Measuring Geographic Differences in Technical Change in the US Manufacturing Sector

Ethan Lewis
Final, 26 March 2004

I. Introduction

A large and growing literature examines the influence of advanced technologies on the relative wages and productivity of different workers (for example, Doms, Dunne and Troske (1997) and Autor, Katz and Krueger (1998)). These studies are motivated by indirect evidence that recent trends in technological change, such as the dissemination of information technology, have increased the relative demand for skilled workers and raised their relative wages and employment. Using microeconomic data on the use of different technologies by individual worker (the computer use supplements to the Current Population Survey) or by individual establishment (such as the Surveys of Manufacturing Technology), researchers generally find an association between technical change and increases in wage dispersion.

There is some evidence, however, that the pace of technological advances has varied across regions of the US. Two key facts are that changes in relative skill ratios across different US cities are nearly uncorrelated with changes in relative wages, and that the use of computers on the job shows significant differences across local labor markets.

Motivated by this evidence, the purpose of this project is to use the Surveys of Manufacturing Technology (1988, 1991 and 1993 – hereafter SMT) to develop a new geographic area series for release to the public on the use, plans to use, and reasons for using advanced manufacturing technologies, and to generate regionally representative statistical weights for future SMT micro data users interested in constructing their own regional data. As these are fairly large surveys (around 10,000 establishments were surveyed in each year) it will be feasible

1

to report statistics on at least the major categories of advanced production technology (described below) by state and in large metropolitan areas without the risk of revealing confidential information. (See Appendix 2.) A further goal will be to create more refined cross-tabulations by plant characteristics such as 2-digit industry, plant size (which Dunne (1994) shows is an important determinant of technology use), whether the plant is a military contractor.

If successful, these statistics could be of use in a wide variety of research applications. For example, having publicly available regionally stratified technology-use data would be useful to researchers or decision-makers interested in the effect that government regulations (minimum wage, environmental regulations, etc.) have on technology use and technical change. It might also be possible to use these statistics as an alternative means of asking what effect technical change has on outcomes of interest, such as employment or wages.

My own interest – and a secondary goal of this project – will be to try to assess how the characteristics of the local work force affect any observed geographic differences in the use and implementation of advanced manufacturing production technologies. This is motivated by evidence suggesting that technology responds to work force changes in a way that mitigates the impact of supply on wages (Lewis (2002)). Using the SMT micro data, this question will be evaluated by examining the impact of technology choice in a standard production function setup and comparing specifications that treat technological choice as exogenous versus others that treat the choice as driven in part by the relative availability of high skilled labor in the local labor market. Manufacturers' reports of the benefits (for example, improve product quality) and costs (for example, associated costs of training workers) of the technology – which are asked differently in each SMT – will be assessed as possible alternative explanations or channels through which the skills of the local work force operate to effect technological change.

There are two issues that need to be addressed in using the data for these purposes. The first is that since the design of the Surveys of Manufacturing Technology was not geographically stratified, it will be necessary to account for the different sampling rates in different strata when constructing local area statistics. The procedures for doing this will be discussed in detail in this proposal. A related problem is that because the coverage may be thin in smaller geographic units, it will be necessary to assess the geographic coverage from an independent data source. This will serve both to distinguish true "zeros" from instances where there is no data, and to determine if there are particular geographic units where the data are by chance highly unrepresentative so sampling weights can be adjusted to account for this. For the 1988 and 1993 surveys, the establishment counts in matched cells will be compared to counts obtained in the universe of manufacturing establishments – the 1987 and 1992 Censuses of Manufacturers, respectively. Another check will come by comparison to county and state summary data from the 1988, 1991 and 1993 County Business Patterns. From these comparisons a report on the geographic coverage of the SMTs will be produced. The report in particular will note if there are any systematic geographic shortcomings of the survey; none, however, are expected, because the SMT was a random sample.

This proposal is organized as follows. Following this introduction, the next section describes existing evidence that technical change is not geographically uniform within the US. The third section describes how the proposed geographic area series on the use of manufacturing technologies will be created. The fourth section describes the methods used to assess the impact of local workforce characteristics on the implementation of different manufacturing technologies. A fifth section concludes.

II. Evidence of Regional Differences in Technical Change

A large number of studies have documented national trends in technology that appear to raise the relative productivity of skilled workers and explain the rise in wage inequality over the past few decades in the US. Evidence of this "skill-biased technical change" has come in two forms: indirect evidence, in the form of rising wage dispersion, particularly within industries (Katz and Murphy (1992)); and direct evidence, in the form of greater use of technology which appears to raise the wages or productivity of skilled labor. The latter evidence has made use of data on the use of personal computers, R&D investment (Autor, Katz and Krueger (1998)) as well as the detailed manufacturing technologies observed in the Surveys of Manufacturing Technologies (Doms, Dunne and Troske (1997)).

Similar types of evidence can be garnered to provide support for the idea that these "national" trends in technical change, in fact, differ substantially by labor market. A number of studies have shown that changes in wages and employment rates by skill group show little relationship to changes in the supply of workers in those groups locally in comparisons across labor markets. In addition, it appears that a city's mix of industries is relatively unaffected by changes in the local skill composition of the work force – including the unexpected arrival of masses of less-educated Cuban refugees to Miami in 1980 during the Mariel boatlift (Lewis (2002)). What this means is that in cities that experience large shocks in the relative supply of a certain labor type, employers respond by hiring the more abundant type at a higher rate without large wage adjustments. Such a possibility is suggested by models of endogenous technological choice (e.g. Acemoglu (1998)) in which an increase in supply causes employers to adopt different technologies.

