, \]  

(21)

where \( w_{js} \) is the wage in local labor market \( j \) (e.g. city). The “treatment” dummy is \( D_s \), which may be thought of as measuring migration intensity above some critical threshold in market \( j \). That is, if \( \lambda_j \geq \lambda^* \) then \( D_s = 1 \), and 0 otherwise. Also, an important assumption is that \( E(\varepsilon_{js} \mid D_s = 0) = E(\varepsilon_{js} \mid D_s = 1) \) or \( E(\varepsilon_{js} \mid D_s) = 0 \). Equation (21) simply involves comparing wages in cities that have high migrant populations to those that do not. An unbiased estimate of \( \beta \) is obtained by differencing, i.e.,

\[ b = (\bar{w}_1 - \bar{w}_0), \]  

(22)

where overscores denote group means. Under typical assumptions, \( \alpha \) would be consistently estimated as the group size goes to infinity. Of course, we could also simply estimate \( \alpha \) from the regression equation (21). Note that the researcher needs strong evidence that the two types of labor markets would have been comparable, but for the arrival of foreign-born residents.

Occasionally, data are available for the time period before and after a “treatment” (in our case, the treatment is an immigration shock) for a group that does not receive the treatment but experiences some or all of the other influences that affect the treatment group. At the very heart of the quasi-experimental approach to the immigration and labor market literature are the non-policy and non-institutional shocks that can be considered truly exogenous to existing labor
market conditions in the destination country (e.g., Baby Boom, Black Death, Mariel Boatlift). That is, consider:

\[
    w_{jst} = \alpha + \beta D_t + \tau D_s + \gamma D_{st} + \epsilon_{jst},
\]

(23)

where \(D_t\) can be thought of as a time period dummy, \(D_s\) is defined as above, and \(D_{st} = 1\) if \(D_t = D_s = 1\), and 0 otherwise. The key idea is that \(\tau\) summarises the way in which both treatment and non-treatment groups are influenced by time (e.g. such things as macroeconomic conditions and regional growth trends). The time-invariant difference in overall means between the groups is captured by \(\beta\). \(D_{st}\) indicates membership of the experimental group after it receives the treatment and \(\gamma\) is the true causal effect of the treatment on the outcome for this group. Again, the key identifying assumption is that \(E(\epsilon_{jst} \mid D_{st}) = 0\).

Note that \(\gamma\) would be 0 in the absence of the treatment (i.e. the immigration shock). An unbiased estimate of \(\gamma\) can be obtained by the differences-in-differences estimator, i.e.:

\[
    g = (\bar{w}_{11} - \bar{w}_{01}) - (\bar{w}_{10} - \bar{w}_{00}),
\]

(24)

where the first subscript is \(t\) and the second is for treatment \(s\). Without question, the most cited natural experimental paper is Card (1990) which examines the impact of the Mariel Boatlift on Miami’s labor market. In his paper, the first bracketed term in equation (27) represents the difference in wages for black workers in Miami before and after the Boatlift.\(^{41}\) The second bracketed term is the wage difference for the same types of workers in a group of four comparison cities. The latter cites were chosen because they had relatively large populations of black and Hispanic workers and because they exhibited patterns of economic growth similar to those observed in Miami over the late 1970s and early 1980s. As is well known, despite the dramatic and sudden 7 percent increase in the size of Miami’s work force,

\(^{41}\) Card conducts a similar analysis for Hispanic workers, as well. Also, in addition to wages he uses the same methodology to examine whether the Boatlift had any effect on the unemployment rates of less-skilled workers.
Card is unable to detect any adverse impact on the wages or unemployment of less-skilled workers.

Kugler and Yuksel (2006) exploit the natural experiment provided by Hurricane Mitch, which displaced 1.5 million people in Central America in October 1998. It has been widely acknowledged that a large proportion of the displaced responded by moving North in an attempt to start a new life in the United States. Kugler and Yuksel exploit the exogenous variation provided by Hurricane mix by looking at the inflows immigrants from Central America across US states. In order to purge any remaining demand-driven immigration they use the distance from Honduras as an instrument. They find that unskilled migration tends to raise the wage of skilled natives without finding much of an effect for unskilled natives. They explain their findings by pointing out that unskilled immigrants are likely to act as complements to skilled natives (Borjas, 1995).

There are two notable quasi-experimental studies for Europe. Hunt (1992) examines the impact on wage differentials in France in 1968 of the influx of pied noirs from Algeria during the early 1960s; and Carrington and de Lima (1995) study the return of Portuguese colonialists from Africa and examine the wage effects across the provinces of Portugal. Consistent with Card’s findings, these authors were unable to discern adverse wage effects for native workers.

Friedberg’s (2001) study of the Israeli case is an interesting variant on the natural experiment. Israel is an interesting case for a number of reasons: first, the labor market shock was huge (something like 4% increase in population in 1990 alone and 12% from 1990-1994); second, the immigration shock was unusually skilled (relative both to other flows that have been studied and relative to the Israeli labor market); third, of particular importance to constructing a natural experiment, the shock was unexpected; and finally, Israel is a sufficiently small country that it must be considered a single labor market. Like Suen (2000),
Friedberg uses national data and identifies the effect of immigration from variance in immigrant share across occupation cells. However, because immigrants may sort into high wage occupations, or be constrained into low wage occupations, there is a potential endogeneity problem. To deal with this problem, Friedberg uses occupation in the Soviet Union as an instrument for occupation in Israel. Using OLS, Friedberg finds a significant negative effect of immigration on wages, but no effect on unemployment, of native. However, when the IV approach is used, the first result is reversed and there is no evidence of negative effects on wages or employment. That is, the negative effects in the OLS estimates are due to immigrants being forced into occupations with low wages and low wage growth.

In addition to the structural estimates we discussed above, Borjas (2003) also carries out an analysis like Friedberg’s OLS approach. Like Friedberg, Borjas finds a significant negative effect. Unlike Friedberg, Borjas does not proceed to an analysis of potential endogeneity problems of the sort Friedberg found to be significant in the Israel case. Recent papers by Bonin (2005) and Carrasco, Jimeno and Ortega (2006) also apply the skill-correlation approach, to West Germany and Spain respectively, generally finding small negative effects.

The broad conclusion from the first large NBER project on immigration and trade was that immigration had a relatively smaller area impact than increased import penetration on native labor. Overall, the labor market was thought to easily adjust to migrant inflows, absorbing immigrants with little redistributive losses to natives (see Abowd and Freeman, 1991). This conclusion was largely, and somewhat surprisingly, reversed by the second NBER project (Borjas and Freeman, 1992). While the wage and employment effects for natives in local labor markets are small, it was argued that certain groups of workers have been adversely affected by immigration. The augmented factor supplies of less-skilled
workers, due to either the effect of trade with low-income countries or from the immigration of workers from developing countries, were thought to have contributed to the poor outcomes of less-educated American workers during the 1980s and early 1990s.

The finding that certain groups of workers may have been adversely affected by immigration is evident for some European studies as well. For example, De New and Zimmermann (1994) find that greater concentrations of foreign workers in German industries during the 1980s were associated with small wage gains for white-collar workers, but relatively large wage losses for blue-collar workers. Zimmermann (1995) attributes these findings to the greater labor market inflexibility, greater levels of unionisation and low labor mobility in Europe in comparison to the United States. In the case of strong unions or wage inflexibility, the expectation is that immigration is associated with increases in native unemployment. In the case of labor immobility, equations (11) and (12) suggest that skilled wages increase and unskilled wages decrease when unskilled immigration increases.

It should, however, be noted that the results for Europe are quite mixed. For instance, Pischke and Velling (1994) find that immigration had no adverse wage or unemployment effects in German local labor markets. Similarly, Winter-Ebmer and Zweimüller (1996) using both OLS and IV estimation procedures find no detrimental immigration impact upon Austrian industry or regional wages.

---

42 Specifically, De New and Zimmermann find that their IV estimates were substantially more negative than their OLS estimates (in fact, 15 times larger). On one hand, this result may be seen as being consistent with Friedberg’s (2001) occupational crowding finding for Israel, discussed above. On the other hand, at a more practical level there is evidence of some instability in the coefficients of the industry level variables in the IV model specification. (De New and Zimmermann use industry dummies, industry growth rates and industry specific time trends as determinants of share of foreign workers by industry.) As the authors acknowledge, the issue of whether their instrumenting procedure has been able to fully control for the endogeneity of the foreign share of labor may have been insufficient.

43 Interestingly, Zimmermann (1995) notes that there has been little impact of immigration on unemployment rates. The research on the effects of immigration on Australian labor market outcomes has instead focussed on the likelihood of adverse unemployment effects. With a heavily regulated labor market (compared to the United States, at least), the concern has been that labor market adjustments would occur through quantity (length of the dole queue) rather than through prices (wages). However, in surveying the literature, Junankar et al. (1998) conclude that immigration has not increased the Australian unemployment rate.
Overview: For the OECD taken as a whole, immigration flows increased during the 1980s and at the beginning of the 1990s. Migrants have become more “visible” in virtually every OECD economy and immigrant labor has been spreading across an increasing number of sectors. (OECD, 1998, 1999). Of course, this time period is associated with growing wage inequality in OECD countries. During most of the 1990s, there was actually a reduction in legal immigration flows; however, illegal migration steadily increased. The difficulty in controlling the illegal flows has elevated immigration to an important place on the political stage in many developed countries. The “problem” is not just felt in the southern border states of the United States. More than 9 million former Soviet citizens migrated following the collapse of communism. After the fall of the Berlin Wall, Germany absorbed 4 million immigrants, of whom 2 million were ethnic Germans from the former Soviet Union (ILO, 2000). Since the European Union’s Eastern enlargement there have been sizable flows of migrants from Eastern Europe. Given the size Poland, it is not surprising that Polish migration has come in for particular comment.

More skilled workers often find legal entry into a new country easier than their unskilled workers do. However, the most heated debate has undoubtedly concerned the immigration of unskilled labor, particularly from developing countries, and the impact on OECD labor market performance.44 Temporary workers have become increasingly important in Western European countries. For example, Germany recently granted 1,000 new temporary visas for computer programmers and “scarce” IT skills. Such workers are thought to enhance labor market flexibility (particularly, when immigration restrictions are in place) and to help meet temporary sectoral shortfalls of particular types of labor. As with the large numbers of “temporary” Turkish workers who sought to remain in Germany, there are some

44 This not to say that the debate surrounding skilled immigrants is less heated. The computer programmers and doctors who have migrated to the United States from India constitute a brain drain for India. Other less direct issues are important too. For example, the union movement in Australia has voiced a concern that higher intakes of skilled immigrants weaken the incentives for domestic firms to train Australian workers.
potentially thorny political issues that may need to be confronted in the future (see Zimmermann, 1995).

