On-Line Appendix for Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services

Alwyn Young
March 2014

Appendix A: Characteristics of the Model's Equilibrium

This appendix provides the mathematical details behind the assertions made in Section I. All of the proofs are couched in terms of a two sector economy (goods and services). Their extension to the more complex N-sector case is straightforward.

(a): Regardless of the cumulative distribution function describing the paired draw (z_G, z_S), $\xi = (d\bar{z}_i / d\pi_i)(\pi_i / \bar{z}_i) > -1$ [equation (4) in the paper].

Let $G_{j|i}(y|z)$ describe the conditional probability $z_j \leq y$ given that $z_i = z$, i.e. the cumulative distribution function of z_j given z_i, and let $g_{j|i}$ describe the corresponding conditional density and g_i the marginal density of z_i. Then, with $\omega = w_i / w_j$

$$\text{(a1)}\quad \bar{z}_i = \frac{N(\omega)}{\pi_i(\omega)},\ \text{with}\ N = \int_0^\infty z_i g_i(z_i) G_{j|i}(\omega z_i \mid z_i)\,dz_i\ ,\ \pi_i = \int_0^\infty g_i(z_i) G_{j|i}(\omega z_i \mid z_i)\,dz_i$$

$$dN/d\omega = \int_0^\infty z_i^2 g_i(z_i) g_{j|i}(\omega z_i \mid z_i)\,dz_i,\ \text{and}\ d\pi_i/d\omega = \int_0^\infty z_i g_i(z_i) g_{j|i}(\omega z_i \mid z_i)\,dz_i,$$

where, in cases where the domains of g_i and $G_{j|i}$ do not include all positive real numbers I extend them, for the purpose of the integration, by defining g_i and $g_{j|i}$ as equal to zero in the extended region (and similarly for other proofs below). Note that $g_i(z_i)g_{j|i}(\omega z_i|z_i) = g_{i,j}(z_i,\omega z_i)$, the joint distribution of z_i and z_j at the point $(z_i,\omega z_i)$. Assuming this joint distribution has mass along a positive measure of the ray with slope ω from the origin, we have $dN/d\omega > 0$ and $d\pi_i/d\omega > 0$,[1] and it follows that

$$\text{(a2)}\quad \frac{d\bar{z}_i}{d\pi_i}\frac{\pi_i}{\bar{z}_i} = -1 + \frac{\frac{dN}{d\omega}\frac{d\omega}{d\pi_i}}{\bar{z}_i} > -1.$$

As noted in the text, this result is fairly obvious.

For particular distributional forms, it is easy to calculate closed form solutions for ξ illustrating the properties imposed by different distributional assumptions. Thus, for the case where the z_i are independent draws from fréchet distributions with cumulative distribution functions $G_i(z_i) = exp(-(z_i/\lambda_i)^{-\theta})$, we have:

$$\text{(a3)}\quad \pi_i = \int_0^\infty \theta \lambda_i^\theta z_i^{-\theta-1} \exp\left[-(z_i/\lambda_i)^{-\theta}\right] \exp\left[-(w_i z_i / w_j \lambda_j)^{-\theta}\right] dz_i = \frac{(w_i\lambda_i)^\theta}{\sum_i (w_i\lambda_i)^\theta}$$

[1] If it does not, we are at a value of ω where neither π_i nor $\bar{z}_i$ vary with ω, so the derivative of one with respect to the other is not well defined.

and

$$\text{(a4)}\quad \bar{z}_i = \frac{\int_0^\infty \theta \lambda_i^\theta z_i^{-\theta} \exp\left[-z_i^{-\theta}\left(\lambda_i^\theta + (w_j \lambda_j / w_i)^\theta\right)\right] \mathrm{dz_i}}{\pi_i}$$

$$= \frac{\lambda_i^\theta}{\pi_i C^{(\theta-1)/\theta}} \int_\infty^0 - x^{-1/\theta} e^{-x}\ \mathrm{dx} \quad \text{where} \quad C = \lambda_i^\theta + (w_j \lambda_j / w_i)^\theta$$

$$= \frac{\lambda_i (w_i \lambda_i)^{\theta-1}}{\pi_i (w_i^\theta C)^{(\theta-1)/\theta}} \int_0^\infty x^{-1/\theta} e^{-x}\ \mathrm{dx} = \lambda_i \Gamma\left(\frac{\theta-1}{\theta}\right) \pi_i^{-1/\theta}$$

where I have used the substitution $x = z_i^{-\theta}C$ in the second line. Consequently:

$$\text{(a5)}\quad \xi = (d\bar{z}_i / d\pi_i)(\pi_i / \bar{z}_i) = -1/\theta$$

Thus, for independent draws from fréchet distributions ξ is a constant, a function of the distribution's dispersion parameter.

