Alwyn Young

Structural transformation, the mismeasurement of productivity growth, and the cost disease of services

Article (Published version)
(Refereed)

Original citation: Young, Alwyn (2014) *Structural transformation, the mismeasurement of productivity growth, and the cost disease of services.* American Economic Review, 104 (11). pp. 3635-3667. ISSN 0002-8282
DOI: 10.1257/aer.104.11.3635

American Economic Review 2014, 104(11): 3635–3667
http://dx.doi.org/10.1257/aer.104.11.3635

Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services†

By Alwyn Young*

If workers self-select into industries based upon their relative productivity in different tasks, and comparative advantage is aligned with absolute advantage, then the average efficacy of a sector's workforce will be negatively correlated with its employment share. This might explain the difference in the reported productivity growth of contracting goods and expanding services. Instrumenting with defense expenditures, I find the elasticity of worker efficacy with respect to employment shares is substantially negative, albeit estimated imprecisely. The estimates suggest that the view that goods and services have similar productivity growth rates is a plausible alternative characterization of growth in developed economies. (*JEL* E23, E24, H56, J24, O41, O47)

One of the strongest and seemingly most accurate characterizations of the process and problems of growth in advanced economies is William Baumol's "Cost Disease of Services." Baumol's argument, begun in papers as early as 1965 and continuing to this very day (e.g., Baumol and Bowen 1965; Baumol 1967; Baumol, Blackman, and Wolff 1985; and Baumol 2012), starts from the premise that productivity growth is inherently more difficult to achieve in the production of services than in the production of goods. With the two industries competing for factors of production in the same factor markets, the relative cost of producing service output inevitably rises. If the demand for services was income inelastic and price elastic, these trends would not pose a problem, as the share of services in nominal GDP (gross domestic product) would decline. Alas, precisely the opposite is true, and services garner an increasing share of nominal output. Aggregate productivity growth, equal to the nominal output share weighted average of sectoral productivity growths, must decline steadily.[1]

Decades of data on productivity growth in goods and services have confirmed Baumol's thesis, turning it, for all intents and purposes, into a stylized fact of economic growth. Productivity statistics, however, are based on the fundamental

* Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE (e-mail: a.young@lse.ac.uk). I am greatly indebted to Ho Veng-Si and anonymous referees for many helpful comments. The author declares that he has no relevant or material financial interests that relate to the research described in this paper.

† Go to http://dx.doi.org/10.1257/aer.104.11.3635 to visit the article page for additional materials and author disclosure statement(s).

[1] Although not mentioned in the papers cited above, implicit in Baumol's argument is the notion that service output is relatively nontradeable. Otherwise, low productivity growth in services could be met, at least at the individual country level, by exporting more manufactures for services.

assumption that each new worker is qualitatively the same as every old worker.[2] If workers self-select into industries based upon unobservables, this assumption may create a systematic bias, as the type of workers present when an industry is small may not be the same as when the industry becomes large, and vice versa.

In his "Thoughts on the Distribution of Earnings," Roy (1951) identified the mechanism central to this paper. Workers select the industry in which they have the highest relative productivity (i.e., a comparative advantage). If individual productivity in different tasks is uncorrelated or at worst weakly correlated, then individuals having a comparative advantage in an industry will on average also have an absolute advantage in that sector. As a sector expands by offering higher wages to prospective workers elsewhere in the economy, it will draw in individuals with both a lower comparative advantage and a lower absolute advantage in the sector, while leaving individuals with the highest comparative and absolute advantage in competing sectors. Consequently, productivity in expanding sectors will appear to decline and productivity in contracting sectors will appear to rise. In sum, in a Roy world the apparent disparity in the productivity growth of goods and services may come about because services expand by drawing in people who are, as examples, less adept at finance, law, and medicine, while goods sectors contract by shedding the least able farmers, manufacturers, and miners—all of which is not taken into account in measures of productivity growth. Underlying true levels of productivity growth—i.e., taking into account the average efficacy of the workers present in the two sectors—might not be all that different.

Figure 1, which graphs the relative supply and demand for services, summarizes the argument made in this paper. Baumol's supply curve is essentially a horizontal line, determined by the relative productivity of the two sectors.[3] As goods experience more rapid productivity growth, this supply curve shifts up, from S_0^{Baumol} to S_1^{Baumol}, exemplifying the cost disease of services. At the same time, as a consequence of the relatively higher income elasticity of demand for services, the relative demand curve shifts out from D_0 to D_1. The equilibrium moves from E_0 to E_1, with a higher relative output and price of services, which has consequently a growing nominal share of the economy. An alternative hypothesis, however, is that the supply curve is substantially upward-sloping because of the correlation between comparative and absolute advantage Roy describes. As drawn in the figure, the Roy supply curve S^{Roy} intersects both E_0 and E_1. This describes a situation in which productivity growth is the same in both sectors, so the supply curve does not shift, but the relative demand curve shifts out as incomes rise. Here the rise in the relative price of services is driven purely by the changing efficacy of the average worker in each sector.[4]

[2] To be sure, more sophisticated analyses divide workers into categories based upon observable determinants of human capital such as age and education, but within each category the assumption is ultimately made that all workers are identical.

[3] If the capital income shares (i.e., factor intensities) of the two sectors differ, the supply curve will be upward sloping even without the effects Roy describes. However, as discussed in the online Appendix, empirically the capital income shares of goods and services in the US economy are almost identical and the upward slope in the supply curve attributable to this effect is negligible: i.e., an increase in relative prices of 0.4 of 1 percent as relative output goes from 0 to ∞. In the sources cited above, Baumol and his coauthors don't emphasize a relative price effect emanating from relative factor intensities and, in this regard, appear to be completely correct.

[4] As Figure 1 makes clear, for Baumol's argument it does not matter whether or not the relative real output of services is rising (only that its nominal share is increasing), but for the Roy argument it does. Baumol, Blackman, and Wolff (1985) argue that there is no change in the relative output of goods and services. This is actually

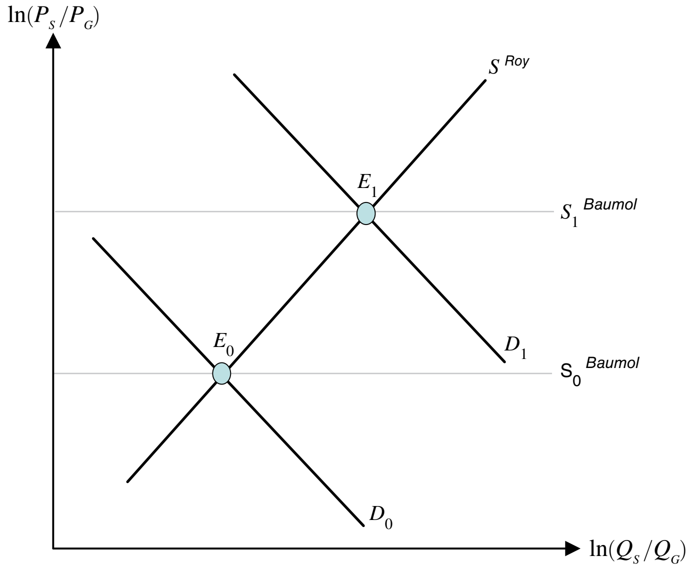

Figure 1. Alternative Views of Relative Supply

Figure 1 makes clear that the Roy model does not deny the rise in the relative price of services, it merely explains it with a different mechanism. Figure 2 illustrates why this matters. Panel A draws the linear production possibilities frontier implied by the Baumol model, which rotates out as goods experience more rapid productivity growth.[5] Panel B draws the Roy production possibilities frontier, which shifts out uniformly when productivity growth is identical in both sectors. This panel shows that the same equilibrium price and quantity relations can be explained with equiproportional shifts of the intercepts of the production possibility frontier and a movement along its concave surface. For the purposes of heuristically illustrating welfare implications, the diagrams also include social indifference curves which, under the assumption of competitive markets, are tangent to the production possibilities frontiers. Aggregate total factor productivity (TFP) growth is the proportional increase in the length of the ray from the origin to the tangent line on the production frontier ($\Delta V/V$ in the figure).[6] In the Baumol model, as the share of services in total

not true. As discussed in Section II, US and OECD (Organisation for Economic Co-operation and Development) data clearly indicate a large rise in the relative real output of services in the postwar era.

[5] For the purposes of this expositional diagram, I assume that factor supplies are constant.

[6] To see this, note that if inputs are constant (as assumed in the diagram), we can describe the problem of maximizing GDP as one of maximizing $P_xX + P_YY$ s.t. $0 \geq F(X, Y, t)$. Differentiating the binding production possibilities constraint, we have (a) $F_XdX + F_YdY + F_tdt = 0$. Rearranging and making use of the first order conditions from the maximization problem ($\lambda F_X = P_X$, etc), one finds that:

$$g_{TFP} = -\lambda F_t/GDP = \theta_X g_x + \theta_Y g_Y,$$

where λ is the value of relaxing the PPF (production-possibility frontier) constraint, θ_i the GDP share of product i, and g_i the growth of the output of i. Thus, total factor productivity growth, the proportionate value of the time trend relaxation of the PPF constraint, equals the GDP share weighted increase of the output of each product. (a) in the paragraph above, however, holds whether the dX and dY are the observed values or imposed values such that $dX = g \times X$ and $dY = g \times Y$. Thus, regardless of the bias of TFP growth, one can equally say that $g_{TFP} = \theta_X g + \theta_Y g = g$. g is the proportionate increase in the length of the vector (with slope determined by the current production bundle) from the origin to the production possibilities frontier.

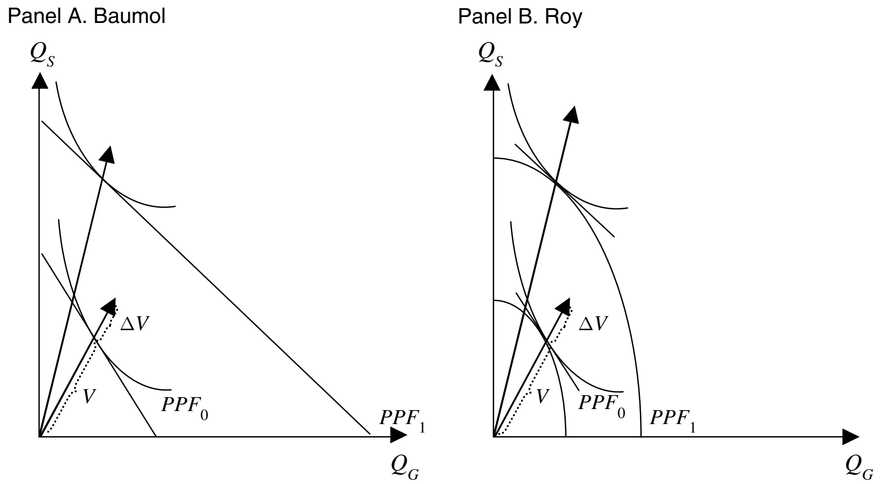

Figure 2. Welfare Implications with the Same Equilibrium Prices and Quantities
($TFP\ growth = \Delta V/V$)

expenditure grows, the growth rate of this vector slows. In the Roy model, the proportional growth rate remains constant. Over time there is a growing discrepancy in the instantaneous rate of welfare growth predicted by the two models.

This paper draws its inspiration from recent interest in the macro-implications of Roy's model. Lagakos and Waugh (2011) argue that selection effects of the type described in this paper can explain the greater relative productivity of agricultural workers to nonagricultural workers in countries with larger nonagricultural sectors. Hsieh, Hurst, Jones, and Klenow (2012) calculate the inefficiency associated with the historical concentration of women and African Americans in particular occupations using a Roy model, and argue that the gradual elimination of barriers to the participation of these groups in other occupations can explain as much as one-fifth of postwar US aggregate wage growth. Kuralbayeva and Stefanski (2013), independent of this paper, argue that the decrease of manufacturing output brought about by the appreciation of the real exchange rate associated with resource windfalls generates a spurious rise in manufacturing productivity as the contraction of the sector leaves only the most productive workers behind. This paper extends these Roy-related analyses to the general consideration of the relative productivity of goods and services. Along the way, I establish the theoretical bias in conventional measures of sectoral productivity and clarify the mathematical conditions necessary for Roy effects to be present (i.e., for average worker efficacy to be declining in a sector's employment share). While the papers above calibrate their models, this paper estimates the size of Roy effects using regression techniques.

There has been limited prior research with regards to estimating empirically the elasticity of average worker efficacy with respect to the sectoral employment share, the key parameter in the macro-implementation of the Roy model. Heckman and Sedlacek (1985), using CPS (Current Population Survey) microdata and instrumental variables, find that the local elasticity of worker efficacy with respect to the employment share is around -0.5 for manufactures and -1 for nonmanufactures (see their Table 3). This brackets roughly the range of estimates found in this paper.