There is a persistent gap in the unemployment rates of foreign and native workers in national labor markets. This is particularly the case for the European OECD countries, rather than the traditional “settlement countries” of Canada, Australia and the United States. In many cases, foreign male laborers are more heavily concentrated in the declining sectors of the economy (mainly import-competing manufacturing industries). The general view that immigrants assimilate relatively quickly is unlikely to be correct for a number of destination countries.

The United States is the primary destination for refugees of all types – political and economic. It is also the country on which most research on the impact of immigration on the labor market has been done. The migration of less-skilled workers to particular “gateway cities” (e.g., Los Angeles, Miami and New York) has attracted the most recent attention. In this section, we focussed on two of the more important “facts” that have seemingly gained widespread acceptance. First, that immigrant flows have small local labor market effects; and second, that immigration has affected certain groups of workers more so than others.

To us, the first conclusion seems inescapable. The same cannot be said for the second. In the case of the United States, such a conclusion seems an overly confident one to reach. Given the sheer size of the U.S. labor market and the quantity of unskilled labor, more broadly defined, it is unlikely that immigration (or trade) would have contributed to the overall increase in wage inequality observed in the United States during that particular period. On the other hand, as Rodrik (1998) notes, there may have been a fundamental change in the underlying demand for unskilled labor that is attributable to the increased

---

45 Similar clustering occurs in Australia as well. Nearly 40 percent of all new immigrants to Australia choose to settle in Sydney. The most visible opponent to the recent calls to liberalise Australia’s immigration policy have been Sydney-based politicians, who claim that the city is ‘bursting at the seams’ and that the city and state social services are inadequate.
availability of unskilled, migrant labor. As argued by Gaston and Nelson (2000b), it may be the case that trade and immigration engender institutional responses that do leave some types of unskilled labor more vulnerable to economic shocks than others. We explore this theme below.

III. How Trade Economists Think About the Effects of Immigration

We have already noted, in section I, that the essential difference between the labor theoretic model and the trade theoretic model, as frameworks for evaluating results like those reported in section II, lies in dimensionality. The one final-output sector model preferred by labor economists for the strong identifying restrictions that it generates for empirical work drives all adjustment to increased supply of factors through factor-price, while the multiple final-output sector model preferred by trade economists embodies output mix adjustment, as well as factor-price adjustment. In the extreme case of the HOS model, all adjustment occurs on the output-mix margin. The essential purpose of the HOS model, for our purposes here, is precisely that it throws the output-mix adjustment mechanism into high relief. In particular, it helps us understand why assertions to the effects that “The laws of supply and demand have unambiguous implications for how immigration should affect labor market conditions in the short run” (Borjas, 2003) are fundamentally wrong.47

46 Useful general surveys of the relationship between trade and immigration can be found in Ethier (1986, 1996), Wong (1995), Razin and Sadka (1997), and Venables (1999). In our focus on labor market effects in OECD countries we will be ignoring two important bodies of research that focus on immigration and trade in developing countries: the literatures focusing on emigration in the context of homogeneous labor (e.g. Kenen, 1971; Rivera-Batiz, 1982a; Djajic, 1986) and heterogeneous labor (e.g. Bhagwati and Rodriguez, 1975; Bhagwati and Wilson, 1989; Miyagiwa, 1991; Wong, 1997).

47 The reference to “the short run” at the end of this quotation might be seen as saving the claim. It is, of course, true that if quantity cannot adjust, price must bear all of the adjustment. However, the labor economists’ underlying model certainly assumes a sufficient period of time in which the aggregate economy, represented by the GNP function that generates the demand for labor curve, can absorb the immigration in an efficient way. More specifically, Borjas’ analysis rests firmly on the notion that the period is sufficiently long that native workers can adjust geographically to the immigration shock.
In this section we discuss the way the relationship between immigration and factor-market outcomes of natives and immigrants are affected by the structure of the model used to evaluate that change. We will consider: undistorted neoclassical models under various assumptions about dimensionality (a.k.a. dimensional generalizations of the HOS model); specific assumptions about product-market microstructure (non-traded goods and intermediate goods); and increasing returns to scale/geography models. However, before turning to these various interpretive frameworks, we briefly review the small body of empirical work that has sought to explicitly incorporate international trade in goods along with international migration.

A. The Empirical Link between Trade and Immigration, and Its Implications

The simplest approach to examining the effects of trade and immigration takes an agnostic position on the nature of the relationship between trade and immigration, and simply includes variables measuring both in a wage equation.\textsuperscript{48} Freeman and Katz (1991) estimate regressions of both hourly wage and annual hours on measures of change domestic demand, foreign demand, imports, and immigration (both stock and change), as well as a number of controls, on a cross-industry data set.\textsuperscript{49} Changes in imports and immigration are negatively related to hourly wages and positively related to annual hours. However, the authors suggest that these regressions generate suspiciously large effects of immigration, leading to an argument that they are picking up the tendency of immigrants to move into low- and

\textsuperscript{48} Borjas, Freeman and Katz (1992, 1997) simulate a partial equilibrium labor market model in which an inelastic labor supply is shifted by a direct immigration shock and an indirect labor import shock calculated via the factor contents of commodity trade. Even in this framework, which is adopted to maximize the labor market effects of globalisation, the authors conclude (BFK, 1997, pg. 66): “The bottom line from our simulations is that the economic impact of immigration is mainly redistributional and primarily affects a small group of the least educated U.S. native workers”.

\textsuperscript{49} These changes are calculated for 1958-1984. As a control, the authors also estimate these models on CPS data, with essentially the same result.
This explanation is consistent with the standard trade theoretic model, due to Mundell (1957), in which trade and factor mobility are substitutes. That is, sectors facing increasing competition from low wage (unskilled intensive) countries can slow the rate at which they decline by importing low wage labor directly.

Similar methodologies have been applied in the cases of Germany and Austria. For the German case, Haisken-DeNew and Zimmermann (1999) use the German Socioeconomic Panel data (SOEP) data to estimate wage regressions on a variety of individual variables and region/sector specific trade deficit and foreigner share variables, in a random effects panel model for 1984-1992. In addition to carrying out the analysis on the sample of all workers, they also segment the sample by skill (under both job title and years experience definitions), by blue v. white collar. In all cases, they find that trade is negatively related to wage, and immigration (in all cases but one) positively related to wage. The first finding parallels that of Freeman and Katz, while the second is directly contradictory. Because the immigration results are generally larger, and more precisely estimated for high-skilled workers, the authors conclude that this is suggestive of complementarity between immigrants and high skilled workers. Consistent with Freeman and Katz’ suggestion of a substitutive relationship between trade and immigration, however, is Haisken-DeNew and Zimmermann’s finding that import-competing sectors employ a larger share of immigrant workers. Winter-Ebmer and Zweimuller (1999a) examine trade and immigration in a cross-section of Austrian workers, finding that immigration increases unemployment duration by a small amount, but has no statistically significant effect of probability of unemployment. In addition, they find no effect

---

50 This tendency is observed directly in a wide variety of research.
51 By the logic of the Rybczynski theorem, as illustrated in figure 2, the import of unskilled labor results in an increase in the output of the unskilled labor-intensive sector and a fall in the output of the skilled labor-intensive sector. Even if the relative endowment of skilled labor is rising as a result of domestic human capital accumulation, possibly driven by increased international competition, an increasingly unskilled labor-abundant immigration will slow down the rate of decline of the unskilled labor-intensive sector.
52 The one exception is a statistically significant negative relationship between number of immigrants in a region/industry and the wages of low-skilled, white collar workers, where skill is defined by level of experience.
of trade on probability of unemployment or unemployment duration. In a related study of young workers, Winter-Ebmer and Zweimuller (1999b) find exports negatively related to unemployment (though exports to the CEEC are positively related), imports having no significant effect (though those from the CEEC have a negative effect), regional stock of immigrants makes unemployment more likely, but immigrants in the sector make it less likely. Again, these effects are generally small. Finally, Winter-Ebmer and Zimmermann (1999) present results, for both Austria and Germany, for changes in overall employment growth, native employment growth, and wage growth, as a function of changes in exports (to CEEC and rest of world), imports, and foreign share. In the Austrian case, immigration has essentially no effect on overall employment growth, and only small negative effects on native employment growth and wage growth. Imports also generally have a negative relationship to employment growth, with imports from the CEEC having a generally larger negative effect. For the German case there is evidence that overall immigration has a small negative effect on native employment growth and a small positive effect on wage growth, while immigration from Eastern Europe has a rather strong effect on native employment growth and a sizable positive effect on wage growth. The effects of growth in imports and exports are uniformly small, mostly insignificant, and perversely signed. Overall, these results are consistent with results reported above that immigration effects are small, even taking into account interactions with international trade.

The results reported to this point take an essentially ad hoc approach to evaluating the effects of globalization on labor market outcomes. An alternative approach extends the production theoretic approach to the simultaneous existence of trade in goods and factors. Following original work by Burgess (1974), empirical trade economists have exploited duality theory to estimate comparative static effects of trade by treating trade as a direct
argument in a GNP function. The marriage of this approach to trade modeling to the production theoretic modeling of immigration seems obvious, but has only rarely been done. Wong (1988) works with an indirect trade utility function which is, itself, a function of the GNP function. This function is estimated, in translog form, on prices for home produced durable goods, home produced nondurable goods and services, and imported goods and services, and endowments of capital, land, and labor, for a number of years between 1948 and 1983. Foreign capital and labor are taken to be perfect substitutes for the domestic factors, so the comparative statics on the indirect utility function can be used to generate the relevant elasticities. These elasticities are all small. At least as interesting, Wong finds that trade and immigration are complements, unlike the results we have discussed to this point which suggest that they are substitutes. Kohli (1993, 1999) develops this sort of analysis in considerably greater detail. Specifically, using annual Swiss data from 1950-1986, Kohli (1999) estimates the translog cost function associated with the primal GNP function and a z vector containing capital, home labor, immigrant labor, and imports. Thus, where Wong treats home and immigrant labor as perfect substitutes, Kohli is able to test this relationship. In fact, Kohli finds that home and immigrant labor are both Allen-Uzawa and Hicks q-substitutes, though not perfect substitutes. Commodity imports and immigrant labor are

53 The underlying idea is to treat trade as an input to final GNP under the argument that virtually all goods in trade must be processed further for final sale. See Kohli (1991) for an excellent development of the theory, econometrics, and results from this research.

54 Letting \( p, z, \) and \( b \) be the parametric price vector, endowment vector, and trade balance, the indirect trade utility function is defined as:

\[
H(p, z, b) := \max_{y} \{u(y) \mid p \cdot y \leq G(p, z) - b\} = V(p; G(p, z) - b),
\]

where \( G(A) \) is a standard GNP function and \( V(A) \) is the indirect utility function of the representative consumer.

