Cointegrated Factor Augmenting Productivity

Alwyn Young*
London School of Economics
February 2025

Abstract

This paper applies tools from macroeconometrics and model selection to uncover the factor augmenting productivity indices pervasive in growth theory, finding that some are rising and some are falling. There is strong statistical evidence of cointegration, with capital and labour productivities trading off against each other, while moving positively with intermediate input productivity. These patterns might perhaps be understood in terms of task based technical change in which the reallocation of tasks generates cost reducing tradeoffs and links between factor productivity growth rates.

*I am grateful to Matthias Doepke, Ethan Ilzetski, Chad Jones, Patrick Kehoe, Luca Neri, Vincenzo Scrutino and seminar participants at Bologna, LSE, and Stanford for helpful comments.

I. Introduction

While technical change might involve any shift in the output associated with given levels of inputs, economic theorists have found it convenient to adopt a narrower factor augmenting conceptualization. Thus, in general form the relation between output, capital, labour and time might be written as $Y(t) = F(K(t), L(t),t)$, where the third "time" argument indicates changes in the mapping between inputs and output. Harrod (1948), however, restricted technical change to be labour augmenting, $Y(t) = F(K(t),A_L(t)L(t))$, while Solow (1962) restricted it to be capital augmenting, $Y(t) = F(A_K(t)K(t),L(t))$. Even earlier, Hicks (1932) described technical change as augmenting the capital-labour aggregator, $Y(t) = A(t)F(K(t),L(t))$ which, given constant returns to scale, can be re-expressed as equal rates of capital and labour augmenting technical change, $Y(t) = F(A(t)K(t),A(t)L(t))$. These three restricted forms of technical change, Harrod-neutral, Solow-neutral and Hicks-neutral, permeate the work of growth theorists. This paper takes the factor augmenting formulation seriously and uses it to guide empirical work to uncover patterns of technical change. It differs slightly in nomenclature, however, substituting the term "productivity growth" for "technical change" so as to emphasize that changes in effective factor productivity might come from sources other than technology, such as the allocation or misallocation (Hsieh & Klenow 2009, Hsieh et al 2019) of factors across tasks.

Simple back of the envelope calculations suggest that if productivity growth is factor augmenting, then in recent decades some factors most likely have experienced substantially negative rates of productivity growth in the US. Consider again the two-factor constant returns to scale capital-labour production function $Y = F(A_K K,A_L L)$, where to reduce clutter we drop the (t) notation indicating the dependence of variables upon time. Price taking firms set the ratio of labour to capital marginal products equal to the wage-rental ratio: $A_L F_2/A_K F_1 = W/R$, where Fj denotes the partial derivative with respect to the j^{th} argument. Differentiating this first order condition with respect to time, we have

$$\text{(I.1)}\quad g\left(\frac{A_K K}{A_L L}\right) = \sigma g\left(\frac{W/A_L}{R/A_K}\right),$$

where σ denotes the elasticity of substitution and we use $g(x)$ here and later to denote the growth rate of x. With perfect competition the elasticity of output with respect to factor j equals its income share θ_j, and so total factor productivity growth, calculated as $g(TFP) = g(Y) - \theta_K g(K) - \theta_L g(L)$, will equal

$$\text{(I.2)}\quad g(TFP) = \theta_K g(A_K) + \theta_L g(A_L).$$

Combining the two equations we get expressions for factor augmenting productivity growth for each factor:

$$\text{(I.3a)}\quad g(A_K) = g(TFP) + \theta_L \frac{g(K/L) - \sigma g(W/R)}{\sigma - 1}$$

$$\text{(I.3b)}\quad g(A_L) = g(TFP) - \theta_K \frac{g(K/L) - \sigma g(W/R)}{\sigma - 1}.$$

Other than σ, all of the objects on the right-hand sides of (I.3) are directly observable in growth accounts data. Thus, one can use (I.3) to calculate the values of σ consistent with positive factor augmenting productivity growth for both factors. Table 1 below presents such calculations for private sector data in the BEA's US KLEMS growth accounts for 1987-2021. For both $g(A_K)$ and $g(A_L)$ to be on average positive over the 34 years an elasticity of substitution greater than 1.38 is needed. Dividing the sample into two, we see that with productivity growth falling and the growth of the capital labour ratio relative to that of the wage rental ratio rising over time, an elasticity of substitution greater than 1.72 is needed in recent years. Unless the elasticity of substitution is quite large, there is simply too much capital-labour substitution relative to the change in the wage-rental ratio to be consistent with both low productivity growth and positive factor augmenting technical change in all factors.

Unfortunately, most estimates of σ lie well below 1. Gechart et al (2022) in a meta study of 3186 estimates of the elasticity of substitution in 121 studies find that ¾ lie below 1. They argue there is substantial evidence of publication bias as smaller point estimates are associated with smaller standard errors. Multiple alternative corrections for publication bias all lead them to a similar point estimate of the elasticity of about .5. In a narrower review of 2419 estimates in 77 studies of the US economy, Knoblach, Roessler & Zwerschke (2020) find a mean estimate of .54 for the aggregate economy and, after correcting for systematic biases arising from the use of particular techniques highlighted in Monte-Carlos by León-Ledesma, McAdam & Willman (2010, 2015), conclude plausible values lie between .45 and .87. Thus, notwithstanding rare estimates greater than or equal to 1.4, and even fewer greater than 1.7 (Figure 3 Gechart et al 2022, Figure 1 Knoblach et al 2020), the evidence strongly suggests the elasticity of capital-labour substitution is below 1, and well below 1 if corrections are made for publication and methodological bias.