McLaughlin and Bils (2001) find milder effects, using PSID (Panel Study of Income Dynamics) microdata to show that the wages of entrants or leavers are 6 to 17 percent lower than those of continuing workers. However, as discussed in the online Appendix, the PSID data used in that paper concern mostly simultaneous entry and exit (a form of employment churning) and are uncorrelated with changes in sectoral employment shares. This paper focuses directly on the impact of changes in sectoral employment, using private sector employment changes driven by changes in military spending to identify the elasticity of average worker efficacy with respect to sectoral employment.

Zvi Griliches, in his AEA presidential address (1994) and earlier (1992), brought to the profession's attention the shortcomings of US measures of service sector output, such as those which extrapolated inputs, eliminating productivity growth by construction. Since his time, however, there have been vast improvements in the national income accounts measures of service sector activity, particularly in regards to the recent time period (1987–2010) which is the focus of this paper's analysis. Triplett and Bosworth (2004) provide a review of these developments and the problems which remain. This paper takes as given the official measures of sectoral output, focusing on the systematic bias brought about by the failure to consider the relation between employment shares and average worker efficacy.

The paper proceeds as follows. I begin in Section I by presenting a simple Roy model, showing how the bias in sectoral measures of total factor productivity growth and the slope of the relative supply curve depend upon a key parameter: the elasticity of average worker efficacy within a sector with respect to that sector's share of total employment. Section I also shows how correlation between an individual's productivity in different activities can eliminate the positive association between comparative advantage and absolute advantage, overturning Roy's prediction that average worker efficacy is inversely related to a sector's employment share. Thus, the relation between worker efficacy and sectoral employment depends upon the process generating individual productivity draws: i.e., it is ultimately something which needs to be estimated empirically rather than identified theoretically.

Section II presents industry-level evidence that the elasticity of worker efficacy with respect to sectoral employment is, indeed, substantially negative. Projecting the Bureau of Labor Statistics KLEMS[7] measures for the United States private sector divided into 60 sectors, and the University of Groningen's KLEMS measures for private sector activity in 18 OECD countries divided into 29 sectors, on a variety of instruments, I find that defense spending is the only instrument which robustly satisfies the dual requirements of first-stage significance and second-stage exogeneity (the exclusion restriction) necessary for two-stage least squares. Estimates of the long-run elasticity of worker efficacy with respect to the sectoral employment share range from -0.5 to -1, with most observations concentrated in the more negative half of this range. I also find that an elasticity of -0.75 equalizes goods and services productivity growth in the United States and the OECD at large. It produces a stable Roy supply curve which matches the historical US and OECD data on relative goods and services price and quantity growth, as heuristically illustrated in Figure 1 above.

[7] Capital (K), labor (L), energy (E), materials (M), and purchased service inputs (S).

Section III concludes the published paper. An online Appendix provides mathematical proofs of the theoretical claims made in Section II. While the BLS (Bureau of Labor Statistics) adjusts its aggregate economy-wide measures of labor input growth for compositional effects, it does not do this in the sectoral KLEMS database. The online Appendix also describes how I develop detailed sectoral measures of labor composition which I use to adjust the BLS measures of total factor productivity growth and the sectoral measures of changing employment shares. Finally, as mentioned above, the online Appendix provides a review of the PSID data used in the McLaughlin and Bils paper, showing that it concerns simultaneous entry and exit, rather than the overall expansion and contraction of sectoral employment, which is the focus of Roy's model and this paper.

I. Structural Transformation and the Mismeasurement of Productivity

In this section I present the main theoretical results of the paper. Throughout the analysis I focus on supply relations alone, leaving the general equilibrium closure of the model with preferences and demand unspecified. This is both because I do not want to take a stand on the nature of preferences and demand (including trade), and because it is unnecessary to do so. All of the implications of the Roy model can be understood in terms of the supply curve and all of the theoretical analysis can be understood in terms of movements along that curve, movements whose causes, while obviously related to demand, do not need to be specified. To focus on intuition, I confine the mathematical proofs of the claims made in this section to the online Appendix.

A. *A Simple Model*

Consider an economy with two perfectly competitive industries, goods (G) and services (S). Value added in industry i $(= G$ or $S)$ is produced with capital and labor

$$Q_i = A_i F_i \left(K_i, \int_{u \in \mathcal{Set}_i} z_i(u) \right), \tag{1}$$

where $\mathcal{Set}_i$ is the set of workers u laboring in industry i and $z_i(u)$ is the efficacy or productivity of individual u when working in industry i. Each worker is endowed with a pair of industry productivities (z_G, z_S) which is drawn from some joint cumulative distribution function $G(z_G, z_S)$.

Workers move to the industry providing the highest financial reward. Thus, with w_i denoting the wage per unit of effective labor offered in industry i, the set of individuals choosing to work in that sector is given by

$$\mathcal{Set}_i = \{ u \,|\, w_i z_i(u) > w_j z_j(u) \}, \tag{2}$$

where j is the sectoral complement of i. Define π_i as the probability a worker selects industry i or, equivalently, the share of the labor force in industry i. With L denoting the total labor force, L_i, the number of workers in industry i equals $\pi_i L$. For a given distribution of (z_G, z_S) draws, π_i is determined in a general equilibrium which includes a specification of demand, with $d\pi_i/d\omega > 0$, where $\omega = w_i/w_j$.

Define the expected efficacy of a worker in sector i (i.e., their productivity conditional on working in that sector) as

$$(3) \qquad \overline{z}_i = E\big(z_i(u) \,|\, u \in \mathcal{Set}_i\big) = \frac{\int_{u \in \mathcal{Set}_i} z_i(u)\,du}{\int_{u \in \mathcal{Set}_i} du} = \frac{\int_{u \in \mathcal{Set}_i} z_i(u)\,du}{L \times \pi_i}.$$

As proven in the online Appendix, regardless of the distribution function generating the paired draws (z_G, z_S), the elasticity of average worker efficacy with respect to the sectoral employment share is greater than -1:

$$(4) \qquad \xi = \frac{d\overline{z}_i}{d\pi_i}\frac{\pi_i}{\overline{z}_i} > -1.$$

From (3), we see that if we ignore the numerator the elasticity of $\overline{z}_i$ with respect to π_i is -1. The numerator, however, is increasing in π_i, as anything which increases the total number of workers will increase the cumulative sum of their productivities. Consequently, the overall elasticity of $\overline{z}_i$ with respect to π_i will be greater than -1 (examples for particular functional forms are provided in the online Appendix). None of the empirical estimates presented later in Section II reject this prediction. While ξ may be positive or negative, Roy (1951), as explained in the introduction, argued that it should be negative: i.e., average worker efficacy declines as a sector expands and draws in less productive workers. For the moment, I will assume this to be true.

Aggregate labor input in an industry is a product of the number of workers times the average efficacy per worker, so the production function is usefully reexpressed as

$$(5) \qquad Q_i = A_i F_i(K_i, L_i \overline{z}_i).$$

From this, we see that total factor productivity growth, calculated properly, is given by[8]

$$(6) \qquad \hat{A}_i(true) = \hat{Q}_i - \Theta_{Ki}\hat{K}_i - \Theta_{Li}(\hat{L}_i + \hat{\overline{z}}_i),$$

where a ^ denotes a proportional change and Θ_{Ki} and Θ_{Li} are the factor income shares of capital and labor in sector i, respectively. Unfortunately, in estimating total

[8] The derivation is the usual one for total factor productivity calculations. With perfect competition the capital rental and wage per unit of effective labor equal the value marginal product of each factor, so the elasticity of the production function with respect to each factor equals the factor income share.

factor productivity growth, growth accountants treat each new worker as the equivalent of existing workers,[9] estimating total factor productivity growth to be

$$(7) \qquad \hat{A}_i(est) = \hat{Q}_i - \Theta_{Ki}\hat{K}_i - \Theta_{Li}\hat{L}_i = \hat{A}_i(true) + \Theta_{Li}\,\hat{\bar{z}}_i$$
$$= \hat{A}_i(true) + \xi\Theta_{Li}\,\hat{\pi}_i\,.$$

If average worker efficacy depends inversely on a sector's share of the labor force ($\xi < 0$), growth accountants will systematically overestimate productivity growth in sectors whose labor share is contracting, such as goods industries, and systematically underestimate it in sectors whose labor share is expanding, such as services.

With the addition of two empirical assumptions it is possible to derive a simple expression for the goods and services relative supply curve. These assumptions, although not universal characteristics of the model, approximately characterize the US and OECD economies (see end of Section II below): (i) average wages per worker are proportional across sectors; and (ii) factor income shares are the same in the two sectors. Mathematically, these amount to

$$(8) \qquad W_G = w_G\,\bar{z}_G \propto w_S\,\bar{z}_S = W_S \quad \text{and} \quad \frac{rK_G}{W_G\,L_G} = \frac{rK_S}{W_S\,L_S},$$

$$\text{so } \hat{w}_G - \hat{w}_S = \hat{\bar{z}}_S - \hat{\bar{z}}_G \quad \text{and} \quad \hat{K}_G - \hat{L}_G = \hat{K}_S - \hat{L}_S,$$

where r is the common rental per unit of capital. Although the marginal worker is indifferent between working in the two sectors, average earnings per worker, $W_i = w_i\,\bar{z}_i$, depend on the inframarginal distribution of heterogeneous efficacy and need not necessarily equalize.[10] Thus (8) is an empirical assumption, rather than a theoretical prediction of the model.

Continuing, as $Q_i = A_i\,F_i(K_i, L_i\,\bar{z}_i)$ and $L_i = L\pi_i$, we have

$$(9) \quad \hat{Q}_G - \hat{Q}_S = \hat{A}_G - \hat{A}_S + \Theta_K(\hat{K}_G - \hat{K}_S) + \Theta_L(\hat{L}_G + \hat{\bar{z}}_G - \hat{L}_S - \hat{\bar{z}}_S)$$
$$= \hat{A}_G - \hat{A}_S + (\hat{\pi}_G - \hat{\pi}_S) + \Theta_L(\hat{\bar{z}}_G - \hat{\bar{z}}_S)$$
$$= \hat{A}_G - \hat{A}_S + (1/\xi + \Theta_L)(\hat{\bar{z}}_G - \hat{\bar{z}}_S)\,.$$

[9] A more refined practice is to differentiate workers into types based upon observable characteristics such as age and education. Within each type, however, marginal workers are treated as identical to average workers, producing the same problem, as I show when I extend the model further below.

[10] For example, when the z_i are draws from independent fréchet distributions average wages by sector always equalize, but when they are draws from exponential distributions they do not (see the online Appendix).

From the dual measure of productivity growth $\hat{A}_i = \Theta_K\,\hat{r} + \Theta_L\,\hat{w}_i - \hat{P}_i$,[11] so

$$\hat{P}_S - \hat{P}_G = \Theta_L(\hat{w}_S - \hat{w}_G) + (\hat{A}_G - \hat{A}_S) \tag{10}$$
$$= \Theta_L(\hat{\bar{z}}_G - \hat{\bar{z}}_S) + (\hat{A}_G - \hat{A}_S)\,.$$

Finally, substituting for $\hat{\bar{z}}_G - \hat{\bar{z}}_S$ using (9), we derive the Roy supply curve:

$$\hat{P}_S - \hat{P}_G = \left(\frac{-\Theta_L\xi}{1 + \Theta_L\xi}\right)\left(\hat{Q}_S - \hat{Q}_G\right) + \left(\frac{1}{1 + \Theta_L\xi}\right)\left(\hat{A}_G - \hat{A}_S\right) \quad [\text{Roy}]\,. \tag{11}$$

The first term on the right-hand side of (11) gives the slope of the supply curve; the second term gives the vertical shift associated with a change in relative total factor productivities. For $0 > \xi > -1$, the supply curve is upward sloping, as drawn in Figure 1 of the introduction. In the special case where $\xi = 0$ and average worker productivity does not vary with the sectoral employment share, labor is, for all intents and purposes, homogeneous and the supply curve reduces to

$$\hat{P}_S - \hat{P}_G = \hat{A}_G - \hat{A}_S \quad [\text{Baumol}]\,. \tag{12}$$

With P_S/P_G independent of Q_S/Q_G, this is, of course, Baumol's horizontal relative supply curve.

Equation (11) highlights the fact that, in the absence of differences in productivity growth rates, there is a limit to the relative price growth which can be explained by Roy's model of self-selection. With the labor share of two-thirds observed in the US and OECD economies, as ξ goes from 0 to -1 the slope parameter $-\Theta_L\xi/(1 + \Theta_L\xi)$ goes from 0 to 2. Thus, the Roy supply curve can be no steeper than 2: i.e., the historical growth of the relative output of services to goods has to be at least one-half the historical growth of the relative price if one wants to eliminate Baumol type effects from the story. As it so happens, the historical growth rates of relative goods and services outputs and prices in the United States and the OECD at large appear to be about equal (see Section II), which can be explained, in the absence of any differences in productivity growth, with a ξ of -0.75. This value is comfortably within the range of long-run estimates using defense spending as an instrument reported later in Section II.