55 Kohli (1993) directly estimates a symmetric normalized quadratic GNP function on the same Swiss data. The results are broadly the same, increased immigration reduces home wage, but only weakly; and trade and immigration are found to be complements.
found the be both Allen-Uzawa and Hicks $q$-complements. Once again, the magnitude of the estimated effect of immigration on native wages is negative, but quite small. However, Kohli simulates a short-run model in which the wage is downward inflexible, and finds the effect on home labor displacement to be large.

Hijzen and Wright (2005) provide an application of the GNP function approach to United Kingdom immigration for the period 1975-1996. This approach allows one, in line with the recent debate on immigration and labor market adjustment, to simultaneously study adjustment through changes in relative factor prices and adjustment through changes in the factor mix. They extend the work by Kohli (1999) by also examining how migration affects the relative returns of skilled domestic labor relative to unskilled domestic labor. They find that an increase in the number of unskilled migrants reduces the wages of unskilled domestic workers. However the quantitative impact of immigration is small. No discernible impact of migration is found for skilled native workers. Moreover, they find some suggestive evidence that changes in the output mix help alleviating the potentially adverse effects of immigration on the structure of wages in an open economy. The results also suggest that unskilled migrant workers and imports are substitutes in production, whilst skilled migrant workers and imports are complements.

The final body of work that seeks to evaluate the effect of immigration in a context that incorporates trade in an explicitly Heckscher-Ohlin framework. An early contribution by Horiba and Kirkpatrick (1983) examined direct and indirect (i.e. trade embodied) flows of labor between the North and South United States in 1965-1970, finding that endowment

---

56 Interestingly, imports and capital are Allen-Uzawa substitutes, but Hicks $q$-complements.
57 The United Kingdom provides an interesting application for a number of reasons. First, and foremost, the UK is generally considered to be a relatively flexible economy. Consequently, we would expect that the channels through which foreign workers are absorbed into the labor market to operate more strongly. One of the underlying assumptions of the GNP function is indeed that markets operate perfectly. Second, the United Kingdom, in contrast to the United States for which most work on migration has been conducted, may be deemed a small open economy.
convergence was relatively small, though in the right direction (i.e. labor and labor-intensive products are Southern exports), while the indirect labor flows were considerably larger and seemed to be doing most of the work in equalizing factor prices between regions. More recently, Horiba (2000) finds essentially the same results for 1975-1980. In particular, this work again finds that, migration and trade flows are consistent with the underlying trade model, the migration channel involves relatively small adjustment while the indirect trade in factors is considerably larger. These results are closely related to a growing body of trade research whose results suggest that the HO model, under various plausible extensions of the model (e.g. the presence of trading costs or Hicks neutral international differences in technology) and generalization of the Rybczynski theorem, does a reasonably good job of accounting for production patterns, and research on growth which fails to find a link between migration and convergence.58

Related to this work is a pair of important papers by Hanson and Slaughter (2000) and Gandal, Hanson, and Slaughter (2004), the first dealing with the US the second with Israel. These papers are based on a clever accounting decomposition that seeks to identify the contributions of output-mix change and technological change in adjusting to endowment shocks. In the US case, Hanson and Slaughter (2000) present results consistent with productivity-adjusted factor-price equalization across states and, further, present evidence suggesting that states have absorbed changes in labor endowments primarily via skill-biased technological change which is common across all states and, secondarily, via changes in output mix. That there should be evidence of output-mix adjustment in a period of rapid and substantial technological change strikes us as important, especially considering Horiba’s

58 On the subject of the endowment-output link, and the ways results vary in moving from inter-regional to international environments, see: Davis, Weinstein, Bradford, and Shimpo (1997); Davis and Weinstein (1997); Bernstein and Weinstein (1998); and Kim (1999). For the lack of a relationship between migration and convergence, see Barro and Sala-i-Martin (1991) and related work by Kim (1998) suggesting an important role for industrial structure, as well as technological change, in accounting for convergence.
findings for a technologically less dynamic period. However, such evidence does not exist in the Israel case, where Gandal, Hanson, and Slaughter (2004) find that global changes in technology were (more than) sufficient to absorb the huge, relatively skilled influx of immigrants from Russia. In addition to the finding that output-mix adjustment was playing a role, there are two important implications of this work for the discussion to follow. First, there is some suggestion that, at least among relatively developed economies, the assumption of a common technology across countries may be less of a distortion that assuming a common technology across a finite period of time (at least during a technologically dynamic period). Second, while appropriately constructed comparative static analysis will identify important forces operating at the level of the economy as a whole, dynamic forces that are not incorporated in the analysis might well overwhelm the static forces. On the other hand, since these forces are both less well understood and less controllable, their relevance for policy analysis is very unclear.

Before turning to a discussion of theoretical models that might be used to evaluate the various econometric results we have discussed to this point, we briefly note a number of papers that have focused primarily on the issue of whether trade and immigration are substitutes or complements. This is an issue of some importance in evaluating theoretical frameworks, and we have already seen that there is some evidence on both sides, though the systematic evidence to this point seems to favor a complementary relationship, though here we must be careful about our meaning of “complement”. In Kohli’s (1993, 1999) work, cited above, he used the standard production theoretic definitions and found trade and immigration to be complements under both the Allen-Uzawa and Hicks definitions. On the

---

59 In addition to technological change, we would also consider factor accumulation to be a dynamic force of considerable significance. It should probably be noted, as Hanson and Slaughter do, that capital accumulation may be playing a large role as well.

60 Wong (1986, appendix A) provides a useful survey the various notions of substitutability/complementarity between factor mobility and trade.
other hand, Horiba’s work suggests a substitutive relationship defined by movements that have the same effect on factor prices–direct and indirect mobility of the sort identified move two economies toward factor-price equalization. This is essentially Mundell’s (1957) definition of substitutes for the trade/factor-mobility relationship. The primary tool used in this literature has been the gravity model, which has a long history of application in the study of both trade and migration separately, and has recently been used to study the relationship between the two. Thus, Molle and van Mourik (1988) estimate a gravity model, on country pairs, in which the dependent variable is migration and, in addition to the standard regressors in gravity equations, various measures of trade intensity were included, and find a statistically significant, positive effect of trade on immigration. Similarly, a number of studies have examined the effect of immigration on trade in a gravity framework, uniformly finding the relationship to be positive (Gould, 1994; Head and Ries, 1998; Wagner, Head and Ries, 2002; Helliwell, 1997; Dunlevy and Hutchinson, 1999). While these results may suggest a complementary relationship, the bilateral data used in these analyses do not really get at the same question as that posed at the beginning of the paragraph.61 A recent paper by Collins, O’Rourke, and Williamson (1999) directly tests for a relationship between trade and migration using data on individual countries and in a panel for 1870-1936. They find no evidence of substitution and find evidence of complementarity in a number of cases.

As with the research in the labor econometric tradition, we consistently find evidence that immigration has had no significant labor market effects in industrial countries, except on other immigrants of similar vintage and origin, and on the very least skilled of native workers. We have also found mixed evidence suggesting that trade and immigration are complements in the production theoretic sense, though possibly substitutes in the Mundellian

---

61 In fairness, with the exception of Molle and van Mourik (1988), this work is primarily trying to identify the presence of network factors in shaping patterns of trade. Thus, the presence of a positive relationship in these regressions is taken as evidence of such network factors. This work is more closely related to the work of Rauch (1999) than to standard models of trade and migration.
factor-price equalization sense. We now turn to a brief overview of trade theoretic models that might be used to interpret these results.

**B. Trade Theory as a Guide to Interpreting Empirical Results**

We have already noted that the HOS interpretation of the general finding of very small measurable effects of immigration on labor market outcomes rests delicately on the HOS assumption structure. In particular, as Jones and Scheinkman (1977) argue, the $2 \times 2$ structure of the HOS model is special for at least two reasons: 2 is a small number; and the number of factors ($m = 2$) is the same as the number of goods ($n = 2$). In the next section we briefly reprise the analysis of Jones and Scheinkman with a focus on the factor-price insensitivity theorem that is the key to the basic comparative static analysis of immigration. We then turn to product market issues that are relevant to evaluating the empirical work—the presence of nontraded goods and intermediate goods. Finally, we briefly discuss current research on increasing returns and agglomeration. In all of this research, the only essential difference between labor mobility and capital mobility is that labor consumes in the country where it produces and capital does not.\(^{62}\) The differences become considerably greater when we introduce policy distortions, but these carry us beyond the limits of this paper.\(^{63}\)

1. **Immigration in the Open, Neoclassical Model: Dimensionality**

For the general case of $m$ factors used to produce $n$ final goods, according to production functions which are linear homogeneous, strictly quasi-concave, and twice

---

\(^{62}\) In the small comparative static literature on temporary migration, even this difference disappears.

\(^{63}\) The two policy distortions that have received the most attention in the general equilibrium analysis of immigration are enforcement activities related to illegal immigration and various aspects of immigration in the context of government transfer policies. The former literature was initiated by Ethier (1986b; also see Bond and Chen, 1987), while the latter is a mainstay of public economics research (e.g. Wildasin, 1991, 1994; Myers and Papageorgiou, 2000; and Bruckner, 2000). Both of these have also been used as the basis of political economy models, which are also outside the scope of this paper.
differentiable, we can represent the technology with the \( m \times n \) matrix \( A = [a_{ij}] \) and the equilibrium by the system of full-employment and zero-profit conditions:\(^{64}\)

\[
Ay = z \\
A'w = p.
\] (25)

Differentiating this system yields:

\[
\sum_{k \in I} s_{ik} dw_k + \sum_{j \in J} a_{ij} dy_j = dz_i, \ \forall i \in I; \\
\sum_{i \in I} a_{ij} dw_i = dp_j, \ \forall j \in J,
\] (26)

where we have used the fact that \( da_{ij} = \sum_{k \in I} \frac{\partial a_{ij}(w)}{\partial w_k} dw_k \), and \( s_{ik} := \sum_{j \in J} y_j \frac{\partial a_{ij}(w)}{\partial w_k} \), to substitute for \( \sum_{j \in J} y_j da_{ij} \) in the first equation; and \( \sum_{i \in I} w_i da_{ij} = 0 \), by cost minimization, in the second.\(^{65}\)

In matrix form (26) is:

\[
\begin{bmatrix} S & A \\ A' & 0 \end{bmatrix} \begin{bmatrix} dw \\ dy \end{bmatrix} = \begin{bmatrix} dz \\ dp \end{bmatrix}.
\] (27)

Now if we denote the first partitioned matrix, \( B := \begin{bmatrix} S & A \\ A' & 0 \end{bmatrix} \), and its inverse

\( B^{-1} := \begin{bmatrix} K & G \\ G' & L \end{bmatrix} \), the comparative static in which we are primarily interested is contained in

\[
\begin{bmatrix} dw \\ dy \end{bmatrix} = \begin{bmatrix} K & G \\ G' & L \end{bmatrix} \begin{bmatrix} dz \\ dp \end{bmatrix}.
\] (28)

The submatrix \( K \) gives us the effect of a change in the endowment vector on the wage vector.