One obvious resolution of this puzzle is to conclude that there is systematic under-statement of productivity growth due, say, to mismeasurement of output growth in services, as was emphasized by Griliches (1994). However, to lower the 2004-2021 restriction on σ needed for universally positive factor productivity growth

Table 1: Values of σ Consistent with $g(A_K) \geq 0$ and $g(A_L) \geq 0$
(US KLEMS private sector data)

	$g(K/L)$	$g(W/R)$	θ_L	$g(TFP)$	σ
1987-2021	.017	.009	.59	.0058	> 1.38
1987-2004	.022	.013	.62	.0084	> 1.25
2004-2021	.013	.004	.57	.0032	> 1.72

Note: Calculated using (I.3) and private sector data from BEA US KLEMS. Growth rates of capital and labour are income share weighted averages of sub-categories, reflecting the relative marginal products of those sub-groups. Growth of wage-rental ratio calculated as growth of nominal income less weighted real factors, i.e. payments per factor adjusted for changing composition.

to the 1.25 needed for 1987-2004 would require a quadrupling of productivity growth in the later period from .032 to .0128. Similarly, satisfying the non-negativity constraints with an elasticity of substitution of .5 would require total factor productivity growth of .0191 in 1987-2004 & .0120 in 2004-2021, with these figures rising to .0300 & .0216, respectively, if σ equals .75. Such large adjustments for mismeasurement seem implausible, as persuasively argued for post-millenium data by Byrne, Fernald & Reinsdorf (2016) and Syverson (2017).

Instead of trying to deny the possibility of negative factor augmenting productivity growth, one might instead embrace it as the natural consequence of task based technical change. Following Acemoglu & Restrepo (2018), consider the CES production function across tasks with elasticity of substitution σ, factor augmenting productivity parameters a_K and a_L, and factor task measures α_K and α_L

$$\text{(I.4)}\quad Y = \left[\alpha_K\left(\frac{a_K K}{\alpha_K}\right)^{\frac{\sigma-1}{\sigma}} + \alpha_L\left(\frac{a_L L}{\alpha_L}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} = \left[\left(\alpha_K^{\frac{1}{\sigma-1}} a_K K\right)^{\frac{\sigma-1}{\sigma}} + \left(\alpha_L^{\frac{1}{\sigma-1}} a_L L\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

$$= \left[(A_K K)^{\frac{\sigma-1}{\sigma}} + (A_L L)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \quad \text{where } A_K = \alpha_K^{\frac{1}{\sigma-1}} a_K \text{ and } A_L = \alpha_L^{\frac{1}{\sigma-1}} a_L,$$

and where we assume that the effective supply of each input is efficiently divided equally across the tasks it performs. In this framework, if the increase in the range of tasks performed by one factor is associated with a decline in the tasks performed by the other, measured factor augmenting productivities will be negatively related, with an increase in one necessarily giving rise to a decline in the other. In particular, with an elasticity of substitution less than one technical innovation which increases the range of tasks performed by capital (α_K) at the expense of those performed by labour (α_L) will appear as a measured decline in capital productivity A_K and increase in labour productivity A_L, as seen in (I.4) above and noted in Aghion, Jones & Jones

(2019). Given the low elasticity of substitution across tasks, when a fixed amount of capital is spread across more tasks it becomes less productive per unit, with opposite effects for labour. Firms will implement task reallocations of this sort if the linked changes in factor productivities lower unit costs, i.e. increase total factor productivity.

This paper explores these issues by borrowing tools from macroeconometrics and model selection to test whether measured factor augmenting productivities are cointegrated and examine the degree to which their evolution can be explained by cost reducing movements along a cointegration frontier generated perhaps by inter-factor task tradeoffs, versus shifts in that limiting frontier brought about perhaps by underlying actual factor augmenting technical change. Across a variety of potential production structures, statistical tests, whose Type I error probabilities are verified in Monte Carlos built around the sample sizes and point estimates, consistently reject a vector autoregression (VAR) framework in favour of a vector error correction (VEC) model of cointegration. This includes models which restrict elasticities of substitution to be greater than 1 and, in some cases, find non-negative factor augmenting productivity growth in all factors. Thus, while the "paradox" of negative factor augmenting productivity growth motivates this paper, the reader with strong priors that elasticities of substitution are very large will still conclude that factor productivity is cointegrated. Model selection techniques, however, select in favour models with elasticities of substitution less than 1, whose results consequently receive the most attention.

This paper is not the first to note the difficulty of reconciling observed factor substitution and productivity growth rates with non-negative factor augmenting productivity growth. Bils, Kaymak & Wu (forthcoming) argue that a radical revision of beliefs on the elasticity of substitution between workers of different schooling levels to a number in excess of 4 is needed to reconcile observed changes in relative factor supplies and returns if one disallows a decline in the absolute efficiency of any group. Muck (2017) using recent EU data estimates an elasticity of capital-labour substitution of about .7 and finds opposite positive and negative factor augmenting growth in labour and capital, respectively, of about 3 percent per annum. As already noted above, Aghion, Jones & Jones (2019) point out that task gains by capital would appear as declines in capital augmenting productivity if the elasticity of substitution is less than one. This paper builds on this prior work by establishing statistically the links between negative factor augmenting productivity growth in one factor and positive growth in another brought about by cointegrating relationships and decomposing observed changes into plausible cost reducing movements along cointegration frontiers and outward movements of those frontiers.

While the competition between humans and machines in the performance of tasks has been apparent to the public since at least the Luddites of the early 18th century, the modern formal economic conceptualization can be traced to the seminal paper of Zeira (1998), with further development in Acemoglu & Autor (2011) and foundation in a model of directed technical change in Acemoglu & Restrepo (2018). Microdata evidence given in Autor, Levy & Murnane (2003) and Spitz-Oener (2006) established its empirical importance in the determination of labour demand, wages and inequality, with further support found, e.g., in Autor & Dorn (2013) and Acemoglu & Restrepo (2020). This paper adopts a panel data macro-econometrics perspective and argues that cointegration and large linked positive and negative movements of factor productivities are suggestive of the importance of task based technical change in recent decades. While cointegrated relationships need not only come from the reallocation of tasks, the task based model does provide a natural explanation of observed tradeoffs between labour and capital. At the same time, as shown by economic history and formalized and emphasized by Acemoglu & Restrepo (2018), the allocation of tasks need not be zero sum, and the acquisition of tasks by one factor often creates new tasks for others. I find this most clearly and surprisingly in the form of a positive relationship between the cointegrated productivities (interpretable as transformations of task measures) of intermediate inputs with those of primary factors of production.