Other evidence on the local variation in technology comes from looking at the use of personal computers. The 1984 and 1993 October Current Population Surveys include a supplement, which asks individual workers "Do you use a personal computer directly at work?" I have tabulated the fraction of workers who answer this question affirmatively in the 44 metropolitan areas that can be identified and matched between the two surveys. This shows that while personal computer use was growing everywhere, its growth was far from uniform. Some examples of this include the fact that in 1984, in both Indianapolis and New Orleans, around 25 percent of workers were using a personal computer; by 1993, this fraction had risen to 57 percent in Indianapolis, but only to 36 percent in New Orleans. Perhaps a more compelling example is that in both San Jose and San Francisco personal computer use was at around 36 percent in 1984; by 1993 it was at 63 percent in San Jose but only at 51 percent in San Francisco. These differences in computer use in 1993 may be due in part to differences in industry composition, or the direct effect of the education of a city's workers on their individual propensity to use a computer; however, even after adjusting for these (by regressing the individual level data on dummies for industry and education categories) you find substantial differences in computer use rates across labor markets. And you can show that these residual differences are positively related to the education of the average worker in the local market. (Lewis (2002)).

To date, there is no direct evidence that the use of the specific types of technology covered by the Surveys of Manufacturers differs by labor market. However, two pieces of indirect evidence suggest that there may be such differences. First, Davis and Haltiwanger (1991) and Dunne, Foster, Haltiwanger and Troske (2002) find evidence that recent increases in wage inequality in the manufacturing sector is mostly attributable to growing inequality in the wages paid at different manufacturing plants; the latter find that these between plant differences

are associated with differences in the rate of computer investment at those plants. These between plant differences could conceivably be in part due to regional differences in the rate of computer investment or other types of technical change. Second, Moretti (2002) shows that there are regional differences in the productivity of manufacturing plants, and that these are influenced by the average education levels of workers in the local work force *outside* a given plant. These pieces of evidence suggest we may find differences in the use of particular manufacturing technologies across regions related to the skills of the local work force.

III. Geographic Area Statistics[1]

The 1988, 1991 and 1993 Surveys of Manufacturing Technology (SMT) each poll a random sample (described below) of around 10,000 manufacturing establishments in SIC industries 34 through 38 on the use of, plans for use of, reasons for use of (or for not using) categories of advanced manufacturing technologies. These technologies are listed in Appendix 1. Examples include "design and engineering technology" (including, for example, computer aided design), "fabrication and machining technology" (including "pick and place robots") and "inspection and quality control technology" (including "programmable controllers"). The survey also records establishment characteristics, such as plant size, plant age, ownership, production type, military contractor status. The 1988 and 1993 surveys ask whether not an establishment uses 17 individual technologies listed in Appendix 1; the 1991 survey asks only about four broad categories of technology (I-IV in Appendix 1), but asks (in categories) how intensively the technology is used. Each of the surveys is a stratified random sample of manufacturing establishments with at least 20 employees. Within the 3-digit SIC by 3 class size cells (20 to 99,

---

[1] Much of the descriptions of the SMT survey design in this section comes from reports written by the US Bureau of the Census (1989, 1993, 1994).

100 to 499, more than 500 employees) that make up the strata, a simple random sample was

taken, and a weight was recorded equal to one over the sampling rate for that strata.

This paper proposes to tabulate for public release, the answers given to these survey

questions by state and metropolitan area. So letting j index industry, s index class size, r index

geographic regions (states or metropolitan areas or counties), and n index individual

manufacturing establishments within a js strata, a simple way of representing that statistics this

project aims to produce is:[2]

(1) $N_{Yr} = \Sigma_{jsn} w_{js} \, y_{jsn} \cdot \mathbf{1}\{R_{jsn} = r\}$

…where w_{js} is the sample weight for the industry j, size s establishments (hereafter referred to as

"strata js" establishments); y_{jsn} is a dummy variable that is 1 if establishment n in strata js reports

using some technology (or answers some other survey question affirmatively) and 0 otherwise;

and $\mathbf{1}\{R_{jsn} = r\}$ indicates whether or not establishment n in strata js is in region r. N_{Yr} therefore

represents an estimate of the total number of establishments in region r meeting the $y_{jsn} = 1$

criteria (such as the use of a particular technology). Also reported will be the prevalence of such

establishments, estimated by dividing by an estimate of the total number of establishments in

region r:

(2) $P_{Yr} = N_{Yr}/N_r$

(3) $N_r = \Sigma_{jsn} w_{js} \mathbf{1}\{R_{jsn} = r\}$

---

[2] Modified from US Bureau of the Census (1991), p. C-1.

Assessing Geographic Coverage

The SMT was a random sample, but because it was not a geographically stratified a legitimate concern is that particular regions by chance have more or fewer establishments than would be representative of the population of establishments in that region. Using outside data sources – namely the Censuses of Manufacturers, and County Business Patterns data – it will be possible to assess the magnitude of this problem, and if necessary adjust the weights to be more regionally representative.

The way this will work is as follows. Using the SMT data, an estimate of a count of establishments in region *r* and strata *js* is:

$$N_{jsr} = \Sigma_n w_{js} \boldsymbol{I}\{R_{jsn} = r\}$$

For the purpose of the 1988 or 1993 SMT, this can be compared to counts of establishments in region *r* and strata *js* from the 1987 and 1992 Censuses of Manufacturers, respectively, denoted N_{jsr}^{CM}; each of the SMT counts can also be compared to and County Business Patterns establishment count data for the same year, denoted N_{jsr}^{CBP}. By chance, we expect that some of the SMT estimated N_{jsr} will differ substantially from the counts in the universe data, but examination of the cells will allow a determination of whether there are any systematic regional or other patterns to the cells with poorly estimated counts. (Discrepancies will also occur because of the overall nonresponse rate, which can be adjusted for – see note 3.) In general, the problem could be addressed by multiplying the terms inside the sum of (1) and (3) by a factor that adjusts for the discrepancy in establishment counts in cell *jsr:*

(4) $N_{Yr}' = \Sigma_{jsn} w_{js} \, y_{jsn} \cdot \mathbf{1}\{R_{jsn} = r\} \cdot a_{jsr}$

(5) $P_{Yr}' = N_{Yr}'/N_r'$

(6) $N_r' = \Sigma_{jsn} w_{js} \mathbf{1}\{R_{jsn} = r\} \cdot a_{jsr}$

…where $a_{jsr} = \dfrac{N_{jsr}^{CM}}{N_{jsr}}$ or $\dfrac{N_{jsr}^{CBP}}{N_{jsr}}$.[3] Another way of writing this is by defining "regionally representative" weights, which are the product of the original strata weights and the regional adjustment factor, i.e. $\widetilde{w}_{jsr} \equiv w_{js} a_{jsr}$. Using this definition, (4) – (6) could be rewritten as:

$N_{Yr}' = \Sigma_{jsn} \widetilde{w}_{jsr} \, y_{jsn} \cdot \mathbf{1}\{R_{jsn} = r\}$

$P_{Yr}' = N_{Yr}'/N_r'$

$N_r' = \Sigma_{jsn} \widetilde{w}_{jsr} \, \mathbf{1}\{R_{jsn} = r\}$

The construction of these regionally representative weights $\widetilde{w}_{jsr}$ is one of the benefits of this project, and it will be described again in the benefits section below. To repeat, the idea again is that because the SMT was not a stratified sample, by chance some regions' establishments may be underrepresented or overrepresented in the survey; using the data on the universe of establishments I create weights that are regionally representative.[4] Adding these regionally

---

[3] This can also be multiplied by a factor to adjust for the overall nonresponse rates in each *js* strata, which if necessary can be estimated by $\dfrac{N_{js}}{N_{js}^{CM}}$ or $\dfrac{N_{js}}{N_{js}^{CBP}}$. (Note that the "r" subscript is dropped for the *overall* nonresponse rate). The SMT documentation, however appears to indicate that the sample weights w_{js} may already be adjusted for the overall nonresponse rate for each strata, or at least the fraction of firms not responding is recorded (US Bureau of the Census (1991)).

[4] A simple example: if, say there were 400 small (20-99 employees) textile plants in the US and these were sampled at a rate of 1 in 4 for the SMT, then we would see 100 small textile plants in the survey, and the strata weight for the

adjusted weights to the SMT micro data will allow future users of the data to construct

representative statistics for arbitrary regions that are comprised of counties and/or states.

A related set of problems is that N_{jsr} may sometimes be zero (which is a problem only

when N_{jsr}^{CM} or N_{jsr}^{CBP} is not zero) or the data may simply be nonrepresentative in other

unobserved ways for some regions. This is a problem about which little can be done, though the

standard errors will give at least some idea of the reliability of particular estimates. Treating

these regional weights as constants, the variance of the adjusted cell counts is approximately:[5]

$$V(N_{Yr}') = n_{Yr}P_{Yr}(1-P_{Yr}) \, \widetilde{w}_{jsr} \, (\, \widetilde{w}_{jsr} -1)$$

The variance of the prevalence is this expression divided by the squared number of

establishments in the region. In keeping with the practice of Census Bureau reports, the relative

standard error, computed by taking the square root of this expression and dividing by the level,

will be reported.

Cross-Tabulations

If it is feasible (given the sample sizes) it will also be useful to tabulate the answers to

survey questions by region *and* by other plant characteristics. Results shown in Appendix 2

---

strata "j = textiles, s = 20 to 99 employees" would be 4. Suppose now that Miami had 24 such small textile plants in total. In order to be representative of Miami, the survey would need to have 6 of Miami's small textile plants, but nothing guarantees this will be the case. If instead only 3 of Miami's plants were in the survey, for example, then Miami's small textiles would be underrepresented by a factor of 2, and so we would need to weight Miami's small textile plants at double the overall strata weight in order for the Miami observations to be represented in proportion to their presence in the population. We would therefore create a small textile weight for Miami of 4 x 2 = 8. Similarly, if Miami's small textile establishments happened to be overrepresented in the survey, we would give them a smaller weight.

[5] Cochran (1977) discusses finite population sample techniques that could in principle produce a standard error formula that does not rely on the assumption of fixed weights.

suggest that cross-tabulations by two-digit industry are likely to be feasible. Cross-tabulations can be added to the formulas above by adding an additional indicator function for j ∈ {some two digit industry} to both the numerator and denominator of the prevalence formulas.

Cross-tabulations by employment class sizes would also be useful, especially in light of evidence presented in Dunne (1994) that establishment size is a major determinant of technology use. Results in Appendix 2 suggest that this occasionally runs into small cell sizes for very large establishments. An alternative way of presenting the data that would avoid this problem, and which would be useful in particular to researchers interested in the impact of technology on the typical *worker* (rather than establishment) would be to create prevalence estimates that were *employment* weighted, rather than establishment weighted. In the 1988 and 1993 data, it may be possible to do this by matching the establishments by their id's to the exact employment counts in the 1987 and 1992 Censuses of Manufacturers. Letting E_{jsn} be the employment for establishment *n* in strata *js*, the (adjusted) employment-weighted prevalence is:

(7) $P_{Yr}{}^e = N_{Yr}{}^e / N_r{}^e$

$N_{Yr}{}^e = \Sigma_{jsn} w_{js} \, y_{jsn} \cdot \mathbf{1}\{R_{jsn} = r\} \cdot a_{jsr} \cdot E_{jsn}$

$N_r{}^e = \Sigma_{jsn} w_{js} \mathbf{1}\{R_{jsn} = r\} \cdot a_{jsr} \cdot E_{jsn}$

If this approach is not possible (and it is not in the 1991 survey) one could use the cell-level (strata-county or strata-state) employment counts from the county business patterns data, E_{jsr}^{CBP}, in place of the establishment level counts. This should also give an unbiased estimate of the employment weighted prevalences. Again, regionally representative employment weights would be constructed for the micro data users of the SMT, defined by $\widetilde{w}_{jsr}^e \equiv w_{js} \cdot a_{jsr} \cdot E_{jsn}$.

Other plant characteristics, which would be desirable and may be feasible for cross-tabulations include union contract status (of the establishment), military contactor status, foreign-ownership status, plant age (in categories),[6] exporter status, and value of shipments (in categories).

IV. Do Local Work Force Skills Determine the Pace of Technical Change?

In addition to tabulation of regional differences in the pace of technological change, this project aims to try to explain regional differences in manufacturing technology use with regional differences in the skill composition of the work force. Previous evidence of this includes the fact that personal computers were implemented most rapidly during the 1980s and 1990s in metropolitan areas with a growing relative supply of educated labor (Lewis (2002)) and that the average education of the workers in the local work outside a manufacturing plant force appear to raise the productivity of that plant (Moretti (2002)).