The initial result of interest can be had by considering:

\[
BB' = \begin{bmatrix} S & A \\ A' & 0 \end{bmatrix} \begin{bmatrix} K & G \\ G' & L \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}.
\] (29)

\(^{64}\) The \( j \)’th column of the \( A \) matrix gives the technology in use in the \( j \)’th sector. Recall that the \( a_{ij} \) are derived from cost minimization and are, unique, homogeneous of degree zero, functions of \( w: a_{ij}(w) \). The analysis sketched in this paragraph is drawn directly from the appendix to Jones and Scheinkman (1977).

\(^{65}\) Note that \( s_{ik} \) shows how economy-wide demand for factor \( i \) changes in response to an increase in \( w_k \). In proportional changes this is the economy-wide elasticity: \( \sigma_i = \frac{w_i s_{ik}}{y_i} \).
Note that $A'K = KA = 0$. Thus, all vectors of endowment changes which are linear combinations of the columns of $A$ leave factor prices unchanged (Jones and Scheinkman, 1977, pg. 927). Recall that the $j$'th column of $A$ gives the technique in use in sector $j$.

Together, the optimal input vectors described by the columns define a cone in input space. When each of these vectors is unique, $A$ has rank $m = n$, and the cone is non-degenerate. As long as the new endowment remains in the interior of that cone, i.e. $m = n$ before and after a change in the endowment vector, such an endowment change will have no effect on factor prices as changes in output-mix will suffice to absorb the change. This generalizes both the result and the economic logic of the factor-price insensitivity theorem sketched in section I.66 67

When there are more goods than factors, $m < n$, factor-price insensitivity continues to be the expected outcome. In this case, $m$ of the $n$ technologies will define a non-degenerate cone in $m$-dimensional factor-space. As in the $m = n$ case, the endowment cannot change so much that the new endowment falls outside the initial cone of diversification, implying that some commodities cease production. Subject to that caveat, however, factor-price insensitivity continues to hold in the $m < n$ case.

When there are more factors than goods, $m > n$, the situation is very different. There are now insufficient distinct technologies in use to create a non-degenerate cone in the $m$-dimensional input-space. It is still possible to find changes in the endowment vector that fall within the cone, and thus satisfy factor-price insensitivity, but they are in the nature of very special changes. This is easily illustrated in figure 1, where there is a 2-factor × 1-good economy. The (degenerate) cone is the ray through the initial endowment point and any

---

66 The notion of technology in use requires that each of the $n$ sectors applies a distinct technique. Thus, another way of saying this is that the rank of $K$ is $m - n$, so when $m = n$, $K$ is the zero matrix.

67 There is a sizable literature relating to Mundell’s (1957) analysis of the extent to which factor mobility can substitute for commodity mobility, the reverse of Samuelson’s factor-price equalization question (Markusen, 1983; Wong, 1986; Ethier and Svensson, 1986). Ethier (1996) provides a nice overview. While these results are an important addition to the literature on factor-price equalization, their relationship to factor-price insensitivity (the single-country result) is weak. As a result, we do not pursue these results here.
change in endowment that falls on this line will satisfy factor-price insensitivity, but any
other pattern of change will result in a change in \( \omega \). Thus, when \( m > n \) a change in the
endowment will generally produce a change in the endowment vector.

Perhaps not surprisingly, since there is really nothing to say in the perfectly
competitive, \( m \leq n \) case, there has been some analysis of the labor market effects of
immigration in the \( m > n \) case. A particularly popular model in this regard has been the 3-
factor \( \times 2 \)-good model. Ruffin (1981) extends the “friends and enemies” language of Jones
and Scheinkman (1977), to include: factors \( i \) and \( k \) are friends if \( \frac{\partial w_i}{\partial z_k} > 0 \) and enemies if

\[
\frac{\partial w_i}{\partial z_k} < 0.
\]

For the general \( m > n \) case, Ruffin uses Samuelson’s reciprocity relationship, to
display that friendship is reciprocal (i.e. if \( i \) is a friend to \( k \), the \( k \) is a friend to \( i \)); and that every
factor has a friend. For the 3-factor \( \times 2 \)-good case, Ruffin then shows that, without loss of
generality, there is always an assignment of labels to factors such that, for \( I = \{1,2,3\} \) and \( J =
\{1,2\} \):

\[
\frac{a_{i1}}{a_{i2}} \geq \frac{a_{k1}}{a_{k2}} \geq \frac{a_{31}}{a_{32}},
\]

and the weak inequalities can be replaced by strict inequalities if no two factors are used in
the same proportions across sectors. In the case of strict inequalities, Ruffin refers to factors
1 and 3 as extreme factors, and 2 as the middle factor—as a result, in the case of strict
inequalities, Ruffin refers to (24) as the factor extremity condition. Under this condition,
Ruffin proves:

*Theorem:* If there are three factors and two goods produced in a competitive, small
open economy operating under constant returns to scale, an increase in the supply of
an extreme factor will benefit the middle factor and hurt the other extreme factor.

Thus, under the labeling convention of the factor extremity condition, factors 1 and 3 are
enemies to each other and friends to 2, while factor 2 is a friend to both 1 and 3. One of the
attractive features of this model is that the result depends only on factor-intensity ranking, and not at all on details of the substitution elasticities.\textsuperscript{68}

Thompson and Clark (1983) apply this framework to US data, with $y = \{\text{agriculture, manufacturing}\}$ and $z = \{\text{skilled labor, unskilled labor, capital}\}$, finding capital and skilled labor to be extreme factors, while unskilled labor is the middle factor. Thus, if current immigration in the US raises the endowment of unskilled labor, Ruffin’s theorem implies that the wage of unskilled labor should fall and that of skilled labor rise, producing a change in the wage premium of the sort that was observed in the 1980s. Thompson and Clark (1990) extend this analysis to a $4 \times 3$ model–$y = \{\text{agriculture, services, manufacturing}\}$ and $z = \{\text{unskilled labor, semiskilled labor, skilled labor, and capital}\}$–finding unskilled labor a friend to capital and skilled labor, and an enemy to semi-skilled labor. The magnitudes, in both cases, are such as to suggest that the only sizable effects will be that of unskilled labor on itself. These results seem loosely consistent with those reported in both the labor and trade research frameworks we have reported already.\textsuperscript{69}

It should be noted that all of the implementations of the $m > n$ model we have mentioned were of the form $m = n + 1$. One might reasonably wonder whether the results applied in this work are sensitive to this particular dimensional assumption. Jones (1985) shows that the answer is “yes”. Specifically, Jones notes that when $m = n + 2$ (or more) it is no longer possible to identify the qualitative effect of endowment change on factor wage

\textsuperscript{68} Of course, the substitution elasticities affect the magnitudes. Jones and Easton (1983) is an exceptionally useful development and exposition of this model, with a particular emphasis on the role of economywide elasticities and their conceptualization. Thompson (1983 a,b) and Davies and Wooton (1990) develop the application of this model to migration in some detail. Clark and Thompson (1990) and Davies and Wooton also consider the effect of migration on the source country in this model.

\textsuperscript{69} Clark and Thompson (1986) carry out a similar analysis on Canadian data, but use 5 job categories and capital for a $6 \times 5$ model. In that case, all but highest skill category (L1: professional, technical, managerial, and administrative) bear the same relationship to capital and L1, and to each other: enemies to each other and friend to capital and skilled labor (i.e. L1).
from factor intensity. It is now necessary to have specific knowledge of factor
substitutability as well.70

We have spent considerable time on factor-price insensitivity in the standard HO
environment because there seems to be quite a bit of misinterpretation. We start by recalling
that the sole relevant difference between the basic frameworks in use by labor and trade
economists is dimensionality. First, dimensionality is not nearly so damaging of factor-price
insensitivity as it is of factor-price equalization. The former is a one-economy comparative
static result, while the second seeks to make a multi-country comparison, requiring both
strong assumptions about internationally common technology and global univalence to make
the comparisons. While Hanson and Slaughter’s work suggests that technological change
within a country may interfere with inference in periods of large-scale technological change,
the multi-good framework seems quite appropriate as the basis of but-for analyses of
immigration shock. Second, contrary to some of the assertions by both trade and labor
economists, it does not seem to us that the choice between $m \leq n$ and $m > n$, as interpretive
frameworks, should rest on whether or not the framework generates income distribution
effects from immigration.71 Given the very weak evidence in favor such income distribution
effects, this seems doubtful in any event. But it seems that, on any but fairly short-term
interpretations of the concepts of commodity and factor, there are massively more
commodities than factors, and in this case the logic of factor-price insensitivity holds quite
straightforwardly.72 Note that we are not arguing that factor-price insensitivity actually
obtains, but that, within the parameters that are commonly agreed in the basic labor and trade

70 Jones then offers a useful result which rules out certain patterns of response of the wage vector to changes in
the endowment vector, and which generalizes Ruffin’s theorem, given above.
71 Trade economists like Thompson and Wooton seem to make this argument as the entering wedge of a political
economy argument, while labor economists make the argument to shore up the foundations of their estimating
framework.
72 See Bernstein and Weinstein (1998) for a recent development of the dimensionality argument, and its
implications for tests of directions of trade predictions.
theoretic traditions, \( m \leq n \) seems a more plausible assumption, from which factor-price insensitivity follows. We should generally expect adjustment at the output-mix margin to play a considerable role in responding to factor immigration. If the mechanism breaks down, it must be as a result of deviations from those elements of the basic model that are shared between trade and labor economists, and not on dimensionality. Thus, we now turn to several plausible sources of such deviation.

2. Immigration in Economies with Nontraded and Intermediate Goods

The most obvious place to start looking for deviations from the model developed in the previous section is nontraded goods. It is at least arguable that a substantial portion of any economy, and any OECD economy in particular, is nontraded.\(^{73}\) We begin, as in the previous section with a brief discussion of the effect of nontraded goods on factor-price insensitivity in general, and then consider several specific versions that have been applied to the analysis of immigration. In generalizing the analysis of factor-price insensitivity to the case of nontraded goods, there are at least two important considerations, both related to dimensionality. In the pure generalization of the comparative static analysis, we retain the assumption that factors are not immobile internationally and treat immigration as a comparative static increase in the endowment of some factor. In that case, we want to know whether the dimensionality of the model affects conclusions with respect to the sign of \( \frac{\partial w_j}{\partial z_k} \).