Estimating elasticities of substitution and rates of factor augmenting productivity growth is more than challenging. Diamond, McFadden & Rodriguez (1978), as well as Sato (1977), famously showed that if one only assumes the existence of a constant returns to scale production function and observable marginal products, the elasticity of substitution and bias of factor augmenting technical change cannot both be identified from observational data. This is most obvious in (I.1) earlier, where observed growth rates of K/L and W/R can be reconciled with different combinations of σ and the growth of A_K/A_L. Identification requires additional assumptions, and most approaches to estimating the elasticity of substitution explicity or implicitly restrict the nature and bias of factor augmenting technical change, allowing for substantive biases (León-Ledesma, McAdam & Willman 2010, 2015). However, for production functions with trend stationary inputs and factor augmenting parameters, Klump, McAdam & Willman (2007, 2012) , buttressed by Monte Carlos in León-Ledesma, McAdam & Willman (2010), show that the assumptions of a non-linear CES structure and specific trend stationary functional forms for the relationship between factor augmenting productivity and time, as well as systems estimation using both first order optimality conditions and the level of non-linear CES output with normalization at points of interest, are in combination enough to yield credible identification of both elasticities and factor augmenting productivity growth. This levels approach, however, is not useful when total factor productivity, factor ratios and factor augmenting productivity are difference stationary, as suggested by the data, invalidating the levels analysis used in those papers and, for example, Muck's (2017) work noted above.

This paper bases its estimates on differenced data with systems estimation based upon the derivatives of cost minimization conditions and total factor productivity with respect to time. Estimating equations for discrete differences are identical to those for continuous time, with the exception that average factor shares take the place of instantaneous factor shares, a substitution that can be motivated as a second order approximation to arbitrary production functions with constant elasticities within the range of the sample data. The complex non-linear structure of the global CES in differences, however, is not used. Factor augmenting productivity growth is estimated with unrestricted year and industry fixed effects within a VAR or VEC structure. The VAR model is tested against the VEC alternative agnostically assuming a variety of factor nesting and grouping structures and restrictions on elasticities of substitution to be greater or less than 1. Across this universe of potential models, the data always reject the VAR model in favour of the cointegrated VEC structure, while indicating that they are difference stationary.[1]

Absent any of the standard restrictions, this paper "solves" the Diamond et al/Sato identification problem by showing how systems estimation in differences closely resembles a structural VAR, and then adopting the standard structural VAR assumption of assuming the zero mean (factor augmenting) shocks are orthogonal to each other. Identification is achieved by matching, as closely as possible, the unrestricted covariance matrix of relative factor input and total factor productivity growth rates by passing the orthogonal shocks through the structural model. I find that a production structure with a capital and labour value added aggregator separable from intermediates provides the highest likelihoods. Point estimates of the capital-labour elasticity of substitution are around ½, consistent with the values suggested by

[1]Monte Carlos in the on-line appendix show that standard cointegration tests when evaluated using their asymptotic distribution have very large positive size distortions, so wild bootstraps under the null of no or lower order cointegration, which in Monte Carlos show more accurate finite sample rejection rates, are used in all tests.

Gechart et al (2022) and Knoblach, Roessler & Zwerschke (2020) in their reviews.

The heroic assumption of a diagonal covariance matrix is not left unexamined and its impact on results is repeatedly (locally) stress tested. Having identified elasticities assuming a diagonal shock structure, I then take them as known and reestimate all the key parameters of the model, including the cointegration structure, allowing for an unrestricted covariance structure to the shocks. Parameter estimates and conclusions regarding movements along and shifts of cointegration frontiers are very similar. In implementing inference using a wild bootstrap I use two procedures, one which retains the empirical off-diagonal covariance of the estimated shocks and the other which constrains it to be zero. Decisions reached using the two methods are again very similar, with rare highlighted exceptions. The bias of maximum likelihood procedures is evaluated using bootstraps with and without off-diagonal covariance of shocks, finding almost no impact on substantive conclusions. These results should not be taken as affirming the validity of the diagonal covariance assumption, because estimation using an erroneous assumption could bias the covariance structure of the estimated residuals. Rather they should be taken as local to the point estimates, showing that for most factor nesting structures and > or < 1 elasticity restrictions there exist elasticity estimates such that the empirical covariance of the shocks, while not zero, is rarely of any import to the principal conclusions drawn. These elasticity estimates and production function structures, however, run the gamut of priors, and all produce the same conclusion: that factor augmenting productivity growth is cointegrated, producing tradeoffs across factors. As implied by the Diamond et al/Sato result, all estimates of elastiticies of substitution based on observational data make strong assumptions, either explicitly or implicitly. This paper similarly makes one, drawn from the structural VAR literature, using confirmatory tests that, contingent upon given elasticity point estimates, relax the assumption. Combined with the examination of results for a range of production structures and > or < than 1 restrictions on elasticities, this hopefully convinces the reader that while belief in the highest likelihood model and its precise point estimates requires restrictive articles of faith, the overall conclusions are more ecumenical.

The paper proceeds as follows: Section II describes the data, highlighting its heavy tailed and difference stationary distribution, motivating empirical specifications that use the multivariate t-distribution on differenced data, i.e. growth rates. Section III lays out the methodology in what is hopefully a "hands above the table" manner, revealing how, in the light of the Diamond, McFadden & Rodgriquez (1978) and Sato (1977) results, standard structural VAR assumptions regarding diagonal shocks achieve the miracle of identification, and how the wild bootstrap is used to test how binding, practically, these assumptions are in the neighbourhood of point estimates, select among models and provide better finite sample inference. Section IV lays out aforementioned results for a model of capital, labour and intermediates, with the model with the highest likelihood, in particular, finding that capital and labour productivity trade off against each other while varying positively with intermediate input productivity. Section V decomposes productivity growth into movements along the cointegration frontier and shifts of that frontier. While movements along the frontier account from most of the gross changes in factor productivity, in the highest ranked model they account for at most ⅙, or if one insists ⅓, of (net) total factor productivity growth. Shifts of the frontier fuel the remainder, suggesting that while task based technical change is heavily influencing relative factor ratios and returns, more traditional factor augmenting technical change accounts for the lion's share of actual TFP growth. These conclusions consider adjustments for bias as estimated by wild and parameteric bootstraps with and without off-diagonal covariance of shocks. Section VI concludes with observations on how findings of negative productivity growth and cointegration that links difference stationary productivities to moving trend stationary frontiers challenge theoretical models of growth, including those emphasizing task based technical change. An on-line appendix provides Monte Carlos for all empirical techniques using data generating processes based on the practical sample sizes and point estimates found in the paper and theoretical proofs of claims made regarding Taylor approximations and consistency.