Public use data will be used to construct measures of work force skills by labor market, defined here as metropolitan area. The initial approach taken will be to estimate the cross-sectional impact of local work force characteristics on technology use measured in the 1988 and 1991 surveys, which are quite different in the types of technology questions they ask. For both sets of these regressions, 1990 Census of Population 5% PUMS data will be used to construct measures of local work force composition, because they are richer in geographic detail than contemporaneous data, such as the CPS.

Technology use rates, the dependent variable in these regressions, will be measured at the industry by metropolitan area level, which will allow unobserved factors at the city and industry

---

[6] Though it is worth noting that Dunne (1994) finds no evidence of any plant age effects on technology use.

level affecting technology use to be controlled for using fixed effects. Thus, regressions will be of the form:[7]

$$P_{Yjr}{}^e = \alpha_j + \beta_r + \delta_y{}'X_r + \varepsilon_{jr}$$

…where $P_{Yjr}{}^e$ represents the (adjusted) employment-weighted prevalence of technology y in industry j and metropolitan area r (see previous section for definition); α_j and β_r are industry and metro area fixed effects, respectively; X_r is a vector of city-r work force characteristics (such as the fraction of workers with a college degree). δ_y, the coefficient vector of interest, measures the impact of local work force characteristics on the use rate of technology y.

In theory, it is also possible to do this type of regression in first differences, which asks the question whether *changes* in work force characteristics affect the rate of adoption of technology. Because the 1988 and 1993 SMTs ask technology questions a similar fashion, one approach to first differences will be to use the 1988 and 1993 Current Population Surveys (CPS) to estimate changes in work force characteristics by metropolitan area between those two years, allowing estimation of regressions of the form:

$$\Delta_{88\text{-}93}P_{Yjr}{}^e = \alpha_j{}^* + \beta_r{}^* + \delta_y{}^*{}'\Delta_{88\text{-}93}X_r + \varepsilon_{jr}{}^*$$

…where $\Delta_{88\text{-}93}$ represents an operator for the change between 1988 and 1993. The problem with doing this is that there are a limited number of metropolitan areas that can be observed in any

---

[7] Note that because this is a regression, the fact that some data points will be generated by small cell sizes does not pose a disclosure risk (Merrell and Reznek (2002)), though as a practical matter it may be necessary to use two-digit industry cells rather than three-digit. Only the slope coefficient, and not the fixed effects, will be reported.

CPS. So another approach that will be taken is to regress the 1988 or 1991 levels of technology use on the 1980-1990 changes in work force composition by metropolitan area:

$$P_{Yjr}{}^e = \alpha_j{}^{**} + \beta_r{}^{**} + \delta_y{}^{**}{}'\Delta_{80\text{-}90}X_r + \varepsilon_{jr}{}^{**}$$

…where $\Delta_{80\text{-}90}$ represents an operator for the change between 1980 and 1990. Under the assumption that use of these technologies was limited in 1980, $P_{Yjr}{}^e \approx \Delta_{80\text{-}90}P_{Yjr}{}^e$ giving $\delta_y{}^{**}$ a similar interpretation $\delta_y{}^*$ of the previous regression.[8]

Another set of regressions would take advantage of the survey questions that ask establishments *when* a particular technology was implemented, or if it *intends* to implement a technology in the near future. Using the prevalence of these survey responses as alternative dependent variables will allow some sense of the timing of the response to work force composition.

Identification

The slope estimates from the regressions presented so far may have interpretations other than the causal effect of work force composition on technology use. For example, implementation of advanced technologies may raise demand for skilled labor and draw such labor to markets where the technology is being implemented. To account for the first problem, instruments for workforce characteristics will be used, derived from the strong impact that foreign immigration tends to have on the work force characteristics of different local labor markets in the US and the historical tendency of immigrants from different parts of the world to

---

[8] To the extent that the technologies were in use in 1980, and the use in particular markets is positively correlated over time, estimates of $\delta_y{}^{**}$ will be biased towards 0.

settle into particular US labor markets. The instrument, developed by Card (2001), can be expressed as:

$$(8) \ Z_r = \frac{\sum_g \frac{F_{gr}^{70}}{F_g^{70}} z_g}{P_r}$$

…where $\frac{F_{gr}^{70}}{F_g^{70}}$ represents the fraction of all foreign-born residents from country g living in city r in 1970 (measured with 1970 PUMS data), z_g represents a vector of skill categories of immigrant arrivals from country g for a relevant recent period, and P_r represents the population of city r in the present period. z_g might, for example, be a vector representing the 1980-1990 arrivals of Mexican high school graduates, Mexican some college-educated and Mexican college graduates; $\frac{F_{gr}^{70}}{F_g^{70}} z_g$ therefore assigns these three groups of Mexicans to cities r in the same proportions as the settlement patterns of Mexicans living in the US in 1970. Summing over countries produces an estimate of the expected impact of recent foreign immigration on the supply of different skill groups in each city, based on 1970, rather than present-day, settlement patterns. An assumption for the validity of the instrument is that these 1970 settlement patterns are on average orthogonal to unobserved shocks that might effect technology adoption in the present day. The skills of recent immigrants will be measured in the 1990 Census of Population or in the 1994 Current Population Surveys depending on the regression being run.

A second issue confounding the regressions presented so far is it may actually be an *individual*-level relationship between education and technology use that drives the relationship,

rather than a market level relationship. One way make sure this is not the case is to measure the vector of city-level work force skills, X_r, only for those *outside* the industry; i.e. replace X_r with X_{-jr}, the vector of worker characteristics outside industry *j*. This is similar to the approach taken by Moretti (2002), and should give a lower bound estimate of the effect of local work force skills on technology use.

Channels By Which Education Affects Technology

   The SMTs also ask questions about the reasons why technology was implemented or not implemented (such as "improve product quality" for why implemented and "cost of training workers" for why not implemented). If work force skills affect the probability of implementing technology, then controlling for the prevalence of the various reasons in the regressions will allow a determination of the channels through which education affects the probability of technology adoption. If the effect gets loaded onto these variables, then it suggests education operates through them; otherwise it suggests education affects technology adoption by other means.