B. *Comparative and Absolute Advantage and the Sign of* ξ

In the online Appendix I prove that sufficient conditions for ξ, the elasticity of average worker efficacy with respect to a sector's share of total employment, to be less than zero are that (i) the sectoral productivity draws z_i are independent of each other; and (ii) the elasticity of the cumulative distribution function for each of the draws, $(dG/dz) \times (z/G)$, is decreasing in the productivity of the draw.

[11] Totally differentiating $P_i\,Q_i = rK_i + w_i L_i \bar{z}_i$: $\hat{P}_i + \hat{Q}_i = \Theta_K(\hat{r} + \hat{K}_i) + \Theta_L(\hat{w}_i + \hat{L}_i + \hat{\bar{z}}_i)$. Substituting for $\hat{Q}_i$ gives the equation in the text.

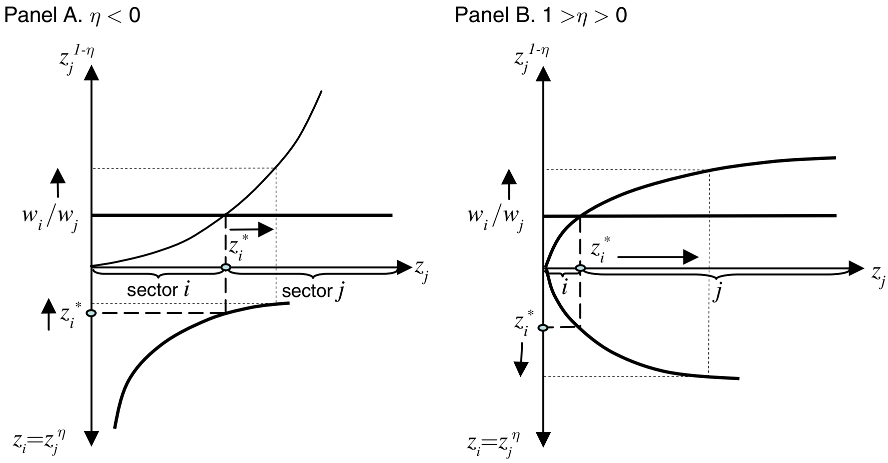

FIGURE 3. CORRELATED DRAWS, $z_i = z_j^{\eta}$

The latter characteristic is true of all of the popular distribution functions defined on nonnegative numbers: i.e., the chi-squared, exponential, F, fréchet, gamma, lognormal, pareto, rayleigh, uniform, and weibull distributions,[12] so I relegate a discussion of its role to the online Appendix. The assumption of independence is more problematic, so I explore its role here with a simple example and diagram.

Consider a two-sector example where the draw for sector i is deterministically related to that of sector j by the equation $z_i = z_j^{\eta}$, with z_j drawn from any distribution function. Workers will select sector i if $w_i z_i > w_j z_j$ or, equivalently, $w_i/w_j > z_j^{1-\eta}$. Figure 3 illustrates how the characteristics of the resulting equilibrium vary with η. Panel A considers the case where $\eta < 0$: i.e., the productivity draws are negatively correlated. The upper quadrant of the diagram shows that there exists a marginal draw z_j^* such that all workers with draws greater than z_j^* work in sector j and all workers with draws less than z_j^* work in sector i. The productivity of workers in sector i is illustrated in the lower quadrant, where the axis, despite its location below the horizontal line, should be read as representing positive numbers. With $\eta < 0$, the productivity of workers in sector i is negatively related to the z_j draws. $\overline{z}_j$ is given by the average of the workers to the right of z_j^*, while $\overline{z}_i$ is given by the average of the workers below (i.e., south of) $z_i^* = z_j^{*\eta}$. As w_i/w_j increases, sector j sheds workers with less than the average productivity in that industry, while sector i gains workers with less than the average productivity in that sector. Average productivity rises in the sector losing workers and falls in the sector gaining workers, so $\xi < 0$.

[12] While this condition may be true for all of the well-known distributions, I should note that it isn't hard to think of distribution functions where it is not. Thus, the distribution function $G(z) = (\exp(z) - 1)/(\exp(1) - 1)$ defined on $[0, 1]$ violates the condition and, in a simple two sector example, produces regions where the average productivity of workers in a sector is rising in the sector's share of total employment. I should also note that for the uniform distribution defined on $[a, b]$, for $a > 0$ the elasticity of the cumulative distribution is decreasing strictly in z but for $a = 0$ it is constant and a weaker form of the theorem applies (ξ is nonpositive).

Turning to panel B of the figure, we consider the case where the draws are positively correlated, so $1 > \eta > 0$.[13] With a positive relationship between z_i and z_j, $\overline{z}_j$ is once again given by the average of the workers to the right of z_j^*, but $\overline{z}_i$ now equals the average of the workers *above* (i.e., north of) z_i^*. As w_i/w_j rises, industry j sheds workers with less than the average productivity in that sector, but industry i gains workers with more than its average sectoral productivity. ξ is still negative for sector j, but it is now positive for sector i.

Returning to the trade terminology used in the introduction of this paper, if there is a positive correlation between comparative advantage and absolute advantage, then marginal workers entering or exiting an industry will have less than the average sectoral productivity. If, however, the correlation between comparative and absolute advantage is negative, marginal workers will have more than the average productivity. In panel A of Figure 3, workers who choose to work in industry i or j (a consequence of comparative advantage) are absolutely more productive in that sector than workers who choose to work in the other sector, so comparative advantage is positively correlated with absolute advantage. In panel B, this is true for sector j, but it is no longer true for sector i. In the case of sector i, workers who choose to work in the industry (those with z_i lying north of z_i^* on the vertical axis) are absolutely less productive in that sector than those who choose to work elsewhere (those with z_i lying south of z_i^* on the vertical axis), so comparative advantage is negatively correlated with absolute advantage.

Roy argued that if a worker's productivities in different sectors are independent of each other, then the marginal worker entering or exiting an industry will be less efficient than the average worker in that sector. The theorem described above and proven in the online Appendix shows that, modulo a technical density condition, Roy's conjecture is true. Figure 3 shows that positive correlation between an individual's productivity in different sectors undermines the association between comparative and absolute advantage, producing an indeterminate association between average and marginal productivities. In constructing total factor productivity growth estimates, as discussed shortly below, the growth accountant typically adjusts for observables such as age and education that create positive correlations in individual productivity across industries and tasks. These adjustments are, however, by no means exhaustive and it remains an empirical question whether or not comparative advantage is positively or negatively correlated with absolute advantage. The empirical results of the next section, interpreted in the light of the Roy model, provide some evidence in favor of the view that the elasticity of average worker efficacy with respect to a sector's employment share is negative: i.e., that by and large comparative and absolute advantage are indeed positively correlated.

C. *Practical Extensions*

A modest amount of notational and algebraic complexity must be added to the model to bring it to the data. To this end, imagine that there are N sectors with gross

[13] For the case $\eta > 1$, rearrange $z_i = z_j^\eta$ as $z_j = z_i^{1/\eta}$, rename i as j and j as i and proceed with panel B.

output in each sector i a function of J types of labor input and M types of other inputs:

$$(13)\quad Q_i = A_i F_i\left(\int_{u\in\mathcal{O}_i^1} z_i^1(u),\ \int_{u\in\mathcal{O}_i^2} z_i^2(u),\ldots,\int_{u\in\mathcal{O}_i^J} z_i^J(u),\ M_i^1,\ M_i^2,\ \ldots,\ M_i^M\right),$$

where I now use superscripts to denote the type of input and subscripts the industry. The switch from value added to gross output reflects the fact that my data sources, the BLS and Groningen KLEMS, measure total factor productivity growth at the sectoral level, using the gross output concept, so the list of M additional inputs moves beyond capital and includes intermediate inputs such as materials, services, and energy. Good estimates of total factor productivity growth typically adjust for "labor quality" by decomposing labor into mutually exclusive categories based upon observable determinants of human capital such as sex, age, and education. This decomposition not only produces more accurate measures of total factor productivity growth, it also implicitly controls for factors that produce a positive correlation in individual productivity across tasks, as noted above.

While the Groningen KLEMS adjust for labor quality, the BLS KLEMS measures do not adjust for labor quality, using only total labor hours as the measure of labor input. Using Current Population Survey data, I have constructed measures of labor input for each of the 60 KLEMS sectors cross-classified by sex, age (six categories), and education (five categories). I follow a methodology very similar to that used by the BLS in producing its measures of labor quality for the aggregate economy, using the CPS data to determine the distribution of workers by characteristic, but benchmarking the sectoral totals of hours and workers using the BLS Current Employment Statistics data. Details are provided in the online Appendix. I use these estimates to adjust the BLS TFP growth measures for the changing composition of the workforce and to calculate the changing shares of workers by characteristic, as in (15) below. The main results, however, can just as easily be found with the unadjusted BLS data, as reported in footnotes later.

To extend the model to this environment, let each worker of type j be endowed with a set of N industry productivities $(z_1^j, z_2^j, \ldots, z_N^j)$ drawn from some joint distribution function and let w_i^j denote the wage per unit of effective labor of type j in industry i. A worker of type j chooses to work in sector i if $w_i^j z_i^j(u) > w_k^j z_k^j(u)\ \forall\ k \neq i$. Total factor productivity growth in each sector is given by

$$(14)\qquad \hat{A}_i(true) = \hat{Q}_i - \sum_j \Theta_{Li}^j(\hat{L}_i^j + \hat{\bar{z}}_i^j) - \sum_m \Theta_{Mi}^m \hat{M}_i^m,$$

where L_i^j is the number of workers of type j employed in sector i, $\bar{z}_i^j$ is their average efficacy, and the Θ_{Li}^j and Θ_{Mi}^m represent the gross output factor income shares of workers of type j and other inputs of type m in sector i, respectively. Conventional measures of total factor productivity growth, by ignoring changes in the average efficacy of workers, have a bias equal to:

$$(15)\quad \hat{A}_i(est) = \hat{Q}_i - \sum_j \Theta_{Li}^j \hat{L}_i^j - \sum_m \Theta_{Mi}^m \hat{M}_i^m = \hat{A}_i(true) + \sum_j \Theta_{Li}^j \hat{\bar{z}}_i^j$$
$$= \hat{A}_i(true) + \xi \sum_j \Theta_{Li}^j \hat{\pi}_i^j.$$

Growth accounting calculations assume intrinsically that all workers of a given type are the same. Unless the list of observable worker characteristics completely exhausts the determinants of individual productivity, the productivity of the marginal worker entering or exiting an industry will generally be different than that of the sectoral average for that type of worker. If the elasticity of average worker efficacy with respect to the employment share is negative ($\xi < 0$),[14] conventional growth accounting will under- or overstate productivity growth in sectors with expanding or contracting employment shares, respectively.

Finally, I note that the gross output TFP measures of multiple subsectors can be combined to form goods and services value added aggregates using the formula:

$$\hat{A}_j = \sum_{i \in \mathscr{I}(j)} \frac{VA_i}{GDP_j} \frac{GO_i}{VA_i} \hat{A}_i \quad \text{where} \quad GDP_j = \sum_{i \in \mathscr{I}(j)} VA_i, \tag{16}$$

and where $j =$ goods or services and $\mathscr{I}(j)$ is the set of subsectors in j. TFP measures calculated using the gross output approach equal TFP measures calculated using the value added approach times the ratio of the value of gross output to value added (GO_i/VA_i), so (16) converts subsectoral gross output TFP measures to value added TFP measures and aggregates to sectoral totals by weighting by shares of sectoral value added. I use this measure to summarize goods and services productivity growth further below.