It is not difficult to find distributions with different characteristics. Thus, if z_G and z_S are independent draws from exponential distributions with densities $\lambda_i exp(-z_i \lambda_i)$, allowing $\tilde{\omega} = w_i \lambda_j / w_j \lambda_i$ we have:

$$\text{(a6)}\quad \bar{z}_i = \frac{\int_0^\infty z_i \lambda_i \exp(-\lambda_i z_i)[1 - \exp(-\lambda_i \tilde{\omega} z_i)] \mathrm{dz_i}}{\int_0^\infty \lambda_i \exp(-\lambda_i z_i)[1 - \exp(-\lambda_i \tilde{\omega} z_i)] \mathrm{dz_i}} = \frac{1}{\lambda_i} \frac{\tilde{\omega}+2}{\tilde{\omega}+1}$$

$$\pi_i = \int_0^\infty \lambda_i \exp(-\lambda_i z_i)[1 - \exp(-\lambda_i \tilde{\omega} z_i)] \mathrm{dz_i} = \frac{\tilde{\omega}}{1+\tilde{\omega}}$$

$$\frac{d\bar{z}_i}{d\pi_i} \frac{\pi_i}{\bar{z}_i} = \frac{d\bar{z}_i / d\omega}{d\pi_i / d\omega} \frac{\pi_i}{\bar{z}_i} = -\frac{\tilde{\omega}}{\tilde{\omega}+2}$$

so ξ once varies between 0 and -1 as ω goes from 0 to ∞ or, equivalently, π_i goes from 0 to 1. In this case ξ is a function of sectoral size alone. The log-normal distribution provides an interesting third example. With independent productivity draws with ln means μ_i and (for simplicity) common standard deviation σ, we have:[2]

[2]Proven by applying corollary 2.2b and theorem 2.6 of J. Aitchison and J.A.C. Brown, The LogNormal Distribution with special reference to its uses in economics, Cambridge: Cambridge University Press, 1957.

(a7) $\bar{z}_i = \exp[\mu_i + .5\sigma^2]\dfrac{N\left(\dfrac{\tilde{\omega}}{\sqrt{2}\sigma} + \dfrac{\sigma}{\sqrt{2}}\right)}{N\left(\dfrac{\tilde{\omega}}{\sqrt{2}\sigma}\right)}, \ \pi_i = N\left(\dfrac{\tilde{\omega}}{\sqrt{2}\sigma}\right)$

$$\xi = \exp\left[-\frac{\tilde{\omega}}{\sigma}\frac{\sigma}{2} - \frac{\sigma^2}{4}\right]\frac{N\left(\dfrac{\tilde{\omega}}{\sqrt{2}\sigma}\right)}{N\left(\dfrac{\tilde{\omega}}{\sqrt{2}\sigma} + \dfrac{\sigma}{\sqrt{2}}\right)} - 1$$

where $N()$ is the cumulative standard normal and $\tilde{\omega} = ln(w_i/w_j) + \mu_i - \mu_j$. Holding constant the size of each sector (i.e. $\tilde{\omega}/\sigma$), as σ goes from 0 to ∞ ξ goes from 0 to -1 in both sectors.

(b): Equality of mean sectoral wages with different distributions.

In the case of independent draws from fréchet distributions, equilibrium wages per worker equalize across sectors. Using (a4) earlier:

(b1) $w_i\bar{z}_i = w_i\lambda_i\Gamma\left(\dfrac{\theta-1}{\theta}\right)\left(\dfrac{(w_i\lambda_i)^\theta}{\sum_i (w_i\lambda_i)^\theta}\right)^{-1/\theta} = \Gamma\left(\dfrac{\theta-1}{\theta}\right)\left(\dfrac{1}{\sum_i (w_i\lambda_i)^\theta}\right)^{-1/\theta}$

which is independent of i. This is not a general characteristic of this type of model. For example, for the case wher e the productivities are independent draws from the exponential distribution, we use (a6) and see:

(b2) $w_i\bar{z}_i = \dfrac{w_i}{\lambda_i}\left[\dfrac{(w_i\lambda_j / w_j\lambda_i) + 2}{(w_i\lambda_j / w_j\lambda_i) + 1}\right] \neq \dfrac{w_j}{\lambda_j}\left[\dfrac{(w_j\lambda_i / w_i\lambda_j) + 2}{(w_j\lambda_i / w_i\lambda_j) + 1}\right] = w_j\bar{z}_j$ unless $w_i/\lambda_i = w_j/\lambda_j$

(c): Independence of the paired productivity draws and $\eta(z) = zg(z)/G(z)$, the elasticity of the distribution function generating the draws, declining in z are, together, sufficient conditions for $d\bar{z}_i / d\pi_i < 0$ and $\xi < 0$, i.e. for average labour efficacy to be declining in a sector's share of total employment.