Productivity Impact

   One motivation for asking how the skills of the work force affects technology use is wage evidence that the demand for workers in a local labor market is affected by the relative supplies of different skill groups. Moretti (2003) shows that increases in the relative supply of college educated in a metropolitan area raise the wages of all workers in the area. And in another paper (Moretti (2002)), the same author shows using Census of Manufacturers data that the

productivity of manufacturing plant increases in the average education of workers in the local work force *outside* the plant.

Merging the SMT data to the Censuses of Manufacturers provides an opportunity to extend this previous research by potentially describing a technological mechanism by which education levels of workers in a local market of workers affects the productivity and wages of other local workers. Adding dummies for the use of different advanced technologies to the production function similar to the one specified in Moretti (2002) allows the skills of the local work force to affect productivity both directly and through a higher prevalence of these technologies. The importance of observed technological shifts on productivity can be assessed by estimating production functions of the form: [9]

$$ln \ Q_{jsrn} \ = \ \alpha_1 ln \ L^P_{jsrn} \ + \ \alpha_2 ln \ L^{NP}_{jsrn} \ + \ \beta ln \ K_{jsrn} \ + \ \phi_j \ + \ \phi_r \ + \ \gamma X_{-jr} \ + \ \Sigma_y \theta_y T_{yjsrn}$$
$$+ \ \xi_{jsrn}$$

…where Q_{jsrn} is the output (value added) at plant *n* in strata js (which is region *r*). This is a function of the amount of production workers (L^P), non-production workers (L^{NP}) and capital (K); industry (ϕ_j) and city (ϕ_r) fixed effects; a vector of characteristics of other workers in the labor market outside the plant (X_{-jr} which is actually measured outside the *industry* of that plant, since detailed worker characteristics are only observed at the industry level) and dummies for a vector of dummies for different technologies (T) in the SMT data (likely the four broad categories of technology – see Appendix 1), which are indexed by *y*.[10] Estimation with and without the technology dummies will allow an assessment of whether local work force

---

[9] The regression will be weighted by the regional employment weights, defined above.
[10] It is also common in the use of the SMT data to measure technological intensity of a plant with the number of technologies in use, and this approach may also be taken.

characteristics affect manufacturing productivity through adoption of these manufacturing technologies, or by other means. To the extent a more skilled work force raises manufacturing productivity through greater adoption of skill complementary technology the coefficient γ will be driven to 0 when the technology controls are included.

Endogeneity is also an issue with these plant-level estimates: there may be unobserved heterogeneity (modeled as ξ_{jsrn}, which could include things such as "management quality") which affects both the level of inputs used in some city-industry's manufacturing plants and the level of output at those plants. It is unfortunately difficult to conceive of instruments for plant-level inputs. One change that might reduce bias would be to control for city-industry fixed effects. The feasibility of this can only be determined by looking at the micro data (but is worth noting that Moretti (2002) suggested that it was unfeasible). Another approach could potentially reduce bias would be to control for plant level fixed effects by matching plants across surveys (the 1988 and 1993 surveys) by id. This is the approach taken by McGuckin et al. (1998). The main drawbacks of this approach are a reduction in the sample size, and an increase in error in the measurement of inputs. A second approach to first differencing would be to impute past data for all the plants in the 1993 survey. Retrospective questions on the when the technology was implemented, available in the 1993 survey, could be used to construct estimates of technology use at the same plant 5 years ago; and growth rates of other inputs in a given strata-region cell (jsr) could be used to impute past levels (albeit noisily measured) of the other inputs.

V. Benefits to the Bureau

Each part of this project constitutes a benefit to the Census Bureau under Title 13, Chapter 5. The details of the benefits are given in this proposal's PPS. See the abstract to this proposal for a brief description.

VI. Conclusion

This document has proposed using the micro data in the Surveys of Manufacturing Technology to tabulate the use, plans for use, and reasons for use of different manufacturing technologies by state and metropolitan areas, and proposes releasing the tabulated statistics to the public to the extent it is feasible. Existing evidence indicates that technology use may differ substantially by region, so having publicly available statistics that measure this would be of interest to decision-makers and researchers. The SMT appears to be a large enough survey to feasibly generate regional statistics that do not risk disclosure of individual establishments' confidential survey responses.[11] Whether or not it will be also possible to further cross tabulate the survey responses by plant characteristics, another proposed goal, can only be determined by examination of the micro data, but tabulations by plant size and two-digit industry appear at least to be possible. In addition to these public statistics, two sets of sample weights that will allow future users of the SMT to construct statistics that are representative of the establishments or workers in arbitrary US regions (comprised of counties or states) will be constructed and added to the SMT micro data.

A secondary purpose of this project is to determine if any regional differences in manufacturing technology use can be causally attributed to regional differences in the skills of the local work force: either directly, or indirectly through the effect of local work force skills on

---

[11] See Appendix 2 for evidence of this.

the costs or benefits of implementing new technologies. The effect of this on productivity will also be examined. Other evidence suggests that technologies used by industries respond to changes in local work force composition. The level of detail in the SMT is suited to give a stronger answer to the question of how local work force skills affect industries.

Project Details

This paper proposes access to the following confidential Census Bureau datasets:

- *Surveys of Manufacturing Technology* (SMT) in years 1988, 1991, and 1993, all class sizes and industries. Variables being requested include the answers to all survey responses, and industry, employment class size, state, county, sample weight. An establishment identifier that can be matched to the Census of Manufacturing and across surveys of manufacturing technology (such as Census establishment ID or permanent plant number) is also requested.
- *Census of Manufacturers* (CM) in years 1987 and 1992, all class sizes and industries. Variables being requested include total employment, employment of production and nonproduction workers, value added, book value of capital (equipment and structures), and SIC industry. Also being requested is an establishment identifier that will allow a link to the 1988 and 1993 SMTs by establishment.