II. Industry Evidence on the Elasticity of Worker Efficacy with Respect to Employment Shares

A. *Empirical Specification*

I use the following two-stage least squares (2SLS) specification to explore the bias in sectoral measures of total factor productivity growth brought about by changing labor allocations:

$$\hat{Y}_{ict} = \alpha_{ic} + \delta_{ct} + \gamma_{ic} \hat{U}_{ct} + \xi \hat{X}_{ict} + \varepsilon_{ict} \tag{17}$$

$$\hat{X}_{ict} = \alpha^X_{ic} + \delta^X_{ct} + \gamma^X_{ic} \hat{U}_{ct} + \beta_{ic} \hat{Z}_{ct} + \eta_{ict} \quad E(\varepsilon_{ict} \eta_{ict}) \neq 0,$$

where $\hat{Y}_{ict}$ is total factor productivity growth in industry i of country c in period t, the α_{ic} are industry $\times$ country dummies capturing mean productivity growth by sector, and the δ_{ct} are country $\times$ year dummies capturing economy-wide fluctuations in average productivity growth. There is a well-known association between the business cycle and measured productivity growth, driven perhaps by mismeasurement due to changes in capacity utilization and the role real technology shocks play

[14] Equation (15) assumes that ξ is the same for all sectors and types at all times. This is true precisely for some distribution functions (e.g., independent draws from fréchet distributions with the same dispersion parameter). Otherwise, one must take ξ as an average of the differing elasticities.

in producing the business cycle. While the country $\times$ year dummies account for mean economy-wide changes, the ln change in the national unemployment rate $\hat{U}_{ct}$, entered separately by industry $\times$ country (γ_{ic} is an industry $\times$ country effect), corrects for the cyclical variation in relative industry productivity growth which might otherwise appear as correlation with other variables. Finally, $\hat{X}_{ict}$ equals the labor-income-share-weighted sum of the change in national employment shares by worker type, as shown in the right-hand side of (15) earlier. The coefficient ξ, by the theory described earlier above, is the elasticity of worker efficacy with respect to employment shares, the principal object of interest in the regression.[15]

The OLS (ordinary least squares) relation between productivity and employment shares potentially has both exogenous and endogenous components. On the one hand, movements in relative industry demand, due to the growth of aggregate income and non-homothetic preferences, will lead to exogenous changes in relative employment shares. On the other hand, the response of relative demand to relative price movements brought about by productivity growth may lead to an endogenous response of employment shares to productivity growth. There are special cases where these effects disappear, such as with homothetic utility and unitary income elasticities of demand (no exogenous variation of relative demand) or with Hicks-Neutral technical change and unitary price elasticities of demand (no endogenous variation of factor allocations with sectoral productivity growth), but it seems reasonable to allow for the existence of both in the data.[16] As shown in the second line of (17), to correct for potential endogeneity I run a first-stage regression in which the labor-income-share-weighted changes in sectoral employment shares are regressed on the exogenous variables of the total factor productivity equation plus an excluded instrument. The relation of the excluded instrument with $\hat{X}_{ict}$ is allowed to vary across industries and countries (β_{ic} varies by industry $\times$ country). Variation by industry is necessary, as for an instrument to influence employment shares it must raise employment in some industries at the expense of others, and variation by countries allows for differences in the composition of otherwise nominally "identical" sectoral aggregates. Because the instrument is interacted by industry $\times$ country (i.e., appears multiple times in the regression), it is possible to perform a valid overidentification test of the exclusion restriction, even though only "one" instrument appears in the regression.[17]

[15] This specification estimates a single ξ, but should be compatible with a world in which ξ_i varies by industry and we are estimating an average effect, as the panels are balanced (having the same number of observations for each industry) and there are industry dummies, so ξ is being estimated by the equally weighted variation (exclusive of the business cycle) within industries in rates of employment share changes. I should note that estimating ξ industry by industry is not sensible, as the resulting sample sizes are tiny (e.g., 20+ observations per industry in the United States), while 2SLS relies on asymptotics.

[16] Ngai and Pissarides (2007) provide an analysis of the case with homothetic utility, Hicks-Neutral technical change, and inelastic demand, where all of the relation between labor allocations and productivity is endogenous. Homothetic utility, however, provides a poor characterization of demand, as it implies that relative quantities fall with relative prices whereas, as discussed below, the overwhelming trend in the OECD is for relative quantity to rise with relative price (reflecting non-unitary income elasticities). Hicks-Neutral technical change misses interesting interactions between factor-biased technical change and the elasticity of substitution. For example, Bustos, Caprettini, and Ponticelli (2013) show that despite an infinite elasticity of demand (free trade), labor augmenting technical change in the presence of a low elasticity of factor substitution can actually lead to a reduction in sectoral employment.

[17] Lest the reader think there is an error here, I confirm the distribution of the overidentification test using simulated data which satisfy the exclusion restriction, as discussed further below.

I draw on two datasets which provide comprehensive measures of private sector total factor productivity broken down by sector ($\hat{Y}_{ict}$ above). First, I use data on total factor productivity growth by sector drawn from the Bureau of Labor Statistics' KLEMS (capital, labor, energy, materials, and business services) database, which provides estimates of US private sector productivity growth disaggregated into 60 comprehensive industries from 1987 to 2010.[18] As noted earlier, these data do not adjust for the changing composition of the labor force, so I use Current Population Survey data to develop industry-level measures of the distribution of workers by sex $\times$ age $\times$ education and use these to adjust the total factor productivity growth and calculate a compositionally adjusted measure of changing labor shares, as described in the online Appendix.[19] Second, I use the EU KLEMS database, developed by the University of Groningen with a consortium of diverse partners, which divides private sector productivity growth in a variety of advanced economies into 29 comprehensive sectors.[20] After removing transition economies, where productivity growth and factor allocations are likely to be driven by considerations outside this paper, the sample consists of 18 countries, namely: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Italy, Japan, Korea, Luxembourg, Netherlands, Portugal, Spain, Sweden, the United Kingdom, and the United States. The productivity estimates run from 1970 to 2005, with the available years varying by country. I shall refer to these data as the OECD or OECD 18 sample, notwithstanding their development in the European Union. Measures of annual unemployment for the United States and the OECD countries are drawn from the Federal Reserve Bank of St. Louis, FRED database.

Turning to potential instruments, I consider simple measures of my own alongside the more sophisticated constructions of others. Using FRED, Stockholm International Peace Research Institute (SIPRI) and World Bank data, the instruments I prepare are: (i) the ln change in country defense expenditures over GDP; (ii) the average ln change in metal prices (aluminum, copper, iron ore, lead, nickel, platinum, tin, and zinc); and (iii) the average ln change in oil prices (Dubai and West Texas Intermediate). Changes in defense expenditures, driven by events such as the collapse of the Soviet Union and 9/11, are arguably exogenous to sectoral

[18] Calculating industry-level TFP estimates for the United States is a nontrivial task. Whereas most countries report capital formation by industry of use, the United States reports these by industry of ownership. Marrying these data to the output data and ensuring that the proper value added reallocations are being made in the national accounts, while simultaneously dealing with historical changes in sectoral definitions, requires a great deal of inside information. The BLS, with its official status and resources, is well positioned to have access to the requisite data and knowledge. Given all of the difficulties involved, however, it is not surprising that the BLS, while producing aggregate private sector numbers going back to the early postwar era, has only been able to extend the comprehensive sectoral breakdown back to 1987.

[19] The adjusted and unadjusted industry measures of total factor productivity growth are available on my website. My calculations indicate that adjustments for the changing sex $\times$ age $\times$ education composition of the labor force lower economy-wide private sector total factor productivity growth between 1987 and 2010 from an average of 1.25 percent per annum to 0.97 percent per annum.

[20] There are actually 31 private sectors, but 2 ("private households with employed persons" and "extra-territorial organizations") are relatively minor and do not appear in all instances. Employment shares are always calculated relative to national totals (including the public sector). Although the EU KLEMS TFP calculations adjust for the composition of the workforce, the data provided only allow for the calculation of the distribution of total workers by sector (not workers by type), so I use the labor income share times the change in the total employment share as the X variable, as in equation (7) above. Equations (7) and (15) are identical if the distribution of workers by type is proportional to the industry share of total employment: i.e., $L_i^j = L^j(L_i/L)$. For the US KLEMS, I find that substituting the changing shares of total employment for the changing shares of employment by worker type yields virtually identical results, as reported in a footnote below.

productivity growth. There is less reason to feel confident in the exogeneity of metals and oil prices. Productivity change in key producing or using industries in the US and the OECD countries, which are large actors in the global markets for these materials, might produce endogenous responses in prices. While US defense spending and materials and oil prices are available for all years of my TFP data, because of changes in concepts and coverage, the SIPRI data on OECD country military expenditures only extend back to 1988.[21]

I expand the list of potential instruments by adding all 15 of the nontechnology shock instruments considered by Stock and Watson (2012) in their dynamic factor model analysis of the US economy. Covering oil prices, monetary policy, uncertainty, liquidity, and fiscal policy, these are:[22] (i) Hamilton's (2003) measure of the increase of the oil price PPI relative to the max of the previous three years, available for 1962–2010; (ii) Kilian's (2008) measure of the OPEC production shortfall from wars and civil strife, available for 1971–2004; (iii) the residuals of Ramey and Vine's (2011) measure of full gasoline prices regressed on lagged macroeconomic variables, based on their updated spreadsheet (available 1959–2011); (iv) Romer and Romer's (2004) residual of Fed monetary intentions regressed on internal Fed forecasts (1969–1996); (v) Smets and Wouters' (2007) measure, updated by King and Watson (2012), of the shock to the monetary policy reaction function in a dynamic stochastic general equilibrium model (1959–2004); (vi) Sims and Zha's (2006) monetary policy shock estimated in a structural VAR (vector autoregression) (1960–2002); (vii) Gürkaynak, Sack, and Swanson's (2005) measure of surprise changes in the federal funds rate (1990–2004); (viii) innovations in an AR(2) of the VIX, as suggested by Bloom (2009) (1962–2011); (ix) innovations in an AR(2) of Baker, Bloom, and Davis' (2012) policy uncertainty index calculated from media references to economic policy (1985–2011); (x) innovations in an AR(2) of the TED spread, as provided by Stock and Watson (1971–2011); (xi) innovations in an AR(2) of Gilchrist and Zakrajšek's (2012) bond premium (1973–2010); (xii) Bassett et al.'s (2012) measure of unpredictable changes in bank-level lending standards (1992–2010); (xiii) Ramey's (2011) measure of news of changes in the net present value of military spending divided by nominal GDP (1959–2010); (xiv) Fisher and Peters' (2010) measure of excess returns on stocks of military contractors (1959–2008); and (xv) Romer and Romer's (2010) measure of tax changes relative to GDP (1959–2007). I average quarterly or monthly shocks to annual levels.

With the exception of Kilian's oil production shortfall, the Stock and Watson instruments listed above are US-centered and not appropriate for an OECD analysis. However, as shown in the pages below, none of these instruments performs at all well in the analysis of the US KLEMS. Hence, undertaking the monumental task of developing similar instruments country by country is not likely to be profitable. In

[21] The SIPRI website notes that SIPRI has not been able to construct a consistent series extending back to earlier dates, and the SIPRI data has now become the standard, reproduced in other online sources (such as the World Bank) to the exclusion of any other information. I tried to construct an alternative series of my own using historical paper issues of The Military Balance, but ultimately concluded that SIPRI's concerns about coverage and data quality are correct.

[22] In most cases I use the data provided online by Stock and Watson (2012) and follow their procedures (e.g., AR(2)s, regressions on lagged macro-variables, etc.) to construct the instruments. The dataset, however, contains a major misreporting of the Ramey-Vine figures (formulas rather than values were copied into the Stock and Watson spreadsheet), so I use the updated data from Ramey and Vine (2011).

fact, the only instrument which consistently satisfies the first-stage requirement of significance and the second-stage exclusion restriction is defense spending. Thus, my main point in using Stock and Watson's (2012) extensive list is to highlight the difficulty of finding alternative instruments for sectoral labor allocations.

B. *Results*

I begin by evaluating the suitability of the various instruments to the problem at hand. In Table 1 I run the first-stage regression of the specification of equation (17) using one instrument at a time, reporting the *p*-value of the *F*-test on the instrument[23] and the total number of observations. In the case of the OECD, I only use my instruments and Kilian's oil production shortfall, which can be considered part of global trends. There are two notable aspects of Table 1. First, virtually all of the factors considered by Stock and Watson (2012) (instruments d through r) are not meaningful determinants of labor allocations. Only the oil price max measure and Federal Funds surprises are significant at the 5 percent level, and these results are suspect as other measures of oil prices and monetary policy are quite insignificant. Second, in the OECD sample none of the instruments are even close to being significant.

Table 1's results are perhaps not terribly surprising. To generate a significant reallocation of labor across sectors, an instrument must not merely shift macroeconomic supply and demand, it must substantially alter relative industry supply or demand away from the norm. Many shocks which have strong aggregate macroeconomic consequences and serve as good instruments for the analysis of macro-aggregates might not have sufficiently strong relative effects for the objective of this paper. In this regard it is noteworthy that Ramey's (2011) measure of news of changes in the NPV (net present value) of military spending is insignificant. Ramey argues that, in explaining changes in macroeconomic aggregates in the United States, her news variable dominates actual defense spending changes. The macroeconomic influence of Ramey's news variable, however, most likely represents the response of private economic actors to the foreseen aggregate consequences (e.g., on demand and tax burdens) of that spending. Continuity of private demand suggests that these are unlikely to have large effects on the distribution of economic activity, even if they affect levels. In contrast, actual defense spending shifts the pattern of demand away from the private norm, resulting in more significant changes in sectoral employment shares. Thus, Ramey's finding for macroeconomic aggregates need not extend to my analysis of labor allocations. When entered jointly with actual defense spending changes in the first-stage regression for the US, I find the *p*-value on the *F*-test of Ramey's news variable to be 0.313, while that on actual defense spending changes remains 0.000. The insignificance of defense spending in the OECD regressions stems from the fact that for 3,358 of the 8,049 observations defense spending changes are zero. Defense spending as a share of GDP is extremely stable in most OECD countries and, with low values and one decimal precision in the SIPRI data, the sudden changes that do occur are most likely reflective of rounding error (e.g., moving from 0.9 percent of GDP previously to 1.0 percent ever after in one year in Japan).