II. The Data and their Characteristics

All measures and results reported in this paper pertain to the private sector alone, as I drop data for the government sector. I draw industry x time data on output, inputs and factor incomes from the Bureau of Economic Analysis's KLEMS (capital, labor, energy, materials & services) accounts, which provide a consistent series for 61 private sector industries from 1987 to 2021.[2] Several features of the data related to the methods and results of this paper are worth noting.

---

[2]I use the BEA's KLEMS estimates in preference to those of the BLS, because the latter report instances of negative labour and capital income. The BEA has also produced historical series for 1947-1963 & 1963-2016, but warns these are of doubtful quality. The concern seems valid as, for example, "funds, trusts and other financial vehicles" show college employment increasing 9 fold between 1997 and 1998 only to fall 99 percent between 2006 and 2007. The EU KLEMS have a disturbingly large number of basic inconsistencies. Even after dropping formerly centrally planned economies and less accurate information on detailed sectors, of 4929 11 country (Austria, Belgium, Denmark, Germany, Finland, France, Italy, Netherlands, Spain, Sweden, and the UK) x 18 major industry x year instances in which the growth of tangible capital and its ICT and non-ICT components (as described in Stehrer et al 2019) is provided, the growth of both components is either greater or less than the growth of their

Table 2: "Statistically Significant" Industry Trends in ln Factor Income Shares (1987-2021, for each factor j and industry i, regression: $\ln \theta_{jit} = \alpha_{ji} + \beta_{ji}t$)

	# significant						average $\|\beta\|$ when so					
	capital		labour		inter.		capital		labour		inter.	
level:	.01	.05	.01	.05	.01	.05	.01	.05	.01	.05	.01	.05
	43	51	48	50	40	44	.023	.021	.011	.010	.010	.009

Notes: p-values evaluated using heteroskedasticity robust standard errors. These regressions are designed to show that a Cobb-Douglas specification with constant underlying factor shares and measurement error is not consistent with the data. They do not claim to properly identify trends, as with non-unitary elasticities of substitution difference stationary data produce spuriously significant time trends in regressions of this sort. inter. = intermediates.

Kaldor's (1961) stylized fact on the constancy of national income shares has had an enduring influence on the profession, leading many to conceptualize production as Cobb-Douglas, with observed changes in factor income shares dismissed as measurement error. This view is not consistent with industry level data. If production is Cobb-Douglas, factor shares should be trend stationary with, moreover, no trend whatsoever. Table 2 reports the number of statistically significant time trends found when regressing the ln of industry factor shares on constants and time trends. As shown, the majority of the 61 industries have statistically significant time trends. These trends are quite substantive with the absolute value of the capital, labour and intermediate ln share trends, when statistically significant, averaging 2.3, 1.1 and 1.0 percent per annum for the 34 years of the US KLEMS 1987-2021 series. These point estimates should not be interpreted as identifying true trends in the data since, *provided* elasticities of substitution differ from 1, unit roots in industry level factor inputs could easily produce spuriously "statistically significant" time trends of this sort. The intent is merely to show that the data are totally inconsistent with a Cobb-Douglas view of the world. Karabarbounis & Neiman (2013) demonstrate the existence of substantive changes in factor shares in national data.[3] Table 2 reinforces that message at the US industry level. As seen below, factor augmenting productivity growth can only be identified if elasticities of substitution differ from 1, and Table 2 shows that assumption is not unreasonable.

---

aggregate in 1306 cases. That is, technically speaking, impossible. Turning to the relative incomes of different types of workers, in 15444 11 country x 18 private sector industry x year x gender x 3 age cells, "highly skilled" workers (university graduates) earn 26% less on average than "medium skilled" workers (intermediate qualification) in 2490 instances and 28% less on average than "low skilled" workers (no formal qualification) in 1911 instances, divided roughly equally between genders and individuals aged less than and over 30 (skill definitions as given in Stehrer et al 2019). This seems unlikely.

[3]Insofar as constancy of factor shares was once true for aggregate US data, Koh, Santaeulàlia-Llopis & Zheng (2020) show this can be attributed to the pre-1999 classification of intellectual property products investment as a business expense.

Table 3: Covariance Matrix and Higher Moments of US KLEMS Growth Rates
(residuals from regression on year and industry fixed effects)

	g(*TFP*)	*g*(*K*)	*g*(*L*)	*g*(*I*)	skewness	kurtosis
g(*TFP*)	.0015				-0.04	15.1
g(*K*)	-.0002	.0012			-0.04	58.2
g(*L*)	.0000	.0002	.0021		-0.31	14.0
g(*I*)	-.0013	.0002	.0014	.0136	-0.03	21.2

Note: *TFP*, *K*, *L*, & *I* = total factor productivity, capital, labor, & intermediate input, respectively.