Equation (a1) above gives the formulas for $\bar{z}_i$ and π_i for a general joint distribution function $g_{i,j}(z_i,z_j)$ determining the paired productivity draws (z_i,z_j). (a1) also notes that these are functions of the endogenous variable $\omega = w_i/w_j$. From this we see that:

(c1) $$\frac{d\bar{z}_i}{d\pi_i} = \frac{d\bar{z}_i / d\omega}{d\pi_i / d\omega} = \frac{1}{\pi_i}\left[\frac{dN/d\omega}{d\pi_i/d\omega} - \bar{z}_i\right]$$

where $N(\omega)$ is defined earlier in (a1). As $dN/d\omega$ divided by $d\pi_i/d\omega$ equals $dN/d\pi_i$, which is the quality of the marginal worker, we see intuitively that the condition we are looking for is that the quality of the marginal worker entering the industry is less than that of the average worker. Substituting using the formulas in (a1) we have

(c2) $$\frac{d\bar{z}_i}{d\pi_i} = \frac{1}{\pi_i}\left[\frac{\int_0^\infty z_i^2 g_i(z_i) g_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}}{\int_0^\infty z_i g_i(z_i) g_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}} - \frac{\int_0^\infty z_i g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}}{\int_0^\infty g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}}\right] = \frac{1}{\pi_i}\left[E(a) - E(b)\right]$$

where $$F_a(x) = \frac{\int_0^x \eta(\omega z_i) g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}}{\int_0^\infty \eta(\omega z_i) g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}},\quad F_b(x) = \frac{\int_0^x g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}}{\int_0^\infty g_i(z_i) G_{j|i}(\omega z_i|z_i)\,\mathrm{dz_i}},\quad \eta(\omega z_i) = \frac{\omega z_i\, g_{j|i}(\omega z_i|z_i)}{G_{j|i}(\omega z_i|z_i)}$$

and where I have redefined the terms in [] as the difference between the expectation of two random variables with cumulative density functions $F_a(x)$ and $F_b(x)$. As is well known, if $F_a(x) \geq F_b(x)$ for all x, then $E(a) \leq E(b)$.[3] Note that $F_a(x)$ is the same as $F_b(x)$ except for the weighting function η. If z_i and z_j are independent, then η becomes

(c3) $$\eta(z) = \frac{z g_j(z)}{G_j(z)}$$

which is the elasticity of the distribution function. If this is non-increasing in its argument, then $F_a(x) \geq F_b(x)$ for all x[4] and $E(a) \leq E(b)$. Strict inequality follows if η is strictly decreasing.[5] Note that this is a sufficient but not necessary condition, as $E(a) < E(b)$ does not imply $F_a(x) > F_b(x)$ for all x.

---

[3] As $\int_0^\infty x f(x)\mathrm{dx} = \int_0^\infty f(x)\int_0^x 1\mathrm{dt}\,\mathrm{dx} = \int_0^\infty\int_t^\infty f(x)\,\mathrm{dx}\,\mathrm{dt} = \int_0^\infty [1-F(t)]\mathrm{dt}.$

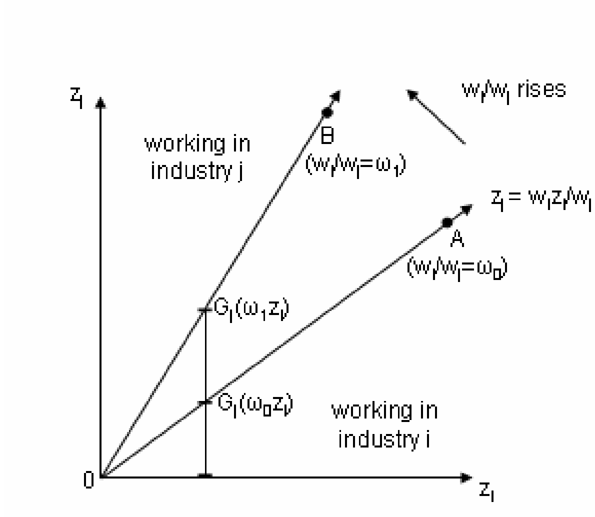
[4] Note that $F_a(x) = A/(A+1)$ and $F_b(x) = B/(B+1)$ where

$$A = \frac{\int_0^x \eta(\omega z_i) h(z_i)\,\mathrm{dz_i}}{\int_x^\infty \eta(\omega z_i) h(z_i)\,\mathrm{dz_i}} \geq \frac{\eta(x)\int_0^x h(z_i)\,\mathrm{dz_i}}{\eta(x)\int_x^\infty h(z_i)\,\mathrm{dz_i}} = \frac{\int_0^x h(z_i)\,\mathrm{dz_i}}{\int_x^\infty h(z_i)\,\mathrm{dz_i}} = B \quad \text{and} \quad h(z_i) = g_i(z_i) G_{j|i}(\omega z_i \mid z_i).$$

[5] Thus, for uniform distributions on [a,b], where η is a constant if a = 0, $d\bar{z}_i / d\pi_i = 0$ for some values of ω.