[23] Although in each case there is only one instrument, its coefficient is allowed to vary by industry $\times$ country, hence an *F*-test rather than a *t*-statistic.

Table 1—First Stage p-value in Regression of Weighted Employment Share Changes on Instruments
(*Instruments evaluated one at a time using specification of equation* (17))

	United States 60 sectors, 1987–2010		OECD 18 29 sectors, 1970–2005	
	F p-value	N	F p-value	N
(a) Δ ln country defense expenditures/GDP	0.000	1,380	0.192	8,049
(b) Δ ln metals prices	0.000	1,380	0.374	12,109
(c) Δ ln oil prices	0.833	1,380	0.367	12,109
(d) Oil price increase over prior maximum (Hamilton 2003)	0.005	1,380		
(e) OPEC oil production shortfall (Kilian 2008)	0.253	1,020	0.762	11,617
(f) Residual of US gasoline prices (Ramey and Vine 2011)	0.965	1,380		
(g) Monetary policy shock (Romer and Romer 2004)	0.866	540		
(h) Monetary policy reaction shock (Smets and Wouters 2007)	0.084	1,020		
(i) Monetary policy shock (Sims and Zha 2006)	0.884	900		
(j) Fed. funds surprises (Gürkaynak, Sack, and Swanson 2005)	0.000	900		
(k) VIX innovation (Bloom 2009)	0.863	1,380		
(l) Policy uncertainty index innovation (Baker, Bloom, and Davis 2012)	0.092	1,380		
(m) TED spread innovation (Stock and Watson 2012)	1.000	1,380		
(n) Bond premium innovation (Gilchrist and Zakrajšek 2012)	1.000	1,380		
(o) Bank lending shocks (Bassett et al. 2012)	0.992	1,140		
(p) NPV defense spending news/GDP (Ramey 2011)	0.104	1,380		
(q) Excess returns on defense stocks (Fisher and Peters 2010)	0.432	1,260		
(r) Tax changes/GDP (Romer and Romer 2010)	0.108	1,200		

Notes: F p-value = F-test p-value on the industry × country coefficients associated with the instrument. N = observations, sample changes with the availability of the instrument. Instruments (d)–(r) calculated using data from Stock and Watson (2012); instruments (a)–(c) based upon FRED, SIPRI, and World Bank data, as described in the text. Each regression follows the first stage specification given in (17), with industry × country and country × year fixed effects and the national unemployment rate change and instruments entered separately for each industry × country. The dependent variable is the labor-share-weighted change in the share of employment by worker type (for the US) or total industry workers (OECD 18, see footnote 20 above). Each row represents a separate analysis with the indicated instrument alone.

Table 2 presents second-stage results using each of the four instruments which are significant at the 5 percent level in the first-stage regressions for the United States in Table 1 (EU KLEMS results are presented later). Aside from the estimate of ξ, the elasticity of worker efficacy with respect to the sectoral employment share, I also report the p-value of the first-stage F-test (which will vary across specifications) and the second-stage χ^2 overidentification test.[24] In the top panel, which follows

[24] As noted earlier, an overidentification test is possible with one instrument because, since it is entered separately for each industry, there are technically actually I (equal to the number of industries) instruments. The overidentification test is whether these instruments have any predictive value in the regression beyond their association with changes in employment shares. As the reader might worry that this is somehow econometrically wrong, I have used Monte Carlo simulations to confirm the accuracy of the test statistic. Using the covariance of the residuals from the first- and second-stage regressions of the baseline specification with defense expenditures, I produce 500,000 simulated draws of the data under the assumption that (modulo their influence on employment shares) defense expenditures are exogenous in the second-stage regression. The resulting test statistic is nearly exact: i.e., the nominal rejection values are very close to the actual rejection probabilities (see the next paragraph).

I should also note that Basmann (1960) argues that the standard (Sargan 1958) χ^2 overidentification test is too conservative (i.e., rejects the null too frequently) in finite samples and proposes a small sample adjustment to the test statistic. I have confirmed his argument, for my case, using the Monte Carlo simulations described above. I find

TABLE 2—ANNUAL TFP GROWTH ON CHANGES IN EMPLOYMENT SHARES
(*United States: 60 sectors* × *1987–2010*)

	OLS	2SLS by type of instrument			
		Δ Defense spending	Δ Metals prices	Oil price maximum	Fed funds surprises
Panel A. Baseline specification (*equation* (17))					
ξ (SE)	−0.218 (0.108)	−0.922 (0.266)	−0.546 (0.318)	0.372 (0.384)	−0.468 (0.318)
F and χ^2 p-value		0.000, 0.148	0.000, 0.004	0.005, 0.000	0.000, 0.000
N/K/L	1,380	1,380/199/59	1,380/199/59	1,380/199/59	900/191/59
Panel B. Dropping unemployment controls by industry (*business cycle adjustment*)					
ξ (SE)	−0.167 (0.100)	−0.359 (0.226)	−0.245 (0.452)	0.359 (0.396)	−0.742 (0.412)
F and χ^2 p-value		0.000, 0.031	0.440, 0.000	0.033, 0.000	0.371, 0.002
Panel C. Substituting ln changes in capacity utilization for unemployment controls					
ξ (SE)	−0.240 (0.100)	−0.689 (0.222)	−0.465 (0.346)	0.363 (0.375)	−0.654 (0.343)
F and χ^2 p-value		0.000, 0.009	0.003, 0.478	0.029, 0.000	0.044, 0.950
Panel D. Adding ln changes in capacity utilization to unemployment controls					
ξ (SE)	−0.207 (0.109)	−0.771 (0.254)	−0.457 (0.332)	0.372 (0.364)	−0.596 (0.319)
F and χ^2 p-value		0.000, 0.260	0.000, 0.427	0.003, 0.000	0.000, 0.663
Panel E. Dropping country × *year dummies* (*common component of TFP growth*)					
ξ (SE)	−0.257 (0.107)	−1.03 (0.263)	−0.738 (0.318)	0.372 (0.390)	−0.541 (0.317)
F and χ^2 p-value		0.000, 0.146	0.000, 0.001	0.007, 0.000	0.000, 0.000
Panel F. Dropping one industry at a time					
Max ξ (SE)	−0.119 (0.107)	−0.812 (0.264)	−0.300 (0.325)	0.636 (0.441)	−0.045 (0.315)
Min ξ (SE)	−0.328 (0.113)	−1.13 (0.312)	−0.915 (0.318)	−0.007 (0.386)	−0.872 (0.363)
Max F p-value		0.000	0.015	0.048	0.003
Min F p-value		0.000	0.000	0.001	0.000
Max χ^2 p-value		0.582	0.067	0.004	0.000
Min χ^2 p-value		0.075	0.001	0.000	0.000
Panel G. Adding four lags of employment share changes					
$\sum \xi$ (SE)	−0.685 (0.209)	−0.750 (0.283)	−0.547 (0.338)	−0.233 (0.348)	−0.621 (0.359)
F and χ^2 p-value		0.000, 0.068	0.048, 0.002	0.083, 0.009	0.000, 0.000

Notes: ξ (SE) = coefficient (standard error) on labor-share-weighted changes of employment shares by worker type. F and χ^2 p-value = p-value on first-stage significance and second-stage overidentification tests. N/K/L = number of observations/number of regressors in first-stage/excluded instruments in second stage. Because of the joint year and industry dummies, one of the industry coefficients for each of the variables entered by industry (i.e., unemployment and capacity changes and instruments) is colinear with other variables and is dropped in all specifications other than those without year dummies. Thus, there are only 59 excluded instruments in the baseline specification. $\sum \xi$ = sum of the coefficients on current and four lags of weighted employment share changes.

the baseline specification of equation (17), three of the four instrumental variables estimates of ξ are substantially negative, although the only statistically significant estimate is that found using defense expenditures. Defense spending, however, is the only instrument which does not strongly reject the second-stage exclusion restriction. I confirm the likely endogeneity of the oil price instrument by correlating its first-stage industry coefficients with the average energy share of gross output in

Sargan's χ^2 test to be grossly conservative (rejecting, as examples, 15.5 percent of the time at the 5 percent level and 4.6 percent of the time at the 1 percent level), while Basmann's small sample correction is only slightly conservative (rejecting 5.8 percent of the time at the 5 percent level and 1.3 percent of the time at the 1 percent level). Consequently, throughout this paper I use Basmann's statistic as the overidentification test.

those industries. If this instrument represents exogenous shifts in prices, then its effect should be substantially negatively correlated with the energy intensity of production: i.e., industries which are more energy intensive should see their relative employment share fall with exogenous increases in oil prices, as their supply curves shift up. In practice, I find a correlation coefficient of 0.232. While not significant (p-value $= 0.077$), the correlation is of the wrong sign. This might occur if some of the increases in the price of oil represent an endogenous positive response to rising energy demand in using industries. In sum, of 18 potential instruments, only 1 (defense expenditures) satisfies the dual requirements of first-stage significance and second-stage exogeneity, and that instrument produces a strongly negative (-0.922) estimate of ξ.

The lower panels of Table 2 examine the sensitivity of the results to the specification. In panel B I remove the unemployment rate entered by industry. This has a very large impact on the estimates, dramatically reducing the estimate of ξ for both defense expenditures and metal prices, raising it for Federal Funds surprises, and rendering both metals prices and Fed surprises completely insignificant in the first-stage regression. In panel C I substitute the Federal Reserve's estimate of aggregate mining, manufacturing, and utilities capacity utilization for the unemployment rate, interacting it by industry as was done for unemployment. As shown, this moves ξ back to the estimates of panel A, although the value using defense expenditures (-0.689) is less extreme than in the baseline specification (-0.922). The Fed's measure of capacity utilization, however, does not exhaust the association of industry productivity and labor allocations with the business cycle. Adding the measure of aggregate capacity utilization to the baseline specification with unemployment and defense spending, I find that the industry coefficients on the unemployment rate in both the first- and second-stage regressions remain highly significant (F p-values of 0.000 and 0.003, respectively), suggesting that the business cycle characteristics of relative industry productivity and employment may go beyond capacity utilization and mismeasurement to something real. The estimate of ξ from defense spending in this specification is -0.771 (panel D). In general, controlling for the association between the business cycle and relative labor allocations and productivity seems appropriate[25] and this matters in the regression because the correlation between defense spending changes and changes in the unemployment rate in this time period is quite strong (0.649 with a p-value of 0.001). Nevertheless, the reader looking to see whether the defense spending results can be rendered insignificant need look no further than panel B. Panel E of Table 2 shows that removing the year dummies, but retaining the unemployment controls, generally increases the magnitude of ξ, with the negative estimate using metals prices now appearing significant.

Panel F of Table 2 explores whether identification and significance come from one particular industry by rerunning the baseline specification 60 times, removing one industry each time, and reporting the maximum-minimum range of the estimates of ξ and the F and χ^2 p-values. As shown, the estimates of ξ based upon the nondefense instruments vary enormously, but the range for defense expenditures is much more limited. Also of note is the stability of the first- and second-stage tests for defense

[25] To see this, the reader might introspect and consider their reaction if I had informed them that the estimate of ξ was substantially negative, but only when measures of the business cycle are *excluded* from the regression.

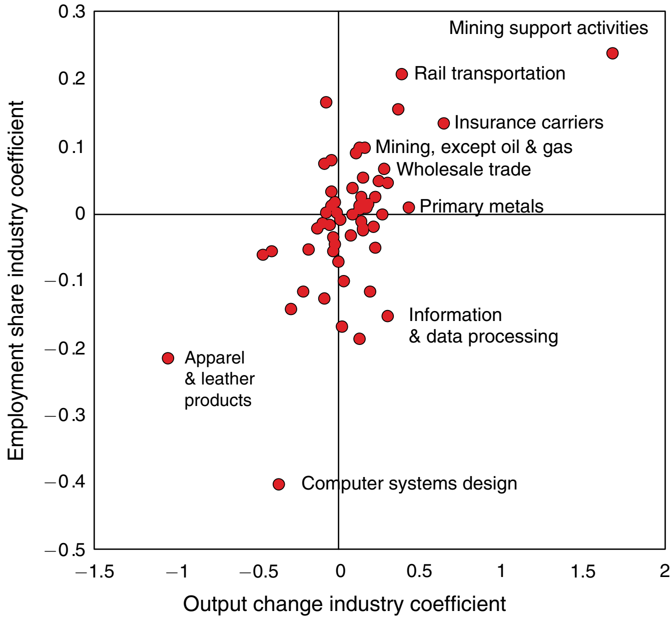

FIGURE 4. FIRST-STAGE REGRESSION ON DEFENSE EXPENDITURES

expenditures. Regardless of which industry is removed, defense spending is always found to be highly significant in the first-stage regression and exogenous in the second-stage overidentification test. In fact, removing all possible combinations of two and even three industries, the first-stage p-value on defense spending never rises above 2.3×10^{-8}, the p-value on its second-stage overidentification test never falls below 0.011, and the coefficient never becomes less negative than -0.590 (0.274). Thus, the correlations between defense expenditures, employment, and productivity which lie behind the significant coefficients reported in the top panel of Table 2 go far beyond one, two, or even three key industries.