Table 3 reports the covariance and higher moments of the residuals of the regression of the ln growth of total factor productivity and factor inputs on year and industry dummies. Most salient is the fact that the variance of the growth of intermediate input is an order of magnitude greater than that of other factor inputs and TFP growth. It is hard to dismiss this variation as measurement error. As TFP growth is constructed by subtracting the growth of intermediate inputs times their income share (averaging .5) from the growth of gross output, were intermediate input variation pure measurement error the covariance of intermediate input growth with TFP would be -½ times the variance of intermediate input growth. In practice the covariance is negative, as is the covariance of capital with total factor productivity growth,[4] but less than 1⁄10th of the variance of intermediate input. Thus, the reported variation of intermediate input appears for the most part to be real.[5] Explaining this through productivity shocks in a way that does not result in excessive variation of other inputs and TFP growth plays a large role in the way the likelihood selects among models below.

Table 3 also reports the skewness and kurtosis of the growth of TFP & factor inputs. As can be seen, skewness is for the most part minimal, but kurtosis is very much greater than the 3 of the normal distribution. The multivariate t-distribution provides a computationally tractable way of modelling distributions with arbitrarily thick tails parameterized by the degrees of freedom and results using this distribution,

[4]These negative covariances need not represent measurement error, as total factor productivity growth could be contractionary, as found for the aggregate economy by Basu, Fernald & Kimball (2006). At the partial equilibrium industry level, with perfect competition where one percent total factor productivity growth translates into a one percent fall in prices, price inelastic demand would be enough to deliver this result.

[5]Systematic measurement error in the form of changes in intermediate input "use" due to inventory accumulation or decumulation (generating a disconnect between inputs purchased and inputs used in current production), that might conceivably be accompanied by offsetting movements in output and hence not generate a large negative covariance with measured TFP growth, does not seem to be an explanation either. In the KLEMS data the residual (net of year and industry fixed effects) variance of non-storable service and energy inputs (.020 and .046, respectively) is of the same order of magnitude as that of material inputs (.031), while services and energy account for an average of 61% of total input costs.

along with those for the multivariate normal, are shown below. As shown in the on-line appendix, parameter estimates using the multivariate t are equivalent to weighted versions of those for the multivariate normal, with weights inversely related to the degrees of freedom plus the covariance adjusted squared observation residuals.[6] These systematically underweight outlier observations with large residuals, with the underweighting growing as the degrees of freedom falls and the tails of the distribution become thicker. In Monte Carlos in the on-line appendix I find that estimation using the outlier resistant multivariate-t provides more reliably accurate rejection probabilities for the wild bootstrap cointegration tests used below. In the sample, point estimates using the normal and t are very similar, although the multivariate-t provides a much better fit to the data (i.e., higher likelihood). Perhaps for these reasons, statistical tests based upon the multivariate-t provide more consistent and unequivocal results.

The KLEMS TFP, capital/labour (*K*/*L*) and intermediates/labour (*I*/*L*) series feature prominently in the analysis below. Table 4 reports the results of unit root and stationarity tests for their lns and ln differences (growth). Panel unit root tests often have ridiculous size distortions in the sample sizes used in this paper, with null rejection probabilities of 0 or 1.00 at the .05 level for normal or t-distributed error processes, as shown in Monte Carlos in the on-line appendix. Table 4 presents results for tests whose null rejection probabilities I find to be close to the nominal .05 level for both normal and thick tailed t-distributed data generating processes parameterized by estimates from my sample. The Hadri (2000) test in panel (A) considers the regression $y_{it} = r_{it} + controls + error$, where $r_{it} = r_{it-1} + u_{it}$, and uses the Lagrange Multiplier test to test the null that the variance of u_{it} equals 0. This is a test of the null that y_{it} is stationary, with r_{it} equal to an industry *i* fixed effect, against the null that it contains a unit root. The test statistic allows for heteroskedasticity. The Harris & Tsavalis (HT, 1999) test runs the regression $y_{it} = \rho y_{it-1} + controls + error$ and tests the unit root null $\rho = 1$. The Im, Pesaran & Shin (IPS, 2003) test regresses the change in the variable on its lagged value plus controls, $g(y_{it}) = \rho_i y_{it-1} + controls + error$, estimating a separate autoregressive parameter ρ_i for each series in the panel, and uses the mean of their t-statistics to test the null that all $\rho_i = 0$, i.e. all panels contain a unit root. While the other tests evaluate the test statistic using its asymptotic distribution,

[6]With **V** denoting the covariance matrix of the *J*-variate normal which divided by an independent chi-squared variable with τ degrees of freedom produces the *J*-variate-t, and ^ estimated parameters, the weights are:

$$(\hat{\tau} + J)/[\hat{\tau} + (\mathbf{y}'_{it} - \mathbf{x}'_{it}\hat{\boldsymbol{\beta}})\hat{\mathbf{V}}^{-1}(\mathbf{y}_{it} - \hat{\boldsymbol{\beta}}'\mathbf{x}_{it})].$$

.

Table 4: Stationarity and Unit Root Tests by KLEMS Series
(p-values or test statistics relative to critical values)

		ln levels			growth rates		
		ln(*TFP*)	ln(*K*/*L*)	ln(*I*/*L*)	g(*TFP*)	g(*K*/*L*)	g(*I*/*L*)
		(A) tests without correction for autocorrelation:					
Hadri	FE	.000	.000	.000	.000	.000	.263
(H_0: stationary)	trend	.000	.000	.000	.267	.000	.714
Harris-Tsavalis	FE	.981	1.00	.023	.000	.000	.000
(H_0: has unit root)	trend	.830	1.00	.016	.000	.000	.000
Im-Pesaran-Shin*	FE	1.0	0.81	.96	3.6	3.0	3.7
(H_0: has unit root)	trend	.97	0.74	.98	2.6	2.3	2.7
		(B) tests with correction for first order autocorrelation					
Im-Pesaran-Shin	FE	.280	.877	.920	.000	.000	.000
(H_0: has unit root)	trend	.773	.998	.912	.000	.000	.000
Fisher-Z	FE	.216	.812	.929	.000	.000	.000
(H_0: has unit root)	trend	.778	.999	.826	.000	.000	.000
Fisher-L	FE	.239	.865	.932	.000	.000	.000
(H_0: has unit root)	trend	.771	1.00	.810	.000	.000	.000

Notes: (*) Reports test statistic relative to finite sample .05 critical value calculated by IPS; all others report p-value based upon asymptotic distribution. Fisher-Z and -L: using the Z and L summary statistics given in the text. FE: industry and year fixed effects as controls; trend adds industry specific time trends as controls.