However, when we turn to nontraded goods, it would seem to be incumbent upon us to be more explicit about factor mobility as well as good mobility. Here we will want to note some

\[^{73}\text{It has been suggested, for example, that the following sectors be considered nontraded: government services; retail trade; wholesale trade; personal household services; restaurant services; health services; and construction. While government services should probably be netted out of any empirical analysis as a non-market sector, the remainder would be a sizable share of any economy.}\]
results that treat the appropriate generalization in terms of numbers of things (goods and factors) that are traded versus number of things that are not.

Suppose that there are \( n^T \) traded goods and \( n^N \) nontraded goods, so \( n = n^T + n^N \). We start by noting that as long as \( m \leq n^T \) the analysis of the previous section is essentially unchanged (Woodland, 1982; section 8.2.6.). That is, as long as there are at least as many traded goods as factors, factor-price insensitivity will continue to obtain, under the same restrictions, and for the same reason, as for the case without nontraded goods.\(^74\) Note the implication that, from factor-price insensitivity and cost minimization by nontraded good producers, nontraded commodity prices are determined by supply condition alone.\(^75\) As above, if \( m > n^T \) endowment shocks will generally have an effect on wages, but this is not so much a consequence of nontraded goods as it is of the dimensionality of the model. Now suppose, instead, that there are \( m^T \) traded factors and \( m^N \) nontraded factors, so that \( m = m^T + m^N \). Ethier and Svensson (1986, pg. 28) give as a condition for factor-price insensitivity that \( n + m^T \geq m \).\(^76\) That is, the total number of international markets (for goods and internationally mobile factors) must be at least as great as the number of factors. Even reducing this condition to \( n^T + m^T \geq m \), it seems to us that this condition is likely to hold, but it also strikes us that this is a considerably more uncertain proposition than that \( n \geq m \). Thus, it is probably not surprising that a number of studies emphasize the role of non-traded goods. Before turning to a brief consideration of these, we again note that the essential thing here is not nontraded-ness, but dimensionality.

\(^74\) Deardorff and Courant (1990) raise the question of the effect of nontraded goods on the size of the cone of diversification, concluding that nontraded goods tend to reduce the size (though, of course, not the dimensionality) of the cone. In the context of factor-price insensitivity, this suggests that nontraded goods narrow the range of endowment shocks that are consistent with factor-price insensitivity, but does not undermine the basic logic. That is, as long as the same traded commodities are produced before and after the endowment shock, the \( \partial w_i / \partial z_j \) will be zero. Factor-price insensitivity will hold.

\(^75\) This follows from XXX including the nontraded goods. Since the \( a_{ij} \) are functions of \( w \), which is unchanged, \( c_j \) is unchanged, and, thus, prices of nontraded goods are fixed.

\(^76\) Note that this condition can also be stated as \( n \geq m - m^T = m^N \), or when \( n^T < n \) to \( n^T \geq m^N \). That is, the number of traded goods must be at least as large as the number of nontraded factors, which is the essential condition we have been noting throughout this section.
One approach to generating wage effects from endowment changes is to assume that $m = n$, but $n^N \neq 0$, ensuring that $m > n^T$. Rivera-Batiz (1982b) develops a 2-factor × 2-good model with one traded good and one nontraded good. The traded good price is locked in by the small country assumption, but the nontraded good price is determined by the interaction of supply and demand conditions in that market. Consider a setup like that in figure 2, with good 1 skilled-labor intensive and good-2 unskilled labor intensive, but now suppose that only good 1 is internationally traded. Once again, we suppose that there is an increase in unskilled labor. Following the logic of the Rybczynski theorem, as illustrated in figure 3, at fixed commodity prices, output of the traded good (1) will fall and output of the nontraded good will expand. Now, however, while the price of the traded good is fixed, the excess supply of the nontraded good at initial prices leads to a fall in its price. This will cause the unit value isoquant for good 2 to shift outward and a new factor-market equilibrium will be established at a lower $w_L$ and a higher $w_S$. This dependence of factor-return on endowment makes this model a popular framework for the analysis of migration. It is useful to note, however, that if the world actually looks like figure 3, treating it like figure 1 for simulation purposes will produce an overestimate of the labor market effect because it will not be taking into account the adjustments in both demand and production that go into the determination of the final equilibrium here.

---figure 3 about here---

77 Bond (1993a) presents an alternative approach, emphasizing cone conditions; Kondoh (1999) extends the Rivera-Batiz analysis to consider demand effects based on duration of migrant stay; while Hatzipanayotou (1994) studies the interaction between various types of policy and immigration. This model has been used, in particular, to study welfare effects of migration in a two-country system, with a sending and a receiving country, e.g.: Krauss (1976); Rivera-Batiz (1983); and Rübel (1994). Interesting generalizations to higher dimensions can be found in Neary (1989) and Bond (1993b).

78 The magnitudes of the new factor wages will depend on both production and demand conditions, but the direction of the changes can be signed unambiguously. For example, in Jones’ $\theta$-notation, Rivera-Batiz gives the effect on the return to $L$ as:

$$\dot{w} = -\frac{\theta_{LX}^L}{\lambda} \dot{P}_X \hat{P}_X$$ and $$\dot{w}_L = -\frac{\theta_{LX}^L}{\lambda} \frac{\hat{P}_X}{(\sigma_D - \sigma_S)}.$$
Kuhn and Wooton (1991) develop an extension of this model to the 3-factor \times 3-good case, with one non-traded good. Let good 1 be the exportable, good 3 the importable, and good 2 the nontraded good. Furthermore, we number factors so that, relative to the two traded good prices, good 1 uses factor 1 intensively, good 2 uses factor 2 intensively, and factor 3 is a middle factor. That is, we have the relationship given in (24). The analysis generates results that are very much like those of Ruffin (1981), referred to above: an increase in the endowment of any factor reduces the return to that factor; the middle factor is a friend to both extreme factors; and both extreme factors are enemies to each other.\(^{79}\) The authors then use US data for 1960, 1970, 1980, and 1984 to construct the vectors \(z = \{L,S,K\}\) and \(y = \{\text{exports, imports, nontraded}\}\), concluding that in nearly every case: skilled labor is the extreme factor in exports; unskilled labor is the extreme factor in imports; and capital is the middle factor (and used intensively in nontraded goods). Thus, given the model structure, the implication is that an increase in the endowment of unskilled labor, via immigration, will tend to reduce the wages of both types of labor and raise the return to capital. The fact that, at this level of aggregation, there is very little evidence of this pattern of effects, might suggest that the dimensionality assumption ensuring such effects are not a part of the world we observe.

In closing this section we briefly note that neither the existence of intermediate goods nor of joint production undermines the general logic of factor-price insensitivity. This has been known since the pioneering analysis of McKenzie (1955).\(^{80}\) Thus, factor-price

\[\text{where the } \theta_{ij} \text{ denote factor distributive shares and the } \lambda_{ij} \text{ are shares of the endowment of factor } i \text{ in use in sector } j; \text{ the hats denote proportional changes, } \sigma_5 \text{ is the elasticity of substitution between goods } 1 \text{ and } 2 \text{ along the transformation curve, and } \sigma_D \text{ is the elasticity of substitution in consumption between good } 1 \text{ and } 2. \text{ Since, } \sigma_5 \text{ and } \sigma_D \text{ are both positive, and the determinants } |\lambda| \text{ and } |\theta| \text{ have the same sign, and the traded good price is unchanged, these imply a fall in the real wage of } L. \text{ The equivalent expressions for } w_S \text{ show a real increase.} \]

\(^{79}\) This result relies on a pair of “normality” conditions—one a restriction on demand and the other on supply.

\(^{80}\) See Woodland (1982; chapter 5) for an exceptionally clear development of these issues. Chang, Ethier, and Kemp (1980) is also very useful on the issue of joint production. Recently, there has been some discussion of the effect of joint production on the likelihood of factor-price equalization paralleling that discussed in note 65. See, in particular: Samuelson (1992); Jones (1992); and Albert and Kohler (1995).
insensitivity is a surprisingly robust property of neoclassical/competitive trade models under a variety of assumptions about the production side of the model.

3. Increasing Returns, Agglomeration, and Immigration

To this point, we have stressed the fundamental theoretical consistency of the frameworks in use by trade and labor economists. As we have noted, both rely on assumptions of perfect mobility within the relevant geographic market, perfect competition within all markets, constant returns to scale in production, *et cetera*. The only essential difference between the two is dimensionality, and here we have argued for a presumption that the number of goods is at least as large as the number of factors of production. Under this assumption, we have seen that factor-price insensitivity generally obtains. This seems broadly consistent with the overwhelming majority of studies which find, at most, small wage/employment effects. At a minimum, something other than wage adjustment is going on—it might be adjustment on the output margin (as implied by factor-price insensitivity), or endogenous technological adjustment (as suggested by Ethan Lewis), or something else, but there just is not much evidence of major adjustment on the wage margin.

This said, the apparent existence of quite distinct local economies is problematic for either approach. Even within the state of California, we can observe quite distinct production structures between a northern economy characterized by high tech industry supporting a high wage structure and a southern economy in which low tech industry supports a low wage structure. If local labor markets are somehow segmented, this multi cone structure is an equilibrium. However, the existence of free trade and free factor mobility between northern and southern California would seem to be completely inconsistent with an account of this

---

81 Similar structures certainly exist in other countries—northern and southern England is another example.
sort. This leads us to the question of whether some form of geographic model supports the endogenous determination of multiple local cones.

Bernard and Jensen (2000) note that although the United States as a whole experienced increasing wage inequality, that this fact disguised very different patterns of wage inequality at the state level. In fact, some states experiencing sharp declines in inequality. They argue that this fact argues against the view of a well-oiled, highly-integrated U.S. economy – at least in the short- to medium-terms. In response to employment shocks, Blanchard and Katz (1992) found that labor markets were integrated after a period of 10 years. Interstate migration does indeed act to smooth shocks, but it only does so very slowly. Wage adjustments were also extremely sluggish, with some effects often lingering beyond 10 years. In other words, shocks to regions are not rapidly transmitted to other regions. This suggests that even the most carefully crafted research, e.g., such as Borjas’, which divides workers into industry, occupation, education, experience, race and sex may be biased because it does not also distinguish between locations – or separate labor markets. There are at least two further implications of this research. Firstly, the measured effects on local labor markets are genuinely indicative of the small adverse impact that new immigrants have on native workers and second, that the skating rink hypothesis, in which each new foreign unskilled worker forces a native off the ice (i.e., to migrate to another region), is dubious.