Figure 4 provides further insight into the variation identifying the coefficients associated with defense expenditures reported in Table 2. For the horizontal axis, I project annual KLEMS industry output growth on industry dummies, time dummies, changes in the national unemployment rate (entered by industry), and changes in defense expenditures over GDP (entered separately by industry). The coefficients reported in the figure are the industry-defense expenditures relationships.[26] For the vertical axis, I run the same specification and report the same type of coefficients, but this time using labor-share-weighted changes in employment shares by type as the dependent variable. Thus, the figure compares the defense expenditure coefficients for the first-stage regression of the results reported above with the same first-stage regression run with output growth as the dependent variable. What the figure shows is that the two sets of coefficients are highly correlated ($\rho = 0.592$, p-value

[26] Given the year and industry dummies in the regression, the defense spending coefficients are not identified for one industry (the base), which I take as Food and Beverage and Tobacco Products. Thus, the coefficients reported in the figure are changes relative to that industry.

$= 0.000$), even when the four outliers with employment change coefficients greater than 0.2 in absolute value are removed ($\rho = 0.357$, p-value $= 0.007$).[27] Thus, the first-stage regressions underlying the results reported above appear to be based upon something real. Changes in defense expenditures change the demand for the output of industries, inducing changes in their employment shares.

The estimates using annual data in Table 2 might not provide an accurate representation of long-run effects. On the one hand, it is possible that short-run coefficients overstate the negative influence of the employment share on sectoral productivity as workers entering a sector are likely to be less productive initially than they will be in the long run, once they acquire sector-specific human capital. On the other hand, it is possible that short-run coefficients actually understate the negative effect of the employment share on sectoral productivity. Worker reallocations come about through changes in equilibrium output, either due to a shift of supply or demand. A sudden increase in output will lead to an influx of workers and, typically, a transitory rise in capacity utilization, producing a transitory overstatement of productivity.[28] Thus, this mismeasurement of productivity will be positively correlated with the movement of workers into a sector, understating the negative influence this otherwise has on measured productivity.[29]

Panel G of Table 2 addresses the issue of long-run effects by adding four lagged values of the labor-share-weighted change in employment shares as predetermined exogenous right-hand side variables to the baseline specification, with current employment reallocations instrumented with the instrument specified in each column. The cumulative effect on long-run-measured productivity is given by the sum of the current and lagged coefficients, which is presented in the table. Comparing these with the baseline results at the top of the table, one sees that ξ is now somewhat smaller in magnitude in the defense expenditures analysis (-0.750 versus -0.922 earlier), while the oil price maximum, which earlier reported an insignificant positive coefficient, is no longer first-stage significant and now produces a negative point estimate of ξ. The metals prices coefficient is unchanged, while that for Fed surprises is more negative.

Table 2 also reports OLS results, running each specification without instruments. Although the baseline OLS relation between employment share changes and productivity (-0.218) is small, the long-run cumulative association, as evidenced by panel G of Table 2's regression with lags of past employment changes, is much more negative (-0.685). It is difficult to explain how past employment changes relate negatively to current productivity growth within a framework where employment shares reflect the endogenous response of demand to shifts of the supply curve

[27] Not shown in the figure, however, is that the average t-statistic of the coefficients on the horizontal axis is 0.63 and the average t-statistic of the coefficients on the vertical axis is 0.92. Thus, while defense expenditures are overall very significantly correlated with industry output growth and employment share changes (F-tests), the estimated relationship, industry by industry, is quite imprecise.

[28] This applies even for instruments which shift the supply curve, provided they satisfy the exclusion restriction: i.e., are not directly correlated with total factor productivity growth. If something shifts the supply curve down without changing fundamental productive capacity, it will lead to an expansion of output which, along with the rise in the employment share, should produce a transitory increase in capacity utilization.

[29] As there are now a variety of mismeasurements, I should clarify. The object of interest in this paper is the mismeasurement of productivity due to the failure to account for the changing efficacy of workers as a sector's employment share expands. The transitory mismeasurement due to capacity utilization, however, works in the opposite direction and may temporarily conceal the effect I'm studying.

brought about by productivity change. The result is easier to comprehend, however, if one moves to a framework where changes in employment shares reflect exogenous shifts of the demand curve brought about by nonunitary income elasticities of demand and other shocks to relative demand. When demand shifts out in an industry, it produces a transitory rise in capacity utilization and a spurious rise in productivity, minimizing the negative effect of employment shifts on measured productivity. Over time, however, capacity adjusts and the full impact is revealed. Evidence in favor of this argument can be found by regressing total factor productivity growth on industry output growth, with industry, year, and unemployment $\times$ industry controls as in the baseline specification. With only current output growth in the regression, the OLS coefficient (standard error) for my 60 industry sample is 0.219 (0.025). With four lags of past output in the regression, the cumulative OLS coefficient is 0.076 (0.058). Thus, past output increases, like past employment increases, lead to lower current productivity growth, which is consistent with the utilization story outlined above.

Table 3 supports the preceding argument using the Federal Reserve Board's industry-level measures of capacity utilization, defined as current output over maximum sustainable output.[30] These measures are only available for the 22 mining, manufacturing, and utilities industries in the 60 sector KLEMS disaggregation of private sector activity. In this table I run regressions with either the change in capacity utilization or total factor productivity growth as the "Y" variables, and either the growth of output or the labor-income-share-weighted change in the share of economy-wide employment by type, the right-hand side variable of interest in the regressions reported above, as the "X" variable. Each regression includes a complete set of industry and time dummies and the change in the unemployment rate entered separately by industry, as in the baseline specification of equation (17). Aside from results with the current value of "X" alone, I also report the cumulative sum of the coefficients in a specification with the current value and four predetermined lags of "*X*."

I begin by taking both X variables as exogenous, running OLS specifications in panels A and B of Table 3. In the first two columns we see that an increase in current output raises both capacity utilization and measured total factor productivity growth, but that the cumulative long-run effect, once lags are allowed, is insignificantly different from zero in both cases. In the third column of the table, we see that a 1 percent increase in a sector's labor-income-share-weighted employment share is associated with a large 1.7 percent short-run rise in capacity utilization, but has no long-run effects. Regarding measured TFP, in the fourth column of Table 3, an increase in the sectoral employment share has no significant short-run impact on productivity, but a very large (-1.1) long-run effect. These results are completely consistent with a view of exogenous demand fluctuations producing transitory

[30] These measures are based upon the Survey of Plant Capacity and are defined as "the greatest level of output the plant can maintain within the framework of a realistic work schedule after factoring in normal downtime and assuming sufficient availability of inputs to operate the capital in place." (Gilbert, Morin, and Raddock 2000, p. 194). The survey measures are then regressed on a time trend, ln capital and dummies which correct for outliers. This suggests that the reported series is basically a smoothed version of the original data, allowing outliers that the Fed believes represents real changes.

Table 3—Response of Capacity Utilization and Productivity to Output and Employment Share Changes (*22 industries, 1987–2010*)

X variable:	Δ Output		Δ Employment share	
Y variable:	Δ Cap U	Δ TFP	Δ Cap U	Δ TFP
Panel A. OLS, current value of X				
Coefficient (SE)	0.548 (0.028)	0.186 (0.036)	1.70 (0.204)	0.137 (0.215)
Panel B. OLS, adding four lagged values of X				
$\sum$ Coefficient (SE)	0.035 (0.050)	−0.096 (0.084)	−0.364 (0.343)	−1.11 (0.420)
Panel C. 2SLS, current value of X instrumented with Δ defense expenditures				
Coefficient (SE)	0.070 (0.086)	−0.192 (0.097)	−0.030 (0.487)	−1.57 (0.509)
F and χ^2 p-value	0.000, 0.891	0.000, 0.103	0.000, 0.929	0.000, 0.236
N/K/L	506/85/21	506/85/21	506/85/21	506/85/21
Panel D. 2SLS, adding four predetermined lagged values of X				
$\sum$ Coefficient (SE)	−0.041 (0.060)	−0.248 (0.100)	−0.234 (0.399)	−1.78 (0.509)
F and χ^2 p-value	0.000, 0.568	0.000, 0.087	0.005, 0.771	0.005, 0.343
N/K/L	418/85/21	418/85/21	418/85/21	418/85/21

Notes: Δ Cap U = ln change in Federal Reserve Board's measure of industry capacity utilization; Δ TFP = ln change in TFP index, adjusted for labor quality (dependent variable in Table 2); Coefficient (SE) = coefficient (standard error) on the current X variable; $\sum$ Coefficient = sum of the coefficients on current and four lags of the X variable. F and χ^2 p-value and N/K/L as in Table 2.

movements in capacity utilization which obscure the true effect of labor allocations on measured productivity.[31]

The preceding is intended to be heuristic, and should not be taken completely literally. In particular, one cannot interpret the results as necessarily indicating that *all* changes in equilibrium quantity demanded (and labor allocations) are exogenous to productivity. To proceed more carefully, panels C and D of Table 3 instrument each X with defense expenditures, the instrument which I have previously found to be consistently first-stage significant and second-stage exogenous. As before, I enter the instrument separately for each industry, and as before the first- and second-stage test statistics satisfy the requirements of 2SLS in an admirably robust and consistent fashion.

Turning to coefficient estimates, the first notable result is that columns 1 and 3 of panels C and D indicate that defense expenditures, while moving around output and labor allocations, have absolutely no effect on industry-level capacity utilization. This is consistent with Ramey's (2011) argument that defense spending changes are well anticipated by public news announcements. While Ramey's news variable is completely insignificant in the first-stage regressions for this sample, as it was before, this merely confirms that the timing of news is different than the timing of

[31] One can try to use the Fed's industry capacity utilization measures to directly adjust productivity, but this raises additional issues. First, an OLS regression approach is unsuitable, because industry capacity utilization is endogenous to industry productivity, but instruments for industry-level capacity utilization are hard to find, as defense spending is uncorrelated with capacity utilization (see below). Second, one can use the utilization estimates to mechanically adjust productivity, but this requires some assumptions about what is being over- and underutilized (capital; capital and labor; or capital, labor, and some material inputs like energy) and what would have to be changed to reach sustainable output. For my purposes, however, it is sufficient to simply show that as capacity utilization effects disappear in the long run, the OLS relation between employment shares and productivity becomes decidedly negative.

actual expenditures.[32] Nevertheless, actual expenditures, when they arrive, may be well anticipated, so that capacity expands evenly with production needs, resulting in no changes in capacity utilization. Because defense spending has no observable impact on capacity utilization, the long- and short-term coefficients for productivity growth (in the second and fourth columns of panels C and D in Table 3) are virtually identical. The elasticity of observed productivity with respect to output is estimated to be around -0.2 (standard error of about 0.1). The coefficient on labor-share-weighted changes in employment shares, which following the theory above is interpretable as the elasticity of average worker efficacy with respect to the employment share, is found to be about -1.5 in this sample of only 22 industries (standard error of about 0.5). This is greater in absolute magnitude than the maximum of -1 allowable by theory, but not (statistically) significantly so.

To summarize the results for the US KLEMS, out of 18 potential instruments, defense spending is the only one which consistently and strongly satisfies the dual requirements of first-stage significance and second-stage exogeneity. Long- and short-term effects for defense spending are quite similar, as defense spending does not have much of an influence on capacity utilization. The long-term OLS association between changing labor allocations and measured productivity is much more negative than the short-term relation, and this appears to reflect transitory capacity utilization changes consistent with exogenous shifts in demand. The long-term OLS estimate of the elasticity of worker efficacy with respect to employment shares in the total US KLEMS sample (-0.685 in Table 2) is not significantly different from that arrived at using defense expenditures as an instrument (-0.750). Thus, while there may be some endogeneity of labor allocations, it probably accounts for a relatively small share of the total variation (exogenous plus endogenous) in this variable.[33]

Turning now to the EU KLEMS OECD data, as I do not have any instrument which is first-stage significant in the analysis of the entire dataset, I focus on country-specific results. Since defense spending is a robustly significant and exogenous instrument in the US KLEMS data, I begin by running country by country first-stage regressions using defense spending as an instrument. I then proceed to the second-stage analysis for the four non-US countries where I find defense spending to be first-stage significant at the 5 percent level (namely Australia, Finland,

[32] Using Ramey's variable as the instrument in the first-stage regressions for output and employment share changes, I get p-values on the F-tests of 0.161 and 0.483, respectively. Running Ramey's instrument jointly with current expenditures in these regressions, I get p-values of 0.837 and 0.992 on her news variable and 0.000 and 0.000 on actual expenditures. As emphasized earlier above, none of this invalidates Ramey's point that her news variable does a better job of explaining changes in macroeconomic aggregates, which will be influenced by the reaction of private economic actors to the anticipated future consequences of those expenditures. This is distinct, however, from moving actual patterns of production away from the private norm, in which actual expenditures have a more significant effect.