IPS used simulation to calculate critical values for a finite sample .05 null rejection rate in fixed sample sizes and the table reports the value of the test statistic relative to that critical value. Reported results include industry and year fixed effects or industry & year fixed effects plus series specific time trends as controls.

As seen in Table 4, panel (A), the HT & IPS tests consistently and emphatically reject the null that the growth rates of the series contain a unit root, while the Hadri test consistently and emphatically rejects the null that their levels are stationary. In most cases the same tests do not reject the null that the level has a unit root or the difference is stationary. However, the available distributions for these tests do not allow for the possibility that the error process is serially correlated and I find, in Monte Carlos in the on-line appendix, that when this is the case they frequently have have either 0 or 1.00 rejection rates at the .05 level, making both Type I and II errors highly likely.

To address the issue of autocorrelation, panel (B) of Table 4 uses tests that control for serial correlation and in Monte Carlos in the on-line appendix produce null rejection probabilities reasonably close to the nominal .05 level. These are the IPS test with the lagged value of $g(y_{it})$ as a right-hand side control, evaluating the test statistic using its asymptotic distribution, and Fisher-type tests that combine the p-values (p_i) of Dickey & Fuller (1979) unit root tests with controls for serial correlation for each of the N industries. The Fisher summary statistics are $Z = N^{-½}\sum_{i=1}^{N}\Phi^{-1}(p_i)$ & $L = \sum_{i=1}^{N}\ln(p_i/1-p_i)$, where Φ^{-1} is the inverse cumulative standard normal. As can be seen in panel (B), at all reasonable statistical levels all three tests unambiguously reject the null of a unit root for the growth rates of the three series and equally unambiguously do not reject the unit root null for the ln series themselves. Further below I also use Johansen's (1995) cointegration sequential testing procedure that allows rejection of the null of unit roots in favour of the conclusion that the ln levels of the series are stationary. Across a dozen models, estimated with normal or t-distributions, this never occurs. In their totality, these results support the notion that the ln series contain a unit root process and are difference (growth) stationary.[7] I treat them as such in the remainder of the paper.

III. Methods

a. Framework

Our starting point is a homogenous of degree one (constant returns to scale) production function for gross output Q_{it} in industry i at time t as a function of J inputs X_{1it} ... X_{Jit} and their corresponding factor augmenting productivity parameters A_{1it} ... A_{Jit}:

$$\text{(III.1)}\quad Q_{it} = F^i(A_{1it}X_{1it}, A_{2it}X_{2it}, \ldots, A_{Jit}X_{Jit}).$$

While the production function F^i may vary by industry, all changes through time are restricted to factor inputs and productivity parameters. With perfect competition, where factor income shares equal the elasticity of output with respect to each input, total factor productivity growth for each industry, calculated as the growth of real output minus the factor income share (θ_{jit}) weighted growth of inputs X_{jit}, j =1 ... J, will equal:

$$\text{(III.2)}\quad g(TFP_{it}) = \sum_{j=1}^{J}\theta_{jit}g(A_{jit}).$$

Cost-minimizing price-taking firms set ratios of marginal products equal to ratios of economy-wide factor prices p_{1t} ... p_{Jt}:

---

[7] The on-line appendix provides results for Table 4 using tests not reported here whose null rejection probabilities in Monte Carlos differ markedly from nominal value. These universally reject the null that the ln series are stationary and the differenced (growth) series have a unit root. Unit root tests that have empirical null rejection probabilities well above nominal value reject the null of a unit root in the ln series in a few instances, while the Hadri test with a homoskedastic covariance estimate rejects the stationary null for the differenced series less frequently than is found using the heteroskedasticity consistent covariance estimate in Table 4, more strongly suggesting that the data are difference stationary.

$$\text{(III.3)}\quad \frac{A_{jit}F^i_j\left(\frac{A_{1it}X_{1it}}{A_{Jit}X_{Jit}},\ldots,\frac{A_{J-1it}X_{J-1it}}{A_{Jit}X_{Jit}},1\right)}{A_{Jit}F^i_J\left(\frac{A_{1it}X_{1it}}{A_{Jit}X_{Jit}},\ldots,\frac{A_{J-1it}X_{J-1it}}{A_{Jit}X_{Jit}},1\right)} = \frac{p_{jt}}{p_{Jt}} \quad \forall\ j,$$

where F^i_j denotes the partial derivative of F^i with respect to its j^{th} argument and we use the homogeneity of degree zero of marginal products to re-express these in terms of effective factor use (A_jX_j) relative to that of the "numeraire" factor J.

Differentiating (III.3 with respect to time gives

$$\text{(III.4)}\quad \sum_{k=1}^{J-1}\left(\xi_{jk}-\xi_{Jk}\right)\left[g\left(\frac{X_{kit}}{X_{Jit}}\right)+g\left(\frac{A_{kit}}{A_{Jit}}\right)\right] = g\left(\frac{p_{jt}}{p_{Jt}}\right)-g\left(\frac{A_{jit}}{A_{Jit}}\right) \quad \forall\ j,\ \text{with}\ \xi_{jk}=\frac{F^i_{jk}}{F^i_j}\frac{A_{kit}X_{kit}}{A_{Jit}X_{Jit}}.$$