Figure A1 provides some intuition as to why independence alone is not sufficient to guarantee $\xi < 0$ and an additional condition on $zg(z)/G(z)$, such as that specified above, is needed. Individual talent is distributed across the (z_i,z_j) space depicted in the diagram. The ray $z_j = w_i z_i / w_j$ determines the division between sectors, with workers with (z_i,z_j) draws above the ray working in industry j and those with draws below the ray working in sector i. Initially, workers below ray $\overrightarrow{OA}$ work in industry i, but as w_i/w_j rises from ω_0 to ω_1 workers in the region encompassed by the rays $\overrightarrow{OA}$ and $\overrightarrow{OB}$ shift to the sector. The average quality of pre-existing sector i workers depends on the z_i weighted integral of the joint density in the area below $\overrightarrow{OA}$, while the quality of marginal workers depends upon the z_i weighted integral of the joint density in the area between $\overrightarrow{OA}$ and $\overrightarrow{OB}$. Even if z_i and z_j are independent, it is possible for the marginal worker to be of higher quality if the ratio $[G_j(\omega_1 z_i)\text{-}\ G_j(\omega_0 z_i)]/\ G_j(\omega_0 z_i)$ (the relative cumulative density for the z_j draws) rises with z_i in some regions, i.e. more relative weight is placed on higher values of z_i in the marginal worker integral. Thinking of ω_1/ω_0 as the same relative change applied for each z_i, avoiding this everywhere amounts to an elasticity restriction on G_j. The condition is sufficient, but not necessary, because it is possible for $[G_j(\omega_1 z_i)\text{-}\ G_j(\omega_0 z_i)]/\ G_j(\omega_0 z_i)$ to be rising in some areas and falling elsewhere and yet, depending upon the distribution of z_i, for the average quality of the marginal worker to still be lower than that of pre-existing workers.

Figure A1: Necessity for Additional Restrictions on the Cumulative Distribution Function



(d): The range of prices supported by the supply curve in the standard Cobb-Douglas model with unequal sectoral factor intensities (footnote in the introduction).

For the standard Cobb-Douglas model with homogenous labour and production functions $Q_i = A_i K_i^{\alpha_i} L_i^{1-\alpha_i}$, the first order conditions for the optimal use of labour and capital imply:

$$\text{(d1)}\quad w = P_i(1-\alpha_i)A_i(K_i/L_i)^{\alpha_i} \quad \text{and} \quad r = P_i\alpha_i A_i(K_i/L_i)^{\alpha_i - 1}$$

where w is the wage and r the rental. From this it follows that (d2) $$\frac{K_i}{L_i} = \frac{\alpha_i}{1-\alpha_i}\frac{w}{r}\ \text{, so}\ \ P_i = \frac{r(K_i/L_i)^{1-\alpha_i}}{\alpha_i A_i} = \frac{r(w/r)^{1-\alpha_i}(\alpha_i/1-\alpha_i)^{1-\alpha_i}}{\alpha_i A_i}$$

Consequently:

(d3) $$\frac{P_i}{P_j} = \frac{A_j}{A_i}\left(\frac{w}{r}\right)^{\alpha_j-\alpha_i}\frac{\alpha_j(\alpha_i/1-\alpha_i)^{1-\alpha_i}}{\alpha_i(\alpha_j/1-\alpha_j)^{1-\alpha_j}} = \frac{A_j}{A_i}\ \text{if}\ \alpha_i = \alpha_j$$

The last equality simply notes that if the factor income shares are identical, the standard model yields a horizontal Baumol supply curve.

Focusing on the first equality in (d3), we see that, holding constant the productivities A_i and A_j, the equilibrium variation in relative prices depends upon the equilibrium variation in *w/r*. The question is what range of variation in *w/r* is possible given constant total factor productivities and a constant endowment of capital and labour. Let sector j be the sector with the higher capital intensity ($\alpha_j > \alpha_i$), and note that in equilibrium it must be the case that $K_j/L_j \geq K/L \geq K_i/L_i$, i.e. the economy-wide capital-labour ratio must lie between the two sectoral capital-labour ratios. From (d2) this implies

(d4) $$\frac{\alpha_j}{1-\alpha_j}\frac{w}{r} \geq \frac{K}{L} \geq \frac{\alpha_i}{1-\alpha_i}\frac{w}{r} \quad \text{or} \quad \frac{1-\alpha_i}{\alpha_i}\frac{K}{L} \geq \frac{w}{r} \geq \frac{1-\alpha_j}{\alpha_j}\frac{K}{L}$$

Since the economy-wide capital labour ratio is the weighted average of the sectoral capital labour ratios (with weights equal to their employment shares), as *w/r* moves to its lower bound the output of sector i goes to 0, while as it reaches its upper bound the output of sector j goes to zero. Consequently, as *w/r* moves from its lower to its upper bound the relative output Q_i/Q_j goes from 0 to ∞. This traces out the supply curve. Combining (d3) and (d4) we see that the relative price change associated with this movement is:

(d5) $$\ln\left[\left(\frac{P_i}{P_j}\right)^{\max} \Big/ \left(\frac{P_i}{P_j}\right)^{\min}\right] = (\alpha_j - \alpha_i)\ln\left[\frac{1-\alpha_i}{\alpha_i} \Big/ \frac{1-\alpha_j}{\alpha_j}\right]$$

In the BLS KLEMS 1987 to 2010 database the average annual capital income shares of value added for aggregate goods and services are $\alpha_G = .35$ and $\alpha_S = .32$, respectively. Plugging these numbers into (d5), we get a variation in the ln relative price of goods to services from the bottom to the top of the supply curve of .03*ln(119/104) = .0040, i.e. 4/10ths of one percent. For all intents and purposes, this is a horizontal supply curve. Thus, operating as if goods and services share the same factor income share provides a very close approximation to the actual relative supply curve generated by their differing factor intensities.

Appendix B: Labour Quality Measures for the BLS KLEMS

As noted in the text, the BLS KLEMS total factor productivity estimates do not differentiate by worker type. For its aggregate private business and private non-farm business TFP measures, however, the BLS constructs measures of differentiated labour input using March Supplement Current Population Survey (CPS) data to construct measures of differentiated labour input and then adjusting the hours totals to match Current Employment Statistics (CES). I use a similar methodology to construct differentiated labour measures for the 60 private sectors in the KLEMS and the government sector.

The first difficulty one encounters lies in matching the industrial sector definitions of the CPS and the KLEMS. From 2003 to 2010, the CPS data uses aggregations of the categories in the 2002 NAICS (North American Industry Classification System), which are a close match to the NAICS categories used in the 60 sector KLEMS. The only exceptions are NAICS 523 (securities, commodity contracts and investments) and 525 (funds, trusts and other financial vehicles), which are separate in the KLEMS but combined in the CPS data. I assume that the distribution of workers by type within the two sectors is the same as in the combined CPS sector. Pre-2003 data, however, are based upon the 1972, 1980 and 1987 SIC (Standard Industrial Classification) codes. While the differences between one SIC and another are minor, and easily reconciled by renumbering and combining a few detailed sub-categories, the differences between the SIC and the NAICS appear more substantial.

The BLS and I address the issue of changing sectoral definitions in labour statistics using 2000-2002 CPS data. In the 2000, 2001 and 2002 iterations of the CPS, industry and occupation data were collected using both the old and new classification systems. In its published labour statistics, the BLS uses the cross-distribution of employment between old industry and new industry in the dual coded data to convert the old data series to the new industrial definitions (http://www.bls.gov/cps/constio198399.htm). I follow a similar methodology, except that I use the cross distribution from old system industry x occupation categories to new industry. However, there are hundreds of industry and occupation categories, so not every industry x occupation cross-classification present in the 1987-2002 data appears in the 2000-2002 sample. For those missing observations, I use higher levels of aggregation, using first the old system industry x detailed (46 categories) occupation cross-classification, then the old system industry x major (14 categories) occupation cross-classification, and, when all else fails, for a handful of observations, simply the old system industry to new system industry distribution.

The second problem that arises is that of zeros. I cross-classify workers by 61 sectors (60 private plus public administration), 2 sexes, 6 age groups, 5 educational categories, and 24 years.[6] Given the limited samples in the CPS, this inevitably creates lots of zeros. Zeros are a serious problem, as total factor productivity calculations involve calculating ln changes. I address this issue by using iterative proportional fitting[7] to estimate the full five-dimensional cross distribution using sub-dimensional totals. Iterative proportional fitting fits a model that assumes independence at higher dimensions. To illustrate with the three dimensional example where X is cross classified by i, j and k, one can use the observed X_{ij}, X_{jk}, and X_{ik} totals to produce estimates $\hat{X}_{ijk}$ which are ln-linearly related to implicit interaction factors λ_{ij}, λ_{jk}, and λ_{ik}, with no interactions at the i x j x k level. By using sub-dimensional totals to estimate the full array, one eliminates the zeros in the detailed cross-classifications. For my estimates of wages per hour, where the samples are particularly sparse as the data are not available for all workers, I use all of the two dimensional cross-classifications to estimate the five dimensional array. I calculate total hours and total income for each two-dimensional sub-array, iteratively proportional fit the entire five dimensional array, and then take ratios of cells to calculate wages per hour. For my worker and hours data, the samples are larger. I begin by defining 12 major sector aggregations (the principal sectors, with manufacturing sub-divided into durables and non-durables) for the 61 detailed sectors. I then iteratively proportionally fit the five dimensional array using every available three dimensional array based upon major industry classification and every two dimensional array based upon detailed industry classification.[8] The use of major industry aggregations allows me to include interactions at higher dimensions without introducing zeros into cells, while the detailed industry two dimensional arrays retain the information on cross-distributions at that level.