The 1987 CM will be matched to the 1988 SMT and the 1992 CM will be matched to the 1993 SMT by establishment identifier. This match will allow two things. First, it will allow a comparison of the number of establishments in each industry-class-size-region cell in the 1988 and 1993 SMT to the corresponding establishment counts in the universe of manufacturing establishments covered by the 1987 and 1992 CMs. This serves the purpose of assessing the geographic coverage of the SMT (see proposal for details). Second, it will allow the estimation of plant-level production functions including more detailed technology indicators from the SMT.

Other data used by this project include

- *County Business Patterns* county and state summary data from 1988, 1991 and 1993. These will be matched to the corresponding SMT by industry, class size and region (county or state). This match will allow another comparison of the number of establishments in each industry-class-size-region cell in the SMT to the same cell in the County Business Patterns data, allowing another assessment of the geographic coverage of the SMT. It will also allow a determination of the average number of workers in each industry-class size-region cell, which can be used to create employment-weighted frequencies of technology use.
- *1970, 1980 1990 PUMS.* Worker characteristics in PUMS data will be aggregated to the metropolitan area level. The 1970 data are useful for an instrument. (See equation (8).) The impact changes in worker characteristics between 1980 and 1990 on technology use will be measured.
- In addition, worker characteristics at the metropolitan area level will be measured in the 1988, 1993 and 1994 current population surveys. Changes in technology use between 1988 and 1993 will be regressed on changes in work force characteristics between these two years. The 1993 and 1994 surveys will be important for measuring the characteristics of recent immigrants. (See equation (8).)

Project Duration

24 months of 40 hour a week access to the CCRDC should be more than sufficient to complete this project. I have experience working with related datasets (the LRD), which should allow me to get to work quickly.

References

Acemoglu, Daron (1996). "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *The Quarterly Journal of Economics*, Vol. 111, No. 3. (Aug., 1996), pp. 779-804

---- (1998). "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics* 113(4): November 1998, p. 1055-89.

---- (2002). "Technical Change, Inequality and the Labor Market," *Journal of Economic Literature* 40(1): March 2002, p. 7-72.

Autor, David H., Lawrence F. Katz and Alan B. Krueger (1998). "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics* 63(4): November 1998, p. 1169 – 1213.

Bound, John and George Johnson (1992). "Changes in the Structure of Wages in the 1980s: An Evaluation of Alternative Explanations." *American Economic Review* 82 (3): June 1992, p. 371-392.

Card, David (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19(1): January 2001, p. 22-64.

Cochran, William G. (1977). *Sampling Techniques.* New York: John Wiley & Sons, 1977.

Dahl, Gordon B. (2002). "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70(6): November 2002.

Davis, Steven J. and John Haltiwanger (1991). "Wage Dispersion between and within U.S. Manufacturing Plants, 1963-86." *Brookings Papers on Economic Activity. Microeconomics*, Vol. 1991. (1991), pp. 115-180.

---- (1996). "Employer Size and the Wage Structure in US Manufacturing." *Annales d'Economie et de Statistique* 41/42: 1996, p. 323 -367.

Davis, Steven J. and John Haltiwanger and Scott Schuh (1991). "Published Versus Sample Statistics From The ASM: Implications For The LRD." *Bureau of the Census Center for Economic Studies Discussion Paper 91-1*: January 1991.

Doms, Mark, Timothy Dunne and Kenneth R. Troske (1997). "Workers, Wages and Technology." *Quarterly Journal of Economics* 62(1): February 1997, p. 253 – 290.

Dunne, Timothy (1994). "Plant Age and Technology Use in US Manufacturing Industries." *RAND Journal of Economics* 25(3): Autumn 1994, p. 488 – 499.

Dunne, Timothy, Lucia Foster, John Haltiwanger and Kenneth R. Troske (2002). "Wage and Productivity Dispersion in US Manufacturing: The Role of Computer Investment." *IZA Discussion Paper #563*: August 2002.

Lewis, Ethan (2002). "Local, Open Economies Within The US: How Do Industries Respond to Immigration?" Mimeo. UC Berkeley, December 2002. (Currently available at http://socrates.berkeley.edu/~gatewood.)

McGuckin, Robert H. Mary L. Streitwieser and Mark Doms (1998). "The Effect of Technology Use on Productivity Growth." *Economics of Innovation and New Technology* 7(1): 1998, p. 1-26.

Merrell, David R. and Arnold P. Reznek (2002). "On Disclosure Protection for Non-Traditional Statistical Outputs." Mimeo. US Bureau of the Census, Center for Economic Studies, 2002.

Moretti, Enrico (2002). "Human Capital Spillovers in Manufacturing: Evidence from Plant-Level Production Functions." *NBER Working Paper #9316* , 2002.

---- (2003). "Estimating the Social Return to Higher Education: Evidence From Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics*, forthcoming 2003. (Currently available at http://www.econ.ucla.edu/moretti.)

Stiroh, Kevin J (2002). "Information Technology and the US Productivity Revival: What Do the Industry Data Say?" *American Economic Review* 92(5): December 2002, p. 1559 – 1576.

U.S. Dept. of Commerce, Bureau of the Census (1989). *Manufacturing Technology 1988*. SMT(88)-1. Washington, DC: US Government Printing Office, 1989.

---- (1993). *Manufacturing Technology: Factors Affecting Adoption 1991*. SMT(91)-2. Washington, DC: US Government Printing Office, 1993.

---- (1994). *Manufacturing Technology: Prevalence and Plans for Use 1993*. SMT(93)-3. Washington, DC: US Government Printing Office, 1994.

---- (1995). *County Business Patterns, 1992*. United States: *U.S. Summary, State, and County Data* [Computer file]. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census [producer], 1994. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1995.

---- (2002). "US Bureau of the Census Center for Economic Studies Research Data Centers Handbook for Researchers." Washington, DC: Center for Economic Studies, Mimeo, January 2002.