Rearranging (III.4) and stacking it on top of (III.2), in matrix and vector form we have:

$$\text{(III.5)}\quad \underbrace{\begin{bmatrix}\mathbf{g}(X_{jit}/X_{Jit})\\ g(TFP_{it})\end{bmatrix}}_{\mathbf{g}(y_{jit})} = \underbrace{\begin{bmatrix}\mathbf{E}^{-1}_{it}\\ \mathbf{0}'_{J-1}\end{bmatrix}}_{\mathbf{B}_{it}}\mathbf{g}(p_{jt}/p_{Jt}) + \underbrace{\begin{bmatrix}-\mathbf{E}^{-1}_{it}-\mathbf{I}_{J-1},\mathbf{E}^{-1}_{it}\mathbf{i}_{J-1}+\mathbf{i}_{J-1}\\ \boldsymbol{\theta}'_{it}\end{bmatrix}}_{\mathbf{C}_{it}}\mathbf{g}(A_{jit}),$$

where $\mathbf{g}(X_{jit}/X_{Jit})$ and $\mathbf{g}(p_{jt}/p_{Jt})$ denote the J-1 x 1 vectors of growth rates of relative inputs and factor prices, $\mathbf{g}(A_{jit})$ the J x 1 vector of growth rates of factor augmenting technical change, $\mathbf{E}_{it}$ the J-1 x J-1 matrix with jk^{th} element ξ_{jk} - ξ_{Jk}, $\boldsymbol{\theta}_{it}$ the J x 1 vector of factor income shares, $\mathbf{I}_{J-1}$ the J-1 identity matrix, $\mathbf{i}_{J-1}$ & $\mathbf{0}_{J-1}$ J-1 x 1 vectors of ones and zeros, respectively, and we define the compact notation $\mathbf{g}(y_{jit})$, $\mathbf{B}_{it}$, and $\mathbf{C}_{it}$ for use later.

The elements of $\mathbf{E}^{-1}_{it}$ are interrelated, as $\xi_{jk}=\xi_{kj}\theta_{kit}/\theta_{jit}$, so the matrix should not be estimated as $(J-1)^2$ independent parameters. I simplify the analysis by assuming that the production function has a nested structure, allowing the expression of $\mathbf{E}^{-1}_{it}$ as a function of a limited number of elasticity parameters and factor shares. Thus, in the three factor model estimated in this paper one might have

$$\text{(III.6)}\quad Q_{it}=F^i(G^i(A_{1it}X_{1it},A_{2it}X_{2it}),A_{3it}X_{3it})),$$

where G^i is a constant returns to scale (X_1,X_2) aggregator, and $\mathbf{E}^{-1}_{it}$ is given by

$$\text{(III.7)}\quad \hat{\mathbf{E}}^{-1}_{it}=\frac{1}{\theta_{1it}+\theta_{2it}}\begin{bmatrix}-\sigma\theta_{2it}-\eta\theta_{1it} & (\sigma-\eta)\theta_{2it}\\ (\sigma-\eta)\theta_{1it} & -\sigma\theta_{1it}-\eta\theta_{2it}\end{bmatrix},$$

where σ is the elasticity of substitution between $A_{1it}X_{1it}$ and $A_{2it}X_{2it}$ and η the elasticity of substitution between $A_{3it}X_{3it}$ and the aggregate G^i.[8] One such production function is

[8]Once the number of factors is greater than 2, there are many definitions of the term elasticity (see Duffy, Papageorgiou & Perez-Sebastian 2004 for a summary). Throughout I use the term

the nested CES:[9]

$$\text{(III.8)}\quad Q_{it} = \{[(A_{1it}X_{1it})^{\frac{\sigma-1}{\sigma}} + (A_{2it}X_{2it})^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma}{\sigma-1}\frac{\eta-1}{\eta}} + (A_{3it}X_{3it})^{\frac{\eta-1}{\eta}}\}^{\frac{\eta}{\eta-1}}.$$

In the empirical implementation I assume that elasticities are constant across the range of the data, which could be motivated with a CES structure such as (III.8), but I do not exploit the non-linear structure of the CES in estimating differences between discrete time periods, as explained further below.

I model factor augmenting technical change using vector autoregression (VAR) and vector error correction (VEC) specifications:

$$\text{(III.9a)}\quad \mathbf{g}(A_{jit}) = \mathbf{\Gamma g}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it}$$

$$\text{(III.9b)}\quad \mathbf{g}(A_{jit}) = \mathbf{\Gamma g}(A_{jit-1}) + \boldsymbol{\alpha\beta}'\mathbf{ln}(A_{jit-1}) + \boldsymbol{\eta}_t + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_{it},$$

were $\mathbf{g}(A_{jit})$ and $\mathbf{ln}(A_{jit})$ denote the J x 1 vectors of $j = 1 \ldots J$ growth rates and levels of factor augmenting productivity, $\mathbf{\Gamma}$ the J x J matrix of coefficients on lagged values of $\mathbf{g}(A_{jit})$, $\boldsymbol{\eta}_t$ and $\boldsymbol{\eta}_i$ J x 1 vectors of fixed effects (dummies) for each industry i and time period t, $\boldsymbol{\varepsilon}_{it}$ the J x 1 vector of iid shocks drawn from the multivariate normal or multivariate-t distribution with diagonal covariance matrix $\mathbf{V}(\boldsymbol{\varepsilon}_{it})$,[10] and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ J x M matrices of coefficients, where M < J. The assumption of a diagonal covariance matrix forms the basis of identification and is stress tested using ancilliary estimates and wild bootstraps that allow correlations between shocks, as explained further below. The $\boldsymbol{\beta}$ define cointegrating frontiers $\boldsymbol{\beta}'\mathbf{ln}(A_{jit-1}) = \mathbf{c}_{it}$ that tie the Jx1 non-stationary vector $\mathbf{ln}(A_{jit-1})$ to the trend stationary Mx1 vector $\mathbf{c}_{it}$, while the $\boldsymbol{\alpha}$ indicate the adjustment process back to the frontiers following a temporary deviation where $\boldsymbol{\beta}'\mathbf{ln}(A_{jit-1}) - \mathbf{c}_{it} \neq \mathbf{0}_M$. The industry dummies for each factor make the estimates invariant with respect to the starting values chosen for the indices of factor productivity A_{ji0} & the associated implicit values of $\mathbf{c}_{i0}$ and the time dummies account for economy-wide movements of $\mathbf{c}_{it}$, while also serving as control functions that ensure consistency despite the endogeneity of factor prices, as discussed further below. Of course, the industry and time dummies for each factor also account for simple mean industry and year differences in productivity growth. Section V later presents interpretations and decompositions of cointegrated factor augmenting

elasticity in the sense of the percentage change in the equilibrium use of two factors or factor aggregates for a percentage change in their relative price.