---

[6]The age categories are 15-24, 25-34, 35-44, 45-54, 55-64, and 65+; the educational categories are less than high school, completed high school, some college, completed college, and more than college; the years are 1987-2010.

[7]See Agresti, Alan. Categorical Data Analysis. New York: John Wiley and Sons, 1990.

[8]Thus, allowing D to denote detailed industry, M major industry, S sex, A age, E education and Y year, I use the sub-dimensional arrays DS, DA, DE, DY, MSA, MSE, MSY, MAE, MAY, MEY, SAE, SAY, SEY, and AEY. In iterative proportional fitting, one can aggregate a dimension into sub-categories. As long as that sub-category contains additional cross-distributions, it is not redundant (i.e. MS is redundant given DS, but MSA is not) and provides an additional interaction factor.

To summarize my procedure, I begin by using the 2000-2002 CPS SIC industry x occupation to NAICS industry population distribution to convert 1987-2002 industry data to 2002 NAICS definitions. I then use the CPS March Supplement individual weights and aggregate to the 60 KLEMS sectors plus the government sector. I treat as a "worker" anyone who reports more than zero hrs of work in the previous week. I then adjust the population totals and hours of work totals by year x industry to match the BLS estimates of workers and hours by year x industry[9] and iteratively proportionally fit workers and total hours to calculate workers and hours by industry x sex x age x education x year. For wages per hour, I take all individuals for which the BLS is able to calculate a wage per hour (based upon the direct report or data on "usual hours"), aggregate into 61 sectors using the CPS weights, adjust hours totals by industry using the BLS CES data, and then iteratively proportionally fit total earnings and hours, taking the ratio of the two five dimensional arrays to calculate wages per hour. The combination of hours and wages per hour then allow me to calculate sub-factor income shares by industry (Θ_{Li}^{j} in the paper) and the data on hours per worker allow me to calculate Tornqvist measures of the growth of labour quality by sector which are comparable to those the BLS calculates for the aggregate private sector:

$$\text{(B1)} \quad \sum_j \left(\frac{\Theta_{Lit}^{j} + \Theta_{Lit-1}^{j}}{2} \right) \ln\left(\frac{H_{it}^{j}}{H_{it-1}^{j}} \right) - \left(\frac{\Theta_{Lit} + \Theta_{Lit-1}}{2} \right) \ln\left(\frac{H_{it}}{H_{it-1}} \right)$$

where H_{it}^{j} denotes total hours of worker type j in industry i at time t and H_{it} denotes total hours in sector i at time t. The measures are added to the growth of labour input and subtracted from the growth of total factor productivity in the BLS data. The data on the distribution of the population by worker characteristic then allow me to calculate weighted and unweighted Tornqvist measures of the changing shares of the labour force:

$$\text{(B2)} \quad \sum_j \left(\frac{\Theta_{Lit}^{j} + \Theta_{Lit-1}^{j}}{2} \right) \ln\left(\frac{\pi_{it}^{j}}{\pi_{it-1}^{j}} \right) \quad \text{and} \quad \left(\frac{\Theta_{Lit} + \Theta_{Lit-1}}{2} \right) \ln\left(\frac{\pi_{it}}{\pi_{it-1}} \right)$$

where π_{it}^{j} denotes the share of the aggregate working population of type j in industry i at time t and π_{it} denotes the share of the aggregate working population in sector i at time t. These

[9] The KLEMS TFP database only contains indices of hours. I take levels of hours and workers from the Industry Employment and Hours Data Tables of the BLS labour productivity database. These are not strictly consistent with the hours indices of the BLS KLEMS total factor productivity database. However, I do not use these totals to change the measure of the growth of total labour input (hours) in the KLEMS database calculations, but only to calculate distributions of workers by characteristic, as shown shortly in (B1) and (B2).

measures are used as the instrumented dependent variable in Section II. Since everything is benchmarked to the BLS totals, the H_{it} and π_{it} measures are simply the original BLS data and are consistent with the totals of H_{it}^{j} and π_{it}^{j} across j. The two measures in (B2) are different, but highly correlated, with a correlation coefficient of .917.

Appendix C: Existing Micro-Data Estimates (McLaughlin & Bils 2001)

McLaughlin & Bils (2001, tables 4 and 5) using PSID data from 1979 to 1992 report that the average ln wage of industry leavers relative to stayers in industries with contracting employment shares and industry entrants relative to stayers (continuing workers) in industries with expanding employment shares is about -16 or -17 percent without adjustment for worker characteristics and -6 or -7 percent with adjustment for worker characteristics. These estimates might lead one to conclude that comparative advantage is indeed aligned with absolute advantage, but that Roy worker efficacy effects are rather small. In this appendix I show that the data examined by McLaughlin & Bils have little to do with the expansion and contraction of industries and are mostly related to a form of "churning" whereby workers simultaneously exit and enter industries.