In more recent work, Bernard et al. (2002, 2003, 2004) show that relative wages vary across regions in the United States and the United Kingdom. They also show that the type of industry varies with the skill wage premium and that skill-abundant regions exhibit lower skill premia than skill-scarce regions. Hence, firms adjust production across and within regions in response to relative wage differences.

In a Heckscher-Ohlin model with multiple cones of diversification one implication is that different regions with different skill endowments have different relative wages. In
equilibrium, regions abundant in skilled workers offer lower skill premia. This finding is at seemingly at odds with economic geography models that have skill abundance and skill premia being positively related as a result of agglomeration externalities.

The existence of multiple cones suggests that economic activity, as well as factors of production, tend to agglomerate. Courant and Deardorff (1992) show that when factors are immobile, regions specialise in the good that uses most intensively its abundant factor. In turn, the “lumpiness” of factor endowments can constitute a basis for trade. The point of course is that regions within countries may often differ more than they do with comparable regions overseas. The lumpiness can be sustained by lower prices for non-tradeable goods or by locational amenities, such as nice weather. In the latter case, real wages wouldn’t equalise, although sunshine-adjusted real wages would. Hence, there would be no FPE. In a similar vein, Quah (1996) finds that regional per capita income varies more strongly with what happens in neighbouring locations than with other regions of the same country. Likewise, Overman and Puga (2002) show the existence of regional unemployment clusters in the EU which often span national borders. These neighbourhood effects are important, regardless of whether the neighbours happen to be domestic or foreign. They argue that the clustering reflects the agglomeration effects of economic integration. Interestingly, the polarisation into high unemployment regions and low unemployment regions cannot be being driven by migration because intra-regional migration has been falling in Western Europe.82

Hanson (2001) provides a very nice review of empirical work on the new economic geography models and describes how increasing returns to scale is one of the main reasons why industries spatially agglomerate. Increasing returns to scale or agglomeration effects can be internal or external to firms. In the former case, transport costs give agents an incentive to

---

82 The EU seems to differ from the United States, where local unemployment significantly influences the migration decisions of the unemployed. On the other hand, local unemployment has little influence on employed individuals in the United States. See Greenwood (1997).
locate close to one another. In the latter case, costs per unit decreases as the industry size increases. This can occur because a larger market lowers cost of key inputs. For example, highly specialised service industries or firms can profitably service a large number of firms in a cluster, but would not be able to do so in isolation. There could also be labor market pooling benefits, where labor is attracted to the opportunities in the cluster. This lowers costs for firms as well. Last, but not least, there may be knowledge spillovers due to the informal and formal diffusion of knowledge afforded by larger clusters. This implies an improved quality and productivity of factor inputs. In fact, individual wages increase in the presence of more-educated workers in the local labor force, which is consistent with localised human-capital externalities. An occasional finding in individual wage regressions is that a measure of average human capital is significant and positive (e.g., Dickens and Katz, 1987). Wages are also known to be higher in urban areas and areas with large consumer markets (Hanson, 2001).

However, it is unclear whether agglomeration has wage benefits for workers. For example, de Blasio and di Addario (2005) finds that industrial agglomeration affects employment probabilities, but affects neither wages nor wage premia. Such a finding suggests a moderation of the forces that lead to agglomeration as agglomeration may be partly sustained by workers migrating to higher wage areas. Worker immobility and equilibrium wage differences that are not eliminated by migration will actually act as a dispersion force (see Krugman, 1991; Ottaviano and Puga, 1998; Puga, 1999 and Neary, 2001).

--Figure 4 about here—

When regions are skilled labor-abundant, they are attractive to skill-intensive industries. This is illustrated in Figure 4, which is taken from Bernard et al. (2004). Unit-value Leontief isoquants for three industries are shown. Endowments of skilled (or non-
production, \( N \) and unskilled (or production, \( P \)) workers for regions \( A \) and \( B \) are also shown. There are two cones of diversification. The relative wage of skilled workers is easily seen to be lower in \( A \), which is relatively skill-abundant. Obviously, the production structure will also differ in regions \( A \) and \( B \).

The authors reject absolute FPE and, unlike Hanson and Slaughter (2002), relative or productivity-adjusted FPE for both the United States and for the United Kingdom. They argue that this is due to the presence of multiple cones of diversification and not to region-industry technology differences, agglomeration or increasing returns to scale. Factor immobility, of at least one of the factors, prevents regional factor prices (and endowments) from converging to a common value across the country. Once again, regions exhibit systematic differences in production structure.\(^{83}\) Thus, one of their other major findings is that firms reallocate production facilities that best match their factor needs, i.e., using their words, “industries move towards workers more readily than workers move towards industries” Bernard et al. (2004, p.14).\(^{84}\)

**IV. How Does Adjustment Occur**

Although subject to varying interpretations, the finding of small local labor market effects has been remarkably robust and consistent with the findings from the econometric studies. LaLonde and Topel (1991) estimate the elasticities of complementarity between immigrants and natives and between new immigrants and older cohorts of immigrants and find both to be very small. Taken in conjunction with their analysis of wages and earnings

\(^{83}\) Kim (1999) finds that factor supplies can account for a large share of the cross-sectional output variation across U.S. states.

\(^{84}\) Hanson (2001) argues that available evidence in the economic geography literature favours demand-driven agglomeration in which firms are attracted to densely concentrated regions and large potential local markets (Krugman, 1991); rather than in which workers benefit from being close to other workers (as in Black and Henderson, 1999).
changes in local labor markets, they conclude that the wage effects of immigration are “quantitatively unimportant”. Based on studies currently in print at the time that we are writing this paper, it appears to us, that such a conclusion is inevitable. Our survey of the research to this point suggests that the real issue now is how economies adjust to rather large immigration flows. Broadly speaking there are three main possibilities: first, compensating flows of natives offset immigration shocks; second, output adjustment and trade between local economies equalizes factor prices; and third, technological change absorbs the immigration shock.

First, some authors argue that there are compensating flows out of labor markets to which immigrants flow, i.e., native out-migration. There are two points here: first, and most important, by spreading the effects throughout the economy, native movements obscure the quantitative effects and must be studied at the level of the economy as a whole (this is the endogeneity problem); and second, a locally large immigration shock might be swallowed up in an economy the size of the United States. The results reported above for Hong Kong and Israel would seem to be prima facie inconsistent with either claim, while the small size of results generally found even at the level of the US as a whole would seem to be inconsistent with the first. Nonetheless, there a sizable literature has developed trying to sort out the existence and magnitude of native migratory response to an immigrant shock.

Studies using U.S. data that support and propound this view include Filer (1992), Frey (1995, 1996), Walker et al. (1995), Borjas et al. (1997); and for the United Kingdom, Hatton and Tani (2005). The allocation of workers across regions within a country experiencing an immigration shock is likely to be affected by two further factors (see Greenwood, 1997; Keeton and Newton, 2005). First, migrants do not randomly allocate themselves across regions within their destination countries and secondly, native workers relocation decisions may be affected by an influx of migrants. If new migrants move to high
labor demand and high wage areas, this will both moderate wage increases in those destinations and possibly reduce the incentives of native workers to move to those areas (Borjas, 2001). On the other hand, migration may be driven by ethnic clustering and enclave-pull factors (based on a common language and culture in certain magnet cities, for instance) rather than by high wages alone (Bartel, 1989; Card, 2001). If these markets happen to have below-average labor demand, then this may encourage some native workers in those regions to move to other regions (or to lower the probability that workers in other regions will move to that particular region). The obvious consequence of out-migration is that wage effects (for the remaining native workers) may be under-estimated.

Borjas (1999) supports the view that native workers migrate out of the cities to which immigrants are attracted. Simplifying the matter, if the number of natives leaving is equivalent to the number of \textit{identically}-skilled immigrants entering, the immigration effect on wages is completely neutralised. Moreover, if the costs associated with moving are substantial, then this constitutes a serious reduction in the welfare of affected native workers; i.e., a dead-weight loss directly attributable to the increased immigration. To compound the welfare cost, if net “in-migration” by native workers to immigrant destinations falls, then this implies that workers who would have otherwise found it beneficial to move also suffer welfare losses by remaining in the jobs and locations they currently occupy (Borjas, 2003).

Borjas (2003) investigates supply shifts for education-experience groups in the economy. By focussing on changes in the relative supplies of these groups for the aggregate labor market at the national level, he avoids the problem of the internal migratory responses of natives. He obtains estimates of own- and cross-price elasticities from a CES production function, which are used to calculate the wage impact of the actual immigrant inflow into the United States between 1980 and 2000. He finds that the 11 per cent increase in the immigrant share reduces the wages of the average native worker by 3.2 per cent, with high-
school drop-outs suffering a very severe 8.9 per cent drop. Using a similar methodology, Borjas (2006) investigates the joint determination of wages and internal migration decisions of natives in local labor markets. He finds that immigration is associated with lower wages, lower in-migration rates, higher out-migration rates and lower labor force participation of natives. He also finds that the smaller the spatial unit of analysis, the larger the out-migration of natives. Borjas argues that this fact explains the small estimated wage elasticities. In fact, he quantifies the under-estimation of immigrant wage effects as being under-estimated by as much as 60 per cent.

Significantly, there have also been studies with diametrically opposed findings, i.e., a number of other authors have found that immigration and domestic internal migration are unrelated (Wright et al., 1997; Card and DiNardo, 2000; Card, 2001; Kritz and Gurak, 2001). Wright et al. (1997)) conclude that the net migration loss of native-born workers living in large metropolitan areas is more likely the result of industrial restructuring than of competition with immigrants. Card and DiNardo (2000) find no evidence of systematic out-migration of native workers from U.S. SMSA’s in response to increases in the population of immigrants. They point to endogenous shifts in industry structure, rather than any rapid redeployments of native workers (i.e., they reject what they term the skating rink hypothesis), as the more important adjustment mechanism. Card (2005) examines whether the overall share of high-school drop-outs in a city is affected by an inflow of immigrant drop-outs. Focussing on dropouts vs. high school graduates, he regresses the relative wage and employment on the relative supply – instrumented by the immigrant dropout share - in 325 SMSA’s and finds no evidence that native migration is related to immigration supply shocks.

85 If anything, they find evidence of the opposite, i.e., native in-migration! This finding may be explicable by recent work which indicates that immigrants add to cultural diversity and to higher productivity (and therefore wages) in their settlement destinations. Ottaviano and Peri (2006) argue that this may be due to the complementarity of immigrant and native worker skills in addition to the externalities generated by skilled migrants. See also Hanson (2001).
In addition, while immigration has affected the relative supplies of unskilled workers in local labor markets, the wages and employment of low-skilled natives are unaffected by these relative supply shocks. Card once again points to the Rybczynski-type changes in industry structure.