[33] As noted earlier, while the preceding analysis is based upon my labor composition adjustment of BLS TFP growth and my estimates of changing sectoral employment shares by type, results are quite similar if I use the original BLS data on productivity and labor allocations without differentiation by worker type. For example, using defense spending as an instrument, I get the following estimates (standard error) of ξ for the panels in Table 2: panel A: -1.06 (0.275); B: -0.373 (0.218); C: -0.722 (0.218); D: -1.03 (0.275); E: -1.17 (0.269); and G: -0.769 (0.292). These follow the patterns presented in the table. The corresponding short-term and long-term OLS results (panels A and G) are -0.377 (0.122) and -0.809 (0.218). In Table 3, looking at the third and fourth columns of panels C and D, where employment share changes are instrumented with defense expenditures, I get insignificant short- and long-term coefficients for capacity utilization of -0.062 (0.475) and -0.390 (0.376) and short- and long-term coefficients for BLS-measured TFP growth of -1.82 (0.502) and -1.84 (0.514). Again, these results parallel those reported above.

Table 4—Country-Level Analysis Using EU KLEMS Data (*29 sectors, 1970–2005*)

	Australia	Finland	Netherlands	United Kingdom
Panel A. Baseline specification (equation (17)) with Δ defense expenditures/GDP as instrument				
ξ (SE)	−1.09 (0.185)	−0.310 (0.359)	−0.264 (0.335)	−0.886 (0.153)
F and χ^2 p-value	0.004, 0.901	0.000, 0.841	0.005, 0.290	0.000, 0.631
N/K/L	493/100/28	493/100/28	493/100/28	493/100/28
Panel B. Dropping one industry at a time				
Max ξ (SE)	−1.03 (0.176)	−0.079 (0.395)	0.152 (0.406)	−0.727 (0.165)
Min ξ (SE)	−1.18 (0.193)	−0.682 (0.406)	−0.472 (0.366)	−1.10 (0.198)
Max F p-value	0.012	0.004	0.038	0.001
Min F p-value	0.002	0.000	0.002	0.000
Max χ^2 p-value	0.960	0.941	0.691	0.809
Min χ^2 p-value	0.653	0.712	0.026	0.534
Panel C. Adding four lags of employment share changes				
$\sum \xi$ (SE)	−1.15 (0.307)	−0.801 (0.424)	−0.560 (0.407)	−0.985 (0.206)
F and χ^2 p-value	0.014, 0.873	0.000, 0.728	0.001, 0.272	0.000, 0.928

Note: As in Table 2.

the Netherlands, and the United Kingdom). As shown in Table 4, in each of these countries defense spending satisfies the second-stage exclusion requirement and produces negative estimates of ξ, although only the large point estimates of Australia and the United Kingdom are statistically significant. Removing one industry at a time, I find that defense spending robustly satisfies the first- and second-stage significance and exclusion requirements. The point estimates of ξ vary greatly for Finland and the Netherlands and much less so for Australia and the United Kingdom, in keeping with their relative standard errors in the baseline specification. Adding lags of employment share changes to the regression produces a much larger estimate of the cumulative negative effect of reallocation on productivity, particularly for Finland and the Netherlands.

The EU KLEMS database has two sets of estimates for the United States—one covering 1977–2005 based upon the current NAICS (North American Industry Classification System) used in the US KLEMS, and another covering 1970–2005 based upon the historical SIC (standard industrial classification). The industrial sectors in both series share the same nominal titles and have TFP estimates grouped into the same 29 private sector divisions which I use in the general analysis of (SIC-based) EU KLEMS data for other countries.[34] Both of these series provide a longer time series than the BLS' US KLEMS (covering 1987–2010) and appear to be developed independently of that source. As in the case of the US KLEMS, I run first-stage regressions for each of the 18 instruments in Table 1 and then proceed to the second stage with those instruments which are significant at the 5 percent level.

Table 5 reports second-stage results for the six instruments which are first-stage significant at the 5 percent level in the EU KLEMS US NAICS data. Defense expenditures operate much as in the analysis of the US KLEMS, producing an extremely large negative estimate of ξ in the baseline specification, first- and second-stage significance, and exclusion test statistics which are quite robust to the removal of one

[34] Whenever I refer to results using all of the EU KLEMS data, as in Table 1's first-stage regressions, I use the SIC version of the US data, in keeping with the SIC definitions used for other countries.

TABLE 5—US ANALYSIS USING NAICS-BASED US DATA IN EU KLEMS (*By instrument, 29 sectors, 1977–2005*)

	Δ Defense spending	Δ Oil prices	Oil price maximum	Smets/Wouters M shock	Sims/Zha M shock	TED spread innovation
Panel A. Baseline specification (*equation* (*17*))						
ξ (SE)	−1.06 (0.425)	−0.838 (0.424)	−1.13 (0.331)	−0.740 (0.413)	0.932 (0.476)	0.047 (0.328)
F and χ^2 p-value	0.003, 0.524	0.003, 0.000	0.000, 0.000	0.002, 0.307	0.007, 0.869	0.000, 0.001
N/K/L	812/111/28	812/111/28	812/111/28	783/110/28	725/108/28	812/111/28
Panel B. Dropping one industry at a time						
Max ξ (SE)	−0.873 (0.427)	0.667 (0.554)	0.282 (0.404)	−0.520 (0.426)	1.08 (0.510)	0.241 (0.361)
Min ξ (SE)	−1.20 (0.455)	−1.43 (0.435)	−1.72 (0.363)	−1.10 (0.463)	0.763 (0.530)	−0.380 (0.333)
Max F p-value	0.012	0.155	0.000	0.079	0.048	0.000
Min F p-value	0.001	0.002	0.000	0.000	0.003	0.000
Max χ^2 p-value	0.697	0.001	0.011	0.440	0.942	0.009
Min χ^2 p-value	0.362	0.000	0.000	0.205	0.693	0.000
Panel C. Adding four lags of employment share changes						
$\sum \xi$ (SE)	−0.851 (0.394)	−2.16 (0.500)	−1.86 (0.404)	−0.944 (0.456)	−0.626 (0.607)	−0.809 (0.457)
F and χ^2 p-value	0.045, 0.528	0.339, 0.020	0.018, 0.082	0.271, 0.066	0.865, 0.988	0.396, 0.000

Note: As in Table 2.

industry at a time, and (once lags are accounted for) a somewhat smaller estimate of the cumulative effect of employment changes. The oil price maximum, which produced a positive point estimate of ξ earlier in Table 1, generates a ξ of -1.1 in this case. However, notwithstanding its statistical significance in the baseline specification, with the removal of one industry this coefficient is easily made positive. The remaining four instruments produce a cornucopia of insignificant results in the baseline regression, are often quite sensitive to the removal of one industry at a time and, when employment change lags are added, produce big cumulative negative estimates of ξ and are found to be utterly insignificant in the first-stage regression. In sum, as in the analysis of the BLS US data, only defense spending consistently satisfies the first- and second-stage tests, and that instrument produces an estimate of ξ close to -1.

Table 6 reports second-stage results for the six instruments which are first-stage significant at the 5 percent level in the EU KLEMS US SIC data. Three of these instruments (defense spending, oil price maximum, and the TED spread) overlap with the list for the EU KLEMS US NAICS data. While the oil price maximum and TED spread produce results which are similar to those in Table 5, those with defense expenditures are dramatically different. Although defense spending is first-stage significant and second-stage exogenous in the baseline specification, it produces a small and statistically insignificant estimate of ξ. With lags, however, the coefficient becomes considerably more negative, albeit not statistically significant. With regards to the remaining instruments, the point estimates are generally quite sensitive to the removal of one industry or the first-stage regression is rendered insignificant once lags are introduced. With the introduction of lags the cumulative effect of employment changes becomes much more negative, although the TED spread is the only instrument in this specification which is strongly significant and exogenous. Its estimate of ξ is both substantially negative (-0.760) and statistically significant.

As Tables 5 and 6 suggest, there are peculiar differences between the SIC-based and NAICS-based EU KLEMS data for the United States. The correlation

TABLE 6—US ANALYSIS USING SIC-BASED US DATA IN EU KLEMS (*By instrument, 29 sectors, 1970–2005*)

	Δ Defense spending	Oil price maximum	Residual gas prices	Romer/Romer M shock	Fed funds surprises	TED spread innovation
Panel A. Baseline specification (equation (17))						
ξ (SE)	−0.118 (0.251)	−0.821 (0.350)	−0.329 (0.413)	0.162 (0.373)	−0.234 (0.485)	0.086 (0.326)
F and χ^2 p-value	0.000, 0.151	0.000, 0.000	0.031, 0.000	0.036, 0.385	0.001, 0.000	0.000, 0.644
N/K/L	1,015/118/28	1,015/118/28	1,015/118/28	754/109/28	435/98/28	1,015/118/28
Panel B. Dropping one industry at a time						
Max ξ (SE)	−0.048 (0.240)	−0.173 (0.369)	−0.035 (0.400)	0.358 (0.472)	0.412 (0.557)	0.251 (0.320)
Min ξ (SE)	−0.191 (0.468)	−1.14 (0.399)	−2.49 (0.981)	−0.006 (0.384)	−0.599 (0.483)	−0.072 (0.446)
Max F p-value	0.000	0.004	0.926	0.058	0.013	0.000
Min F p-value	0.000	0.000	0.019	0.003	0.000	0.000
Max χ^2 p-value	0.723	0.000	0.008	0.676	0.004	0.876
Min χ^2 p-value	0.009	0.000	0.000	0.191	0.000	0.363
Panel C. Adding four lags of employment share changes						
$\sum \xi$ (SE)	−0.550 (0.449)	−1.55 (0.395)	−1.42 (0.459)	−1.14 (0.532)	−0.351 (0.640)	−0.760 (0.384)
F and χ^2 p-value	0.034, 0.540	0.000, 0.000	0.048, 0.000	0.206, 0.748	0.002, 0.000	0.000, 0.879

Note: As in Table 2.

between the annual industry × year total factor productivity growth in one dataset and the other—for the 29 nominally identical[35] large private sector industry groupings and the 28 years which the two datasets overlap—is only 0.502 (i.e., an R^2 of 0.25), despite the fact that they ostensibly measure exactly the same thing. The labor-income-share-weighted labor reallocation measures, however, are much more similar, with a correlation of 0.860 ($R^2 = 0.74$). Not surprisingly, this produces radically different regression results. There are also some disturbing anomalies in the EU KLEMS SIC-based US data and in the EU KLEMS dataset as a whole.[36] Such concerns are, however, somewhat beside the point, as it cannot be taken as altogether surprising that a single instrument, such as defense spending, will in some specifications or some datasets produce weaker results.

Before concluding, I present the OLS results for the EU KLEMS data. As shown in Table 7, the results here closely parallel those for the United States. Whether in the four European countries examined in the tables above, either of the SIC and

[35] For example, "mining and quarrying," "education," "rubber and plastics," etc.

[36] For example, between 1970 and 1981, according to the EU KLEMS SIC data, the relative value-added price of private sector services to goods in the United States fell 27 percent, while the relative quantity rose by 25 percent, for a −2 percent change in relative nominal value added. According to the current official US National Income and Product Accounts, however, during this same period the relative value-added price of private sector services to goods actually declined only 5 percent (reflecting rising energy prices), while the relative quantity rose by 14 percent, for a +9 percent change in relative nominal value added (Chain47on.xls and VA47on.xls available at www.bea.gov). The historical SIC series on the BEA (Bureau of Economic Analysis) website (GDPbyInd_VA_SIC.xls) does not provide real indices back to 1970, but in the 1977–1981 period it shows an 8 percent increase in relative real service quantity and 5 percent decrease in relative price (similar to the current series 7 percent and 3 percent figures for the same period), while the EU KLEMS SIC data show a 15 percent increase in relative quantity and 13 percent decline in relative price. In making these comparisons, I follow the BEA's definition of goods (agriculture, mining, manufacturing, and construction) and services (all other private sector).

As another example, in the EU KLEMS data one finds that in 3 percent of the observations with total factor productivity estimates capital income is negative, averaging −0.14 of value added and −0.05 of gross output and ranging as far as −5.7 times value added or −0.33 of gross output. In these observations, appearing in 16 countries, one gets very close ($R^2 = 0.985$) to the EU KLEMS estimate of total factor productivity growth by dropping capital growth while using the gross output shares of intermediate inputs and labor (i.e., weights which combined now exceed 1) to calculate the contribution of these inputs to output growth.