I work with the annual 1971-1997[10] records of the PSID, using both the low income sample and the census based representative sample, focusing on the industry of employment of household heads. I use two industrial classifications: (a) 9 aggregate sectors, a measure which should eliminate spurious industry shifts brought about by minor errors and misclassifications; (b) 24 sectors, which is the greatest detail I can achieve while keeping industry definitions relatively consistent across the years, and is similar to the detail used by McLaughlin & Bils.[11] For a given industry i, examined in period t, workers are classified as stayers if they were in the same industry i in period t-1, entrants if they were in a different industry j in period t-1, and leavers if they worked in industry i in period t-1 but work in industry j in period t. To be in the sample a worker needs to both report industry of employment and allow the calculation of ln

[10]Prior to 1971 industry is not reported; after 1997 the PSID moves to a biannual framework and hence the calculation of movers and stayers is not comparable.

[11]9 sectors: 1 agriculture, forestry & fishing; 2 mining; 3 manufacturing; 4 construction; 5 transport, communications & utilities; 6 wholesale & retail trade; 7 finance, insurance and real estate; 8 other services (except gov't); and 9 government & armed forces. 24 sectors: 1 agriculture, forestry & fishing; 2 mining; 3 metal industries; 4 machinery (inc. electrical); 5 motor vehicles & other transportation equipment; 6 food & kindred products (inc. tobacco); 7 textile mill products, apparel & other fabricated textile products, plus shoes; 8 paper & allied products; 9 chemical & allied products, petroleum & coal products, and rubber & misc. plastic products; 10 printing & publishing; 11 other manufacturing; 12 construction; 13 transportation; 14 communication; 15 public utilities; 16 wholesale trade; 17 retail trade; 18 finance, insurance and real estate; 19 business services; 20 personal services; 21 health; 22 education; 23 other services (except gov't); 24 government & armed forces.

Table C-1: Entry, Exit and Sectoral Growth in the PSID (observations are industry x year)									
		Representative & Low Income Sample				Representative Sample Alone			
		7 Private Sectors		22 Private Sectors		7 Private Sectors		22 Private Sectors	
		Rates	ln Rates	Rates	ln Rates	Rates	ln Rates	Rates	ln Rates
Correlation Between Entry and Exit Rates									
General (p-value)		.675 (.000)	.717 (.000)	.842 (.000)	.759 (.000)	565 (.000)	.577 (.000)	.747 (.000)	.664 (.000)
Partial (p-value)		.282 (.000)	.167 (.025)	.302 (.000)	.128 (.002)	.224 (.002)	.118 (.114)	.229 (.000)	.106 (.012)
N		182	181	572	571	182	180	572	564
Regression on Change in Industry Employment Share (with industry & year dummies)									
Entry	$\Delta\pi_{it}$	.165	1.07	.217	1.11	.175	.828	.237	.751
	(s.e.)	(.099)	(.506)	(.085)	(.410)	(.132)	(.711)	(.111)	(.544)
	N	182	182	572	572	182	182	572	567
Exit	$\Delta\pi_{it}$	.061	.094	.025	-.077	-.125	-.545	.017	-.245
	(s.e.)	(.099)	(.507)	(.081)	(.415)	(.125)	(.706)	(.105)	(.569)
	N	182	181	572	571	182	181	572	569

Notes: N = number of industry x year observations. An occasional observation is lost when taking the ln of a zero entry or exit rate. Partial correlation = correlation of residuals from regression on industry and year dummies. Regressions = regression of entry or exit rates on industry & year dummies and the change in the share of non-agricultural employment ($\Delta\pi_{it}$).

wage per hour in consecutive years. This eliminates unknown industry and workers who were completely out of employment in one year or the other. Every worker who is an entrant in industry i in period t is a leaver from some industry j in period t-1. Although I use all 9 or 24 sectors to categorize workers, I focus on entry/exit rates in the 7 or 22 private non-agricultural sectors.[12] Overall I have about 61500 individual x year observations (a little over half in representative sample households) in these industries, with about 15% of these being entrants or leavers according to the broad sectoral definitions and 23% according to the narrow sectoral definitions.

I begin by reporting, in the top panel of Table C-1, the correlation between the sample fractions, at the industry x year level, of entrants (in entrants and stayers) and leavers (in leavers and stayers). As shown, there is a very strong positive correlation between the fraction of the

[12]I relate these rates to the BLS Current Employment Statistics based historical SIC measures of employment, which exclude agriculture, while the focus on private sector activity is consistent with the measures examined earlier in the paper.

sample that enters an industry between period t-1 and t and the fraction that leaves the same industry between the same two periods. This holds true whether the measures are in levels or in lns, using both the low income and representative sample or just the representative sample alone. The partial correlation of the entry and exit rates, after removing industry and year fixed effects, is weaker but still generally highly significant. In contrast, in the bottom panel of the table I report the regression of the industry entry and exit rates on the change in the sector's share of non-agricultural employment, with industry and year fixed effects. As shown, the regression coefficients are almost universally insignificant, the only exception being entry rates for the 22 industry measure, and this result largely disappears when the sample is restricted to PSID representative households alone.