Appendix 1
Description of Technologies Covered in Surveys of Manufacturing Technology (1988, 1991, 1993)

I.   **Design and Engineering**

1. *Computer-Aided Design (CAD) and/or Computer-Aided Engineering* – "Use of computers for drawing and designing parts or products and for analysis and testing of designed parts or products."
2. *Computer-Aided Design (CAD)/Computer-Aided Manufacturing (CAM)* – "Use of CAD output for controlling machines used to manufacture the part or product."
3. *Digital Data Representation* – "Use of digital representation of CAD output for controlling machines used to manufacture part or product."

II.   **Fabrication and Machining**

4. *NC/CNC Machine* – "A single machine either numerically controlled (NC) or computer numerically controlled (CNC) with or without automated material handling capabilities. NC machines are controlled by numerical commands, punched on paper or plastic mylar tape while CNC machines are controlled electronically through a computer reading in the machine."
5. *Flexible Manufacturing Cell (FMC)* – "Two or more machines with automated material handling capabilities controlled by computers or programmable controllers, capable of single path acceptance of raw material and single path delivery of finished product."
6. *Flexible Manufacturing System (FMS)* – "Two or more machines with automated material handling capabilities controlled by computers or programmable controllers, capable of multiple path acceptance of raw material and multiple path delivery of finished product. A FMS may also be comprised of two or more FMC linked in series or parallel."
7. *Materials Working Laser* – "Laser technology used for welding, cutting, treating, scribing and marking."
8. *Pick and Place Robots* – "A simple robot, with one, two, or three degrees of freedom, which transfers items from place to place by means of point-to-point moves. Little or no trajectory control is available."
9. *Other Robots* – "A reprogrammable, multifunctional manipulator designed to move materials, parts, tools or specialized devices through variable programmed motions for the performance of a variety of tasks."

III.   **Materials Handling**

10. *Automated Storage and Retrieval Systems (AR/RS)* – "Computer controlled equipment providing for the automatic handling and storage of materials, parts, subassemblies, or finished products."
11. *Automatic Guided Vehicle Systems (AGVS)* – "Vehicles equipped with automatic guidance devices programmed to follow a path that interfaces with work stations for automated or manual loading and unloading of materials, tools, parts or products."

IV. **Inspection and Quality Control**

A. <u>Automated Sensor Based Inspection And/Or Testing Equipment</u> – Includes automated sensor based inspection and/or testing performed on incoming or in-process materials, or performed on the final product.

      12. - *Performed on Incoming or In-Process Materials*
         - *Performed on Final Product*

B. <u>Communications and Control</u>

      13. *Technical Data Network* – "Use of local area network (LAN) technology to exchange technical data with design and engineering documents."

      14. *Factory Network* – "Use of local area network (LAN) technology to link information between different points on the factory floor."

      15. *Intercompany Computer Network* – "Use of network technology to link subcontractors, suppliers and/or customers with the plant."

      16. *Programmable Controllers* – "A solid state industrial control devise that has programmable memory for storage of instructions, which performs functions equivalent to a relay panel or wired solid state logic control system."

      17. *Computers Used for Control on the Factory Floor* – "Exclude computers imbedded within machines, or computers used solely for data acquisitions or monitoring. Include computers that may be dedicated to control but are capable of being programmed for other functions."

Appendix 2
Expected Sample Sizes in Survey of Manufacturing Technology State and Metropolitan Area Tabulations

I. Introduction

The purpose of this appendix is to demonstrate that the proposed geographic tabulations are unlikely to risk disclosure of individual SMT responses. The CES guideline for protecting against disclosure risk when tabulating SMT variables for release to the public requires that the number of firms involved in the computation of each table cell exceed some fixed minimum (U.S. Dept. of Commerce (2002)). The challenge determining whether this minimum count can be met is both the count and the SMT data are confidential. The approach taken here is to use publicly available information – primarily County Business Patterns (CBP) summary data – to estimate the expected number of firms from the SMT sample that will be involved in the proposed tabulations. These expected counts can be reviewed by CES to determine if they will meet the (confidential) minimum firm count.

Note that these are *estimates* of firm counts – actual firm counts in the SMT may differ from the numbers presented here. And it will be the actual number of firms that will determine which statistics can be presented to the public without the risk of disclosure.

II. Obtaining Estimated Firm Counts

The SMT for 1993 included a stratified random sample of 8,432 of the 43,551 manufacturing establishments believed to meet the survey's sample selection criteria (U.S. Department of Commerce (1994)).[12] Larger numbers of establishments were involved in the

---

[12] In order to be included in the survey, a manufacturing establishment had to have at least 20 employees and be classified in SIC 34-38. As will be noted again below, similar numbers of establishments were involved in the 1991 and 1993 SMTs also proposed for use by this project.

1991 and 1988 surveys. The sampling rate was approximately the same in the different strata of each survey.[13]

How can we determine how many of these observations will be in a particular state or metropolitan area? Because the SMT is a random sample, an unbiased estimate can be obtained by multiplying the number of establishments from the *universe* of SMT-eligible firms – which is observable in publicly available data – with the average SMT sampling rate. To see this, let N_r represent the number of SMT-eligible establishments in region r and n_r represent the number of actual number of SMT respondents in region r. Under random sampling,

(1) $E[n_r / N_r] = n/N \equiv \varsigma$

where n and N are the overall number of SMT respondents (8,432) and SMT-eligible establishments (43,551) respectively. This equation says we expect the rate at which establishments were sampled from the overall population to be, on average, the same in each region. An unbiased estimate of n_r is therefore ςN_r, where ς is the average SMT sampling rate ($\frac{n}{N} \approx 0.20$).

The sampling unit in the SMT is the manufacturing establishment. Most manufacturing establishments are single entities; nevertheless, for the purpose of meeting CES confidentiality protections, the count of *firms* rather than establishments is important. So suppose F represents the number of firms with an SMT eligible establishment, the typical SMT establishment will then represent $\theta \equiv F/N$ firms. Given this, an unbiased estimate of the number of SMT firms in region r is:

(2) $\hat{f}_r = \theta \varsigma N_r$

---

[13] The strata were formed by crossing three employment class sizes (20-99, 100-499, >500 employees) with each 3-digit SIC industry (U.S. Dept. of Commerce (1989)). The documentation reports that in certain smaller strata all establishments were surveyed.