If the labor market effects are actually not large, the adjustment must be occurring either via output adjustment or technological change. The former, as we noted in the discussion of trade theoretic models, involves a shift in output and changing structure of trade under fixed relative commodity prices.\textsuperscript{86} Hanson and Slaughter (1999) document the rapid growth in apparel, textiles, food products and other labor-intensive industries in California after the arrival of Mexican migrants. They focus on state-specific endowment shocks and state-specific wage responses. They show that the state output-mix changes broadly match state endowment changes and that variation in state unit factor requirements is consistent with factor price equalization across states. States absorb regional endowment shocks through mechanisms other than changes in regional relative factor price changes. This is consistent with the findings of Blanchard and Katz (1991) which indicate that wages and income per capita converge for American states. However, Blanchard and Katz also find that employment performance diverges, i.e., shocks to employment grow and persist.\textsuperscript{87} Overall, this is consistent with the view that small local labor market effects may be consistent with somewhat larger aggregate labor market effects.

Finally, important recent work by Ethan Lewis (2003, 2004a & b, 2005; Beaudry, Doms and Lewis, 2006) develops an analysis of endogenous technological response to

\textsuperscript{86} Again, there may certainly be commodity price adjustment and, thus, some relative wage effects via the Stolper-Samuelson mechanism. Local labor markets, however, are surely economically small in the world economy and even what look like large immigration flows are unlikely to be large enough to dramatically change national comparative advantage.

\textsuperscript{87} Decressin and Fatás (1995) have similar findings for the regions of Europe. However, they show that changes in labor force participation rates bear proportionately more of the burden of adjustment in response to labor market disturbances. Rowthorn and Glyn (2006) revisit the Blanchard and Katz study finding much less evidence of employment convergence for the US case.
immigration shocks. Drawing on earlier contributions by Acemoglu (1998) and Beaudry and Green (2005), and using firm-level data, Lewis finds little evidence of adjustment on the output margin, but strong evidence of shifts in technology in use by firms to take advantage of newly abundant unskilled labor. In particular, firms in markets characterized by large shocks of unskilled immigration adopt less computerization and less automation—there is even evidence of de-adoption of advanced technologies. Lewis (2002) finds that adjustment to immigration shocks takes place within rather than between industries, i.e., favoring the latter explanation. Hence, he concludes that the standard trade model does not adequately describe how local labor markets adjust. On the other hand, while not specifically concerned with immigration shocks per se, Hanson and Slaughter (2002) find that relative factor prices for U.S. states do not fully absorb changes in state factor supplies and also that industry production techniques are quite similar across states. They identify changes in the output of traded goods and national skill-biased technical change as being the more important adjustment mechanism through which changes in state factor supplies are accommodated. The authors argue that “[t]he combination of common production techniques and unequal wages suggests that state wage differences may be due to neutral differences in state factor productivity” (p.5) and in turn, that this result is consistent with productivity-adjusted FPE across U.S. states (Trefler, 1993). However, they also find that production techniques are most similar in neighbouring states and in states with the most similar factor supplies. Hence, their findings are consistent with the view that industries may agglomerate in certain regions (e.g., areas with large supplies of skilled labor).

IV. Conclusions

As we have stressed throughout this paper, the primary division in the literature on the labor market effects of immigration is not empirical. Unlike the related literature on the labor
market effects of trade, where there are substantial differences over matters of fact, the
impression one gets from the immigration literature is that there is a widely held, and fairly
tight, prior on essentially zero labor market impact. It is also widely agreed that there are
sizable negative effects on migrants of the same origin and vintage, and, perhaps not quite so
widely held, agreement that the small, and shrinking, group of native high school dropouts
experience economically, and statistically, significant negative consequences from
contemporary immigration.

To the extent that there is a dispute in the immigration case, it revolves around the
framework to be used for evaluating the results of the empirical work, and here the division is
very much between labor and trade economists. We have argued that the sole substantive
difference between labor and trade economists relates to the dimensionality of the model used
to evaluate the results—with labor economists preferring an \( m \)-factor × 1-final good model and
trade economists preferring an \( m \times n \) good model (with a modal preference for the \( 2 \times 2 \)
model). As long as \( m \geq 2 \) and \( n \geq 2 \), output-mix adjustment will play a role in adjusting to an
immigration shock, and the failure to account for that role will produce overestimates of the
wage (or unemployment) effects of any given shock. Furthermore, we have also argued for
the fundamental plausibility of the \( m \)-factor \( \leq n \)-good (or, perhaps even more accurately, the
\( m^N \leq n^T \)) model on essentially \textit{a priori} grounds. If this argument is accepted, there is some
presumption that output-mix adjustment fully absorbs the immigration shock. That is, if we
are going to use a perfectly competitive baseline for policy evaluation, as revealed preferred
by both labor and trade economists, our presumption should be that, immigration short of that

---

88 People often talk about a loosely construed “average” opinion on the labor market effects of trade, but this
represents a collective prior with very fat tails. The tails in the immigration case (e.g. Borjas, Briggs, Huddle)
are visible and aggressive in asserting their opinion, but seem to have very small impact on the aggregate
professional opinion.

89 It is, in fact, quite striking in the trade and labor markets case, the extent to which heated disputes about
interpretation take place between people who share a common model. As one example, see the papers by
Leamer, Krugman, Deardorff, and Panagariya in the Journal of International Economics symposium (V.50-#1,
pp. 17-116).
necessary to generate a fundamental shift in production structure has no effect on long-run labor market conditions. Factor-price insensitivity holds.

As a presumption, from which to begin an evaluation of proposed immigration policy, or an evaluation of past immigration policies, this strikes us as the right presumption. And the fact that its key implication, essentially no labor market effect of immigration, is borne out by most empirical work, should strengthen our commitment to this presumption. But it is only a presumption—a point from which we should be willing to be shifted if faced with sufficient evidence in a given case. We have argued that factor-price insensitivity is surprisingly robust to plausible variations on the basic model, but the model is, itself, very simple. There are obviously many relevant facts of economic and social life that are not part of the model, but might well affect our ultimate evaluation of immigration policy. Perhaps the most significant of these relates to short-run adjustment cost. It is now well established that the economic short-run can be chronologically rather a long time, and that these adjustment costs can be substantial.90 We only make two points here. First, these considerations are essentially orthogonal to immigration per se. That is, if we are concerned about adjustment costs borne by citizens, whether as a result of trade, immigration, technological change, or anything else, we have tools for dealing with them, and there is no particularly good reason for worrying about the source of worker dislocation. Second, this does not distinguish the labor and trade approaches. Within either framework, using immigration policy as an instrument for dealing with redistributive concerns is an exercise in (at least) second best.

This leads us to the most difficult question: if immigration is really not relevant to the long-run economic life of citizens, why does it occasionally become such a large political

90 That adjustment to a local labor shock may take a long time is one of the points that we take from the research on local labor markets that we have already mentioned, e.g.: Blanchard and Katz (1992); Decressin and Fatás (1995); and Topel (1986; 1994 a & b). On the economic effects of worker displacement, see: Topel (1990), Ruhm (1991), Kletzer (1991, 1996), and Jacobson, Lalonde, and Sullivan (1993).
issue? In answering this question, it is useful to keep several gross facts about the politics of immigration in mind. First, anti-immigrant sentiment is not a historical constant. For much of human history, relatively large movements of people occurred without particular comment, and often were welcomed by the political authorities (Sassen, 1999). Nationalism, especially as linked to race, is a relatively recent phenomenon (Gellner, 1983; Hobsbawm, 1993; Jacobson, 1998). However, second, there was quite active anti-immigrant politics in the late-19th and early-20th centuries, in the absence of a large, redistributive state (Higham, 1955). Third, although there is considerable evidence that the contemporary median voter prefers less, or at least no growth in, migration, voters do not generally condition their voting behavior on immigration (Gimpel and Edwards, 1999; chapter 2). What does seem to be true, however, is that general politicization of immigration seems to be associated with economic hard times (Higham, 1955; M.A. Jones, 1992). Finally, by comparison to the politics of international trade, the politics of immigration is primarily public politics. Like immigration, trade occasionally becomes a public issue, but trade is also characterized by very active purely professional politics. Furthermore, with the exception of California (in the late-19th century and in current times), the public politics of trade have dominated the public politics of immigration.

In an effort to organize our thoughts about these facts, we begin with a distinction, due to Schattschneider (1960), between democratic politics and group politics. For Schattschneider, democratic politics refers to the public politics through which a democratic civil society constitutes itself and through which it is linked to the policy making apparatus.

---

91 We abstract, of course, from migrations such as those associated with Attila the Hun, William the Conqueror; and the early American colonists.
92 The equivalent to professional politics of trade, is essentially lobbying on behalf of specific immigrants. That is, the activities of the trade bar and the immigration bar are quite different.
93 Thus, the era of high immigration politics, i.e. the late 19th and early 20th centuries, was both shorter lived and of less overall significance to public politics than international trade. We need only recall that the era of classic tariff politics ran from the end of the Civil War until the onset of the New Deal, during which period the tariff was literally the single most important continuing issue in American politics (see, e.g., Hall, Kao, and Nelson, 1998).
While elections are the final defense of democratic politics, as well as the key stimulus to public discourse, as stressed by theorists of deliberative democracy, the core of democratic politics is the public discourse itself. Furthermore, the terms of this discourse emphasize public interest and downplay private/individual interest. There is considerable evidence that the public discourse, and its emphasis on some notion of public interest, affects both attitude formation and voting behaviour. By contrast with democratic politics, group politics is explicitly about private interests. Furthermore, where democratic politics are public politics, group politics happen behind closed doors. Where democratic politics are inclusive, group politics is a game played by insiders. Because group politics are solidly rooted in relatively stable interests, they are predictable and they change in predictable ways in response to the, generally marginal, changes in the environment embedding those interests. Democratic politics are not tied down in the same way. While the location of individuals within the terms of the public discourse is certainly not independent of individual interest, not only do the terms of that discourse change, more quickly and on a much greater scale than the terms of group politics change, but individual locations within that discourse (i.e. preferences) change as well.