Table C-2 follows the McLaughlin & Bils methodology, examining the average relative ln wages of different groups. Without adjustment for worker characteristics, the wages of entrants or leavers are found to be between 11 and 19 percent lower than those of stayers (first four columns). With adjustment for worker characteristics (last four columns), these mean differences are greatly reduced and, in many cases, rendered statistically insignificant. Moreover, in all cases the vast majority of the estimates that underlie the calculation of these averages are insignificant. Thus, for example, while the relative wages of entrants to stayers in expanding sectors are on average 3.1 percent (7 sectors) or 2.9 percent (22 sectors) lower among the representative PSID sample, only about $1/10^{th}$ of the industry x year differences that underlie the calculation of these means are, by themselves, statistically significant at the 5% level. Unlike McLaughlin & Bils, Table C-2 reports relative wages in both expanding and contracting sectors for all measures. As shown, while the relative wages of entrants are lower than stayers in expanding industries, the difference is, generally, even larger in *contracting* industries. Similarly, while the relative wages of leavers are lower than stayers in contracting industries, the difference is generally almost as large in *expanding* industries. These results completely undermine the interpretation of these wage differences as reflecting the relative efficacy of entrants in expanding industries and leavers in contracting industries.

In sum, changes of industrial sector in the PSID appear to reflect a form of "churning", whereby both entry and exit simultaneously occur within industries. It is not hard to motivate such movement, either with models of creative destruction within sectors or with a more general idiosyncratic destruction of existing jobs and appearance of new opportunities. Workers with systematically lower human capital appear to play a disproportionate role in this churning, as adjustment for observable characteristics eliminates most of the relative wage differences. While

Table C-2: Mean Wage Differences Between Industry Entrants or Leavers vs. Stayers in the PSID								
	Average ln Wages Differences				Adjusted for Worker Characteristics			
	7 Private Sectors		22 Private Sectors		7 Private Sectors		22 Private Sectors	
	$\pi_{it} \uparrow$	$\pi_{it} \downarrow$	$\pi_{it} \uparrow$	$\pi_{it} \downarrow$	$\pi_{it} \uparrow$	$\pi_{it} \downarrow$	$\pi_{it} \uparrow$	$\pi_{it} \downarrow$
Entrants vs. Stayers (representative & low income PSID sample)								
Mean Dif.	-.141	-.189	-.128	-.161	-.031	-.065	-.033	-.050
(s.e.)	(.014)	(.014)	(.010)	(.010)	(.009)	(.010)	(.007)	(.007)
Significant/N	27/90	44/92	65/241	84/331	12/90	25/92	25/241	41/331
Leavers vs. Stayers (representative & low income PSID sample)								
Mean Dif.	-.144	-.164	-.128	-.134	-.012	-.032	-.008	-.019
(s.e.)	(.014)	(.015)	(.011)	(.010)	(.010)	(.011)	(.008)	(.007)
Significant/N	25/92	32/89	53/243	68/328	7/92	6/89	21/243	12/328
Entrants vs. Stayers (representative PSID sample)								
Mean Dif.	-.140	-.154	-.118	-.117	-.031	-.059	-.029	-.022
(s.e.)	(.019)	(.019)	(.014)	(.014)	(.012)	(.014)	(.009)	(.010)
Significant/N	18/90	34/91	34/241	59/326	10/90	20/91	22/241	35/326
Leavers vs. Stayers (representative PSID sample)								
Mean Dif.	-.120	-.162	-.108	-.112	-.007	-.045	-.004	-.008
(s.e.)	(.019)	(.019)	(.015)	(.014)	(.014)	(.013)	(.010)	(.011)
Significant/N	11/92	28/89	29/242	50/327	7/92	9/89	21/242	28/327

Notes: Observations are industry x year measures of wage differences. Adjusted for Worker Characteristics = the coefficients on entrant (or leaver) yearly dummies in industry level regressions with controls for sex, age, age2, race (African-American), education (8 categories) and year (dummies), with random effects for PSID individuals. $\pi_{it} \uparrow$ ($\pi_{it} \downarrow$): observations in industries whose share of total employment increased (decreased) in that year. Mean Dif: mean year x industry difference for observations with $\pi_{it} \uparrow$ or $\pi_{it} \downarrow$; s.e. = standard error of the mean difference; N = number of industry x year observations; Significant = number of such observations which are, individually, significantly different from 0 at the 5% level.

these facts are interesting in and of themselves, they provide little insight into the impact of the expansion or contraction of industry employment shares on average worker efficacy.