As the number of firms in the SMT or the SMT universe does not appear to be reported in any of the SMT reports, θ must be estimated from another source. The number of firms with an SMT-eligible establishment can be approximately determined from the Census Bureau's "Statistics of U.S. Businesses," which report County Business Patterns-derived tabulated firm counts. In the 1992 tables, the number of firms in SIC 34-38 with at least 20 employees is 36,218; this probably overstates the number of firms with an SMT-eligible establishment.[14] The number of establishments in the SMT-eligible universe, obtained from County Business Patterns summary files for the same year, is 43,968. These numbers suggest a θ of 0.82 ($= \dfrac{36,218}{43,968}$).

II. Results for States

Column 1 of Table 1 presents estimated SMT firm counts, $\hat{f}_r$, by census division, region, and state by applying equation (2). The total number of SMT-eligible establishments underlying these counts (i.e. N_r) comes from 1993 CBP summary data. (1993 was chosen because it had the fewest number of SMT observations; the results should be considered low estimates of the firm counts in the 1991 and 1988 surveys.) These are multiplied by the average SMT sampling rate for 1993, $\varsigma \equiv 0.20$, and a conservative value for $\theta \equiv 0.75$ (each 4 SMT establishments are assumed, on average, to be owned by 3 firms).

According to these estimates, we should expect the 1993 SMT to contain about 6,500 separate firms. The numbers in the first column of Table 1 show substantial numbers of these firms are likely to be in each state, driven by the fact that manufacturing establishments in SMT industries are present in all states. Smaller states, for example Wyoming and Montana, may not

---

[14] The desired count is of firms that have at least one *establishment* with at least 20 employees. Firms in with greater than 20 employees may in some cases contain more than one establishment, each of which has fewer than 20 employees, making this firm count potentially overstated.

have sufficient numbers of firms in the SMT to meet the reporting criteria; it may be necessary, therefore to not report results for these states. In order to report regional totals, it may also be necessary to exclude some states that individually meet the reporting criteria to avoid complementary disclosure.[15]

Cross-Tabulations

The original proposal also suggested making cross-tabulations of technology use by firm characteristics. For the firm characteristics available in the CBP, it is possible to estimate counts of firms by state *and* characteristics. Letting j index the values of some firm characteristic desirable for cross-tabulation, an estimate of the firms in the SMT table cell jr is simply:

(3) $\hat{f}_{jr} = \theta \varsigma N_{jr}$

where N_{jr} is obtained from CBP data, and ς and θ are as before. Columns (2) – (4) of Table 1 present estimates of firm counts by state and establishment class size (20-99 employees, 100-499 employees, >500 employees). For the most part it appears we can expect substantial numbers of firms in each state for each class size. However, the largest class size, greater than 500 employees, has few firms, making separate counts by state somewhat problematic.[16] Limiting the geographic disaggregation for this class size to region is one approach to addressing this problem. Another alternative would be to break the class size categories at a different point (e.g. 20-99, 100-249, 250+). Lumping the 500+ category with the 100-499 would be less desirable, as Dunne (1994) has shown that plant size is a important determinant of technology use.

---

[15] Avoiding a complementary disclosure in this case would require that the unreported states for each census region *in total* have a sufficient firm count to meet CES reporting criteria.

[16] Since larger establishments are more likely to be a part a multi-unit firm than small ones, even these conservative counts may be overstated (i.e. θ should probably be lower for this class size).

Table 2 presents expected firm counts by state and the five two-digit SIC industries included in the SMT survey. Like in Table 1, reporting means for some smaller states appears to be problematic, but for the most part we should expect firms to be spread across states enough that it will be possible to report statistics by state for medium and large states.

In addition to SIC and class size, it would also be desirable to produce estimates of firm counts by other important plant characteristics (such as plant age) and state. Unfortunately, publicly available data do not allow this. However, there is no a priori reason to expect that other plant characteristics are not also distributed across states in similar proportions to industry or class size, so Table 1 and 2 are informative to the extent that they show the SMT sample size is likely sufficient to develop regional statistics on any major plant characteristic.

III. Results for Metropolitan Areas

Lewis (2003) also proposes tabulations by metropolitan area, at least for major metropolitan areas. So Tables 3 and 4 are similar to Tables 1 and 2, except estimated firm counts are shown for each of 30 large metro areas, ordered by the total number of SIC-eligible establishments in the metro area, ranging from 294 (Los Angeles) down to 43 (Charlotte, NC). This number of firms seems likely sufficient to not risk disclosure. Cross-tabulations may be more problematic. Note, as before, there are probably very few firms covering the largest category of establishment (>500 employees), so a separate breakdown of technology use for this very largest category may again not be possible. The counts by industry also appear to be thin in some places, so it may be necessary to suppress certain cells that do not meet minimum firm counts.

Complementary Disclosure

An additional risk in reporting statistics both by state and metropolitan area is of complementary disclosure. If non-metropolitan parts of a state tend to have very few firms, then by reporting both state and metropolitan area statistics, one may risk disclosure in such non-metropolitan areas. To insure that this will not be a problem, Tables 5 and 6 report estimated firm counts, by state, for the part of the state not in one of the metropolitan areas listed in Tables 3 and 4. Column 1 of Table 5 shows that most states have substantial numbers of firms outside these 30 metropolitan areas. Broken down by class size, again the largest class size is thin, but that was true of the metro area numbers as well.

Table 6 breaks apart the non-metro area firm counts by state and industry. Non-metropolitan Arizona, Washington and Oregon stand out as having few firms in some cells. The small number of firms outside the city may mean that it will not be possible to report statistics for the Portland, Seattle and Phoenix metropolitan areas (at least not by industry). In addition, SIC 37 seems to have very few firms in some other cells, so suppressing statistics for that industry may also be called for.

IV. Conclusion

This appendix has presented conservative estimates of the number of firms likely to be involved in the tabulations of SMT data by state and metropolitan area as described in this proposal. It has shown that it is likely that there are sufficient numbers of firms to allow presentation of such statistics without risking disclosure of confidential survey responses. Cross-tabulations by industry and class size also seem feasible, especially for larger geographic units.