Before turning to the political economy of immigration, we introduce one further element, also essentially from Schattschneider (1960): organization of the public discourse in terms of a relatively stable issue on which parties are seen to differ. Consistent with Downs

---

94 On deliberative democracy, see: Rawls (1993); Habermas (1996; esp. chptrs. 7 and 8); Elster (1998); and Bohman and Rehg (1998). The first two are substantial statements of their authors’ distinctive positions, while the latter two are useful collections of essays.
95 Donald Kinder has been one of the leading scholars of such “sociotropic” politics. For an overview of the issues and evidence, see Kinder and Mebane (1983). Also useful are Markus (1988), Mutz (1992), and Mutz and Mundak (1997).
96 One of the central points of the literature on deliberative politics is that the public discourse is transformative—peoples’ opinions change as a result of taking part in democratic politics. This is the basis of the rejection of results like Arrow’s theorem, which characterizes the relevant community in terms of a fixed profile, from within the deliberative politics framework.
97 This is an essential element of the theory of party systems developed in Schattschneider (1960, chapters 4 and 5). For a more detailed development of the theory of party systems and critical elections, see Burnham (1970).
(1957) analysis of rational ignorance, the presence of such an issue serves to reduce the information costs of participation in democratic politics by providing a frame for the public discourse. When new issues emerge in the public discourse, unless they displace the dominant issue, they are evaluated to some extent in terms of that dominant issue. However, the mapping between the dominant issue and the details of the new issue will always be unclear and people will hold their policy evaluations weakly. This leads to what Nelson (1998) has called footloose policy preferences—individual and aggregate preferences subject to radical change, possibly without a change in fundamentals. This means that an issue, like immigration, that appears as a major public issue only occasionally, may have very different political meaning at different points in time. On the other hand, when an issue is the subject of long-standing, stable group politics, the terms and alignments of those politics will serve to anchor the public politics in a way that occasional issues cannot be anchored.

So what does all of this have to do with the politics of immigration, and its relationship to the politics of international trade? First, to start with the last of our stylized facts, where trade is the archetypal group issue, immigration is an occasional issue. At least since the end of the Civil War, trade has never been off the group politics agenda in the United States, though it was off the public politics agenda from about 1934 until very recently. As a result, much of the detail of tariff, and then trade more generally, policy has been determined by relatively stable and predictable forces that are quite well captured in standard political economy models. Because trade was a core alignment issue during the era

Interestingly, this issue may be in the nature of what Stokes (1963) calls a valence issue, an issue on which essentially the entire electorate agrees, which has the effect of making the party most closely identified with the issue the dominant party. Even if the issue is Downsian, what Stokes calls a choice issue, one party will generally be more closely associated with the median voters most preferred point on that issue, and thus will dominate electoral politics. It is this fact that drives the entrepreneurial search for a new defining issue, reflected in the constant attempt to find and exploit new issues. As Schattschneider explains, this is one of the engines of democratization in a democratic system. It is what renders the public discourse relevant and democracy much more than a figleaf for the system of group politics.

98 This is not to say that there is no ongoing group politics on the migration issue, but that this is a very small part of the politics of migration. The great majority of business on immigration has to do with individual cases, not classes of case.
of classic tariff politics, the public politics of trade during that period were also quite stable, much as the public politics of macroeconomic policy were quite stable in the New Deal era.99 The reemergence of trade in an era without an ongoing public discourse on trade has already revealed the risks to existing trade policy-making institutions as linkages to environmental and labor issues suggest, and the events in Seattle so vividly illustrate. Even the unruly public politics of international trade, however, remain essentially rooted in the definitions of trade policy and general expectations about gainers and losers that have been held at least since the end of the Second World War. In particular, the alignments of most participants are relatively straightforwardly predictable.

The politics of immigration are dramatically different. Following the establishment of the national origin quotas in the Johnson-Reed Act (1924), immigration essentially disappears as a political issue (democratic or group) for forty years. Interestingly, the Immigration and Nationality Act of 1965, which ended the quota system, reflected neither the emergence of new public pressure nor the operation of group politics, but rather derived from its attachment to civil rights issues and, to some extent, to a liberal framing of US international obligations (Gimpel and Edwards, 1999).100 Even the landmark Immigration Reform and Control Act (1986), appears as much as a triumph of a small, intensely interested group (organized primarily around environmental groups) over general disinterest, rather than over other organized interests.101 By 1990, however, immigration appeared briefly as if it might be a successful entrepreneurial issue, possibly even part of a realignment.

---

99 The transition from the era of classic tariff politics to the modern trade policy system was, of course, a special event and needs to be seen as such (Hall, Kao, and Nelson, 1998).
100 Note that there certainly was lobbying on the issue. Our point is simply that the driving force for the change in immigration policy was a changed understanding of domestic and international obligations on the part of Congress. The group politics surrounding the 1965 Act were not particularly active and the public politics were minimal.
While the end of the New Deal electoral system has not produced a clear aligning issue, the Republicans have been most successful in attracting significant parts of the New Deal coalition by attacking policies related to race and taxes.\(^{102}\) In this regard, immigration, and illegal Mexican immigration in particular, offered a fairly natural extension of an effective Republican strategy. As we suggest above, the public is generally not sufficiently interested in immigration to condition its voting behaviour on the issue, but in economic hard times the possibility opens. Thus, in a period of fiscal difficulty in California, Pete Wilson was able to revive a flagging gubernatorial campaign through the aggressive use of a racially charged immigration issue.\(^{103}\) Ultimately, however, while there was a brief period of interest in the issue, the race/fiscal policy mix proved unsuccessful even in other states of high migration (Texas and Florida), and today there seems to be little general interest in the issue. At least as telling, is the small impact of supposedly draconian policy changes contained in the post-1986 legislation (Schuck, 1998). Had there been a substantial change in the foundation, or organizational basis, of the interests involved in immigration policy, we would expect to see substantial change in policy, but the fundamental basis of policy is essentially unchanged from that of 1965.

Thus, the politics of immigration seem to be minimally (at most) about income distribution. Instead, immigration as a public issue appears to be parasitical on broader understandings of the key issues facing the nation. The comparison to the late-19\(^{th}\) century is telling in this regard. Immigration was publicly more prominent, stayed on the public agenda

\(^{102}\) These were essential elements of Richard Nixon’s southern strategy and the core of Ronald Reagan’s appeal to the “Reagan Democrats”. Particularly useful accounts of the general politics of race and taxes can be found in Edsall and Edsall (1991) and Plotkin and Scheuerman (1994).

\(^{103}\) On the role of taxes and race in the politics of immigration in the 1990s, see Calavita (1996) and Huber and Espenshade (1997). This also helps explain the peculiar fact from the public opinion literature that blacks and hispanics, the groups most likely to experience some wage effects, are considerably less negative on immigration than whites. When they evaluate the rhetoric on immigration in terms of a politics organized to no small extent around race, blacks and hispanics naturally evaluate this rhetoric as essentially racist, and are more likely to be concerned about that than about some highly conjectural, and in any event small, effect via the labor market.
longer, and ultimately resulted in considerably more restrictive legislation at a time when there were virtually no fiscal effects of immigration. The politics of the time again had a large component of racial concern (though the races involved were different), but in this case they were tied more to issues of the connection between race and the possibility of democratic governance and/or economic performance (Higham, 1955; Jacobson, 1998).

To this point we have focused primarily on the United States, but the comparison with Europe is informative. While there is no more evidence of distributive effect in Europe than in the US, it is the case that, at least recently, immigration has been more consistently, and more intensely, on the democratic (i.e. public) agenda in Europe than in the US. In addition, the European public discourse has long featured a concern with nationality per se. Europe is, after all, the cradle of both the modern nation state and modern nationalism. The programme of German unification turned, to no small degree, on an essentially racial notion of Germanness, where French nationalism was rooted in a notion of French cultural superiority. Trading nations like the Netherlands and Britain developed more cosmopolitan notions of national identity. Consistent with the argument sketched in this section, Germany, Austria, and France have all developed a more organized and articulated public discourse on immigration, while Britain and the Netherlands seem to experience immigration politics in ways more similar to that in US national politics.104 It would seem to be an interesting area for future research to develop a comparative analysis of the democratic politics of immigration.

To conclude, and running directly contrary to the claim running throughout Borjas (1999b), immigration policy is not about distribution. First, there is virtually no evidence of such effects; and second, the actual politics of immigration seem to be only minimally related to income distribution. Perhaps the best evidence of this is that it is exceptionally hard to

104 Interestingly, California appears to be closer to the European model.
predict anyone’s position on the immigration issue based on the usual predictive dimensions: economic liberals are on both sides of the issue (Bhagwati v. Borjas); social conservatives are on both sides of the issue (Fukuyama v. Buchanan); social liberals are on both sides of the issue (Habermas v. Rawls). None of this suggests that the politics of immigration in the future is likely to be benign. Immigration proportionally on the order of that in the late 19th and early 20th centuries raises difficult questions about the meaning of nation and the possibility of precisely the sort of liberal discourse considered central to democracy by scholars like Schattschneider, Rawls, Walzer, and Habermas. However, we fail as scholars and a citizens if we seek to force these difficult issues into the box of the politics of material interest, when they manifestly do not belong there.
Figure 1: Endowment change and factor-price change in the one-sector model. The isocost line tangent to the unit isoquant has the equation $1 = w_S S + w_L L$. 
Figure 2: Endowment change and factor-price insensitivity in the two-sector model. In this case, with two goods, if commodity prices (and technology) are
Figure 3: Endowment change and factor-price change in the two-sector model with nontraded goods. With commodity 1 traded and commodity 2 nontraded.
Figure 4: Multiple Cones of Diversification

Figure 1: Multiple Cones of Diversification
Table 1: Borjas and Ottaviano compared

<table>
<thead>
<tr>
<th></th>
<th>Change in Real Wages Using</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Our Approach 7</td>
</tr>
<tr>
<td>All Native-Born Workers</td>
<td>+1.10%</td>
</tr>
<tr>
<td>Native-Born Workers…</td>
<td></td>
</tr>
<tr>
<td>...without a High School Diploma</td>
<td>-1.2%</td>
</tr>
<tr>
<td>...with a High School Diploma</td>
<td>+0.8%</td>
</tr>
<tr>
<td>...with Some College</td>
<td>+1.3%</td>
</tr>
<tr>
<td>...with a College Degree or More</td>
<td>+1.5%</td>
</tr>
</tbody>
</table>

Source: Peri (2006)
References


84


*Political Economy*; V.73-#6, pp. 557-572.


*Princeton Studies in International Finance;* #77.

Lerner, Abba (1952). “Factor-Prices in International Trade”. 

*Journal of Economic Literature;* V.30-#3, pp. 1333-1381.

Lewis, Ethan (2003). “Local Open Economies with the US: How Do Industries Respond to Immigration?”. 

Lewis, Ethan (2004a). “How Did the Miami Labor Market Absorb the Mariel Immigrants?”. 

Lewis, Ethan (2004b). “Local, Open Economies Within the US: How Do Industries Respond to Immigration?” 
*Federal Reserve Bank of Philadelphia.*


*Econometrica;* V.23-#3, pp. 239-257.


Molle, Willem and Aad van Mourik (1988). “International Movements of Labor under Conditions of Economic Integration: The Case of Western Europe”. 