TABLE 7—OLS REGRESSIONS USING EU KLEMS (*29 sectors, 1971–2005*)

	Australia	Finland	Netherlands	UK	US NAICS	US SIC	OECD 18
Panel A. Baseline specification with employment share changes (equation (17))							
ξ (SE)	−0.875	−0.266	−0.485	−0.777	−0.518	−0.344	−0.422
	(0.061)	(0.095)	(0.092)	(0.068)	(0.119)	(0.094)	(0.023)
Observations	667	1,015	754	986	812	1,015	12,109
Panel B. Adding four lags of employment share changes							
$\sum \xi$ (SE)	−0.941	−0.726	−0.756	−1.05	−0.946	−0.929	−0.615
	(0.202)	(0.148)	(0.181)	(0.126)	(0.197)	(0.196)	(0.048)
Observations	551	899	638	870	696	899	10,025

Note: As in Table 2.

NAICS versions of EU KLEMS US data, or the EU KLEMS database as a whole, the association between employment share changes and productivity growth is negative, but becomes much more so when past employment share changes are added to the regression. As in the case of the US data, the difference between the current and cumulative coefficients lends itself to the interpretation that exogenous movements in demand produce transitory changes in capacity utilization which obscure the strongly negative long-term association between employment shares and measured productivity. The cumulative OLS coefficients are in most cases quite close to the corresponding cumulative coefficients using 2SLS, suggesting that much of the variation in labor shares is exogenous. I recognize of course that this interpretation—taking employment shares as being exogenous and OLS coefficients as accurate representations of causal relations—is awfully convenient in a paper which struggles to find more than one robust instrument.

The EU KLEMS results, by and large, confirm the analysis using the US KLEMS. Defense spending is the only instrument which is consistently first-stage significant, second-stage exogenous, and robust—both in terms of test statistics and coefficient point estimates—to the selective removal of industries. Long-term OLS elasticities are more negative than short-term relations. The cumulative estimate of ξ, both OLS and 2SLS with defense spending, is always more negative than -0.5 and often much closer to the theoretical limit of -1. Standard errors, however, are very large and coefficient estimates in particular specifications and samples are not significantly different from zero. Thus, while the preponderance of evidence suggests that average worker efficacy does indeed fall with a sector's employment share, there is substantial uncertainty regarding the precise magnitude of the elasticity.

I conclude by simply considering how different values of ξ change our assessment of relative goods and services productivity growth. In Table 8 I combine the 60-sector US KLEMS and 29-sector EU KLEMS sectoral estimates of gross output private sector productivity growth into goods and services value-added aggregates. With a ξ of 0 (i.e., no adjustment for Roy effects) the US and EU KLEMS data indicate that productivity growth is 0.8 percent faster per annum in goods than services in the United States and 1.4 percent faster per annum in the OECD 18 as a whole. Moving down, as ξ becomes more negative the gap between goods and services productivity

Table 8—Average Private Sector Total Factor Productivity Growth with Roy Adjustments
(*Percent per annum*)

	United States (1987–2010, based on US KLEMS)			18 OECD countries (1970–2005, based on EU KLEMS)		
ξ	Goods	Services	Aggregate	Goods	Services	Aggregate
0.00	1.57	0.73	0.97	1.57	0.17	0.70
−0.25	1.34	0.78	0.94	1.31	0.36	0.72
−0.50	1.10	0.84	0.91	1.04	0.55	0.74
−0.75	0.87	0.90	0.88	0.77	0.74	0.76
−1.00	0.64	0.95	0.85	0.51	0.94	0.79

Notes: Goods, services, and aggregate calculated from BLS KLEMS and EU KLEMS 60-sector and 29-sector, respectively; gross output TFP measures using equation (16), with adjustments for bias as indicated by equation (15) earlier above.

growth narrows until, at a value of -0.75, it disappears altogether in both samples.[37] Table 8 also reports aggregate private sector productivity growth, equal to the private sector GDP share weighted sum of sectoral productivity growths, which is quite insensitive to ξ, as increases in one sector are offset by decreases in the other.

In the US National Income and Product Accounts, between 1947 and 2011 the ln relative price of services to goods increases at an average annual rate of 0.83 percent, while the ln relative quantity increases by 0.90 percent. According to the EU KLEMS data, between 1970 and 2005 the ln relative price of services to goods in the OECD 18 increases at an average annual rate of 1.14 percent, while the ln relative quantity rises by 1.06 percent. Thus, the long-run rate of increase of the relative price of services to goods is roughly equal to the long-run rate of increase of their relative quantity. Section I earlier showed that, under the assumptions of equal sectoral factor income shares and proportionate wages—assumptions which are tolerably satisfied in the data[38]—the slope of the Roy supply curve equals $-\Theta_L \xi/(1 + \Theta_L \xi)$. Setting Θ_L equal to $2/3$ and ξ to -0.75, one gets a slope of 1. As the Roy supply curve shows, there are bounds on the explanatory power of the Roy model. If ξ is to lie within its theoretical limit of -1, there must be a sufficient movement of relative quantity and, more precisely, labor allocations relative to the observed sectoral relative price and (measured) productivity movements. Both this simple back-of-the-envelope calculation and the more careful computations of Table 8 show that these movements exist.

The reported difference in goods and services productivity growth in the United States and the OECD is 0.8 and 1.4 percent per annum, respectively. Examining the values in Table 8 for ξ from -0.5 to -1, the range of defense spending-based long-run elasticities found earlier, the adjusted difference ranges from $+0.5$ percent

[37] The Domar-weighted sum of sectoral reallocations is larger in the OECD 18 than in the United States alone, and hence eliminates a larger productivity gap with, interestingly, the same value of ξ.

[38] In the US KLEMS the ln average annual wage per hour is 0.059 higher in goods, with an annual time trend of -0.0014 (0.0005). In the EU KLEMS, across 471 country $\times$ year observations ln relative goods wages are -0.084 lower than in services and, with country dummies, show an annual trend of 0.0052 (0.0004). Regarding factor shares, in the US KLEMS the average annual labor share in goods is 0.65, while in services it is 0.68, and their ln difference has an annual trend of -0.0039 (0.0005). Across the EU KLEMS, the average annual labor share in goods is 0.68 and in services is 0.64 and the ln difference, with country dummies, has an annual trend of 0.0002 (0.0003).

in favor of goods to +0.4 percent in favor of services. Thus, while it provides indications that the productivity growth gap between the two sectors is grossly overstated, this paper does not have a definitive point estimate to deliver to the reader. A value of ξ equal to -0.75, however, lies in the middle of the point estimates, and allows for the reinterpretation of historical productivity, price, and quantity data as representing a world in which true productivity growth in goods and services is roughly equal but Roy worker efficacy effects give rise to relative cost changes and the appearance of productivity growth differences. Thus, the "Roy supply curve" is a plausible, albeit not proven, explanation of the cost disease of services. This is the main point of this paper.

III. Conclusion

William Baumol's cost disease of services has become part of the intellectual landscape of the profession, a truism taught, at least by this author, to generations of students. The profession, however, is also mindful of the fact that total factor productivity growth is a residual—Abramovitz's (1956) famous "measure of our ignorance"—and has constantly sought new ways of explaining it. This paper follows a growing literature showing the role Roy's model of self-selection amongst heterogeneous workers can play in explaining macroeconomic phenomena. It finds evidence in the relation between employment shares and measured productivity that average worker efficacy declines as a sector's employment share increases, systematically biasing standard measures of productivity growth. While there is considerable uncertainty about the precise magnitude of these effects, the depiction of the relative supply of goods and services as being based upon equal goods and services productivity growth—with a rising relative cost brought about by an association between average worker efficacy and sectoral employment shares—is a plausible alternative characterization of developments in the United States and the OECD.

As noted by Jones (2002), barring the Great Depression and World War II, the growth of income per capita in the United States has been a remarkably steady 2 percent per annum for more than 130 years, despite enormous structural changes in the US economy. Theoretically, it is difficult to think about this historical record in a framework in which aggregate economic growth is asymptotically drawn down to that of the slowest, most stagnant, sector. Practically, it is hard to sustain a fear of prospective stagnation in the face of such a lengthy retrospective history of constant growth. The alternative view—that, by and large, a rising tide of technology raises all boats (industries), while changes in relative prices simply reflect movements along a standard classroom concave production possibilities frontier—provides an easier way to think about the past history and future prospects of the US economy.

REFERENCES

Abramovitz, Moses. 1956. "Resource and Output Trends in the United States since 1870." *American Economic Review* 46 (2): 5–23.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2012. "Measuring Economic Policy Uncertainty." Unpublished.

Basmann, R. L. 1960. "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics." *Journal of the American Statistical Association* 55 (292): 650–59.

Bassett, William F., Mary Beth Chosak, John C. Driscoll, and Egon Zakrajšek. 2012. "Changes in Bank Lending Standards and the Macroeconomy." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 2012–24.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review* 57 (3): 415–26.

Baumol, William J. 2012. *The Cost Disease*. New Haven: Yale University Press.

Baumol, William J., Sue Anne Batey Blackman, and Edward N. Wolff. 1985. "Unbalanced Growth Revisited: Asymptotic Stagnancy and New Evidence." *American Economic Review* 75 (4): 806–17.

Baumol, W. J., and W. G. Bowen. 1965. "On the Performing Arts: The Anatomy of Their Economic Problems." *American Economic Review* 55 (1/2): 495–502.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3): 623–85.

Bustos, Paula, Bruno Caprettini, and Jacopo Ponticelli. 2013. "Agricultural Productivity and Structural Transformation: Evidence from Brazil." Unpublished.

Fisher, Jonas D. M., and Ryan Peters. 2010. "Using Stock Returns to Identify Government Spending Shocks." *Economic Journal* 120 (544): 414–36.

Gilbert, Charles, Norman Morin, and Richard Raddock. 2000. "Industrial Production and Capacity Utilization: Recent Developments and the 1999 Revision." *Federal Reserve Bulletin* 86 (3): 188–205.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–1720.

Griliches, Zvi, ed. 1992. *Output Measurement in the Service Sectors*. National Bureau of Economic Research Studies in Income and Wealth. Vol. 56. Chicago: University of Chicago Press.

Griliches, Zvi. 1994. "Productivity, R&D, and the Data Constraint." *American Economic Review* 84 (1): 1–23.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1): 55–93.

Hamilton, James D. 2003. "What Is an Oil Shock?" *Journal of Econometrics* 113 (2): 363–98.

Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market." *Journal of Political Economy* 93 (6): 1077–1125.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2012. "The Allocation of Talent and U.S. Economic Growth." Unpublished.

Jones, Charles I. 2002. "Sources of U.S. Economic Growth in a World of Ideas." *American Economic Review* 92 (1): 220–39.

Kilian, Lutz. 2008. "Exogenous Oil Supply Shocks: How Big Are They and How Much Do They Matter for the U.S. Economy?" *Review of Economics and Statistics* 90 (2): 216–40.

King, Robert G., and Mark W. Watson. 2012. "Inflation and Unit Labor Costs." Unpublished.

Kuralbayeva, Karlygash, and Radoslaw Stefanski. 2013. "Windfalls, Structural Transformation and Specialization." *Journal of International Economics* 90 (2): 273–301.

Lagakos, David, and Michael E. Waugh. 2011. "Selection, Agriculture, and Cross-Country Productivity Differences." Unpublished.

McLaughlin, Kenneth J., and Mark Bils. 2001. "Interindustry Mobility and the Cyclical Upgrading of Labor." *Journal of Labor Economics* 19 (1): 94–135.

Ngai, L. Rachel, and Christopher A. Pissarides. 2007. "Structural Change in a Multisector Model of Growth." *American Economic Review* 97 (1): 429–43.

Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *Quarterly Journal of Economics* 126 (1): 1–50.

Ramey, Valerie A., and Daniel J. Vine. 2011. "Oil, Automobiles, and the U.S. Economy: How Much Have Things Really Changed?" *NBER Macroeconomics Annual 2010*. Vol. 25, edited by Daron Acemoglu and Michael Woodford, 333–67. Chicago: University of Chicago Press.

Romer, Christina D., and David H. Romer. 2004. "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94 (4): 1055–84.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100 (3): 763–801.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–46.

Sargan, J. D. 1958. "The Estimation of Economic Relationships using Instrumental Variables." *Econometrica* 26 (3): 393–415.

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review* 96 (1): 54–81.

Smets, Frank, and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3): 586–606.

Stock, James H., and Mark W. Watson. 2012. "Disentangling the Channels of the 2007–09 Recession." *Brookings Papers on Economic Activity*: 81–135.

Triplett, Jack E., and Barry P. Bosworth. 2004. *Productivity in the U.S. Services Sector: New Sources of Economic Growth*. Washington, DC: Brookings Institution Press.

Young, Alwyn. 2014. "Structural Transformation, the Mismeasurement of Productivity Growth, and the Cost Disease of Services: Dataset." *American Economic Review.* http://dx.doi.org/10.1257/aer.11.3635.