[9]We may think of the usual constants as being absorbed into the industry factor augmenting parameters.

[10]For the multivariate-t, equal to the distribution of a multivariate normal divided by an independent chi-squared variable, second moments may not exist. By its "covariance matrix", in the above I mean that of the multivariate normal on which it is based.

productivity growth, but for now we concentrate on foundational details of identification and testing and (in Section IV) the presentation of results.

b. The Miracle of Identification

This paper accepts the assumptions of constant returns to scale and perfect competition which underlie KLEMS total factor productivity databases.[11] This sidesteps many of the issues in production function estimation, as the elasticity of output with respect to each factor is given by the known values of $\boldsymbol{\theta}_{it}$. There is still, however, the obvious endogeneity of the growth rates of relative factor prices, $\mathbf{g}(p_{jt}/p_{Jt})$, on the right-hand side in (III.5), as well as the question of how, in the light of the Diamond el al/Sato results, identification of both elasticities of substitution and rates of factor augmenting technical change is achieved. In the interest of transparency, this section explores these issues.

I begin by clarifying why I do not exploit restrictions for discrete changes implied by particular production functional forms. The formulae in (III.5) and (III.7), derived through differentiation, are for instantaneous rates of change for arbitrary constant returns to scale production functions. For changes across discrete time periods, where the growth rates in the formulae represent discrete ln changes, as in $g(Q_{it}) = \ln(Q_{it}/Q_{it-1})$, they are generally incorrect. Thus, for example, if the production function is literally CES then with total factor productivity in the US KLEMS calculated as the growth of output minus the average factor income share weighted growth of factor inputs, for discrete time data we have:

$$\text{(III.10)} \quad g(TFP_{it}) = g(Q_{it}) - \overline{\boldsymbol{\theta}}'_{it}\mathbf{g}(X_{jit}) \neq \overline{\boldsymbol{\theta}}'_{it}\mathbf{g}(A_{jit}), \text{ where } \overline{\boldsymbol{\theta}}_{it} = (\boldsymbol{\theta}_{it} + \boldsymbol{\theta}_{it-1})/2.$$

However, using a second order Taylor series approximation in lns (i.e. translog) of the CES, or for that matter any twice continuously differentiable production function, and a first order Taylor series approximation of factor income shares, it is shown in the on-line appendix that for discrete time:

$$\text{(III.11)} \quad g(TFP_{it}) = g(Q_{it}^{TS2}) - \overline{\boldsymbol{\theta}}_{it}'^{TS1}\mathbf{g}(X_{jit}) = \overline{\boldsymbol{\theta}}_{it}'^{TS1}\mathbf{g}(A_{jit}), \text{ where } \overline{\boldsymbol{\theta}}_{it}^{TS1} = (\boldsymbol{\theta}_{it}^{TS1} + \boldsymbol{\theta}_{it-1}^{TS1})/2,$$

and the superscripts TS1 and TS2 indicate first and second order Taylor series approximations as functions of the lns of effective factor inputs. Similarly, the on-line appendix shows that for a second order approximation in lns of the first order optimization conditions (III.3) for any nested production function of the form (III.6) with constant elasticities of substitution σ and η through the range of the data, the line $\mathbf{g}(X_{jit}/X_{Jit}) = \mathbf{E}_{it}^{-1}\mathbf{g}(p_{jt}/p_{Jt}) + [-\mathbf{E}_{it}^{-1} - \mathbf{I}_{J-1}, \mathbf{E}_{it}^{-1}\mathbf{i}_{J-1} + \mathbf{i}_{J-1}]\mathbf{g}(A_{jit})$ in (III.5) holds for

[11] As the methods used here are in themselves already unfamiliar and complex, the additional data and methods needed to relax these assumptions are best left to further work.

discrete changes, provided we calculate the elements of $\mathbf{E}_{it}^{-1}$ in (III.7) using the average of first order approximations of relative factor income shares. In sum, with the use of observed average factor income shares in place of instantaneous income shares, for discrete time (III.5) and (III.7) are correct up to second order approximations of the production function and first order optimization conditions and first order approximations of factor income shares. Further adjustments for specific functional forms rely upon inaccuracies due to higher order terms whose empirical relevance is hard to motivate and whose role in identification is not easily understood. Assuming that the covariance matrix of factor augmenting shocks is diagonal, as done in this paper, is a transparent assumption whose role in identification is easily understood, while the local impact of apparent departures from this assumption in the data on key results can be evaluated, as shown below.

Next, it is important to recognize that estimation of the model described by (III.5) and (III.7) intrinsicially involves model selection. If $\mathbf{V}$ is the covariance matrix of the factor augmenting productivity shocks, then from (III.5) we see that the covariance matrix of the vector of dependent variables in observation *it* is given by $\mathbf{C}_{it}\mathbf{V}\mathbf{C}_{it}'$. When either of the elasticities σ or η in (III.7) equals 1, $\mathbf{C}_{it}$ is singular and the determinant of the covariance matrix 0. Thus, the ln likelihood of the model approaches - ∞ as the point estimates of elasticities approach 1, for the simple reason that at such values there is no way to explain variation in factor ratios based upon variation in factor augmenting parameters.[12] As noted above, observed changes in factor input shares rule out elasticities of substitution equal to 1, so in practice this is not an unreasonable a priori restriction. It does, however, preclude the possibility of consistency of maximum likelihood estimation, as the parameter space is effectively a non-convex set. The solution is to restrict the space for each parameter to lie above or below 1, treating each such restriction as a separate model. For (III.5) and (III.7), for example, we would estimate four models, based upon whether each of σ and η lies in $[0,1)$ or $(1,\infty)$. Different ways of nesting the factors of production, selecting which two factors appear in the G^i aggregator in (III.6), also constitute alternative models. Non-nested model selection techniques, reviewed further below, then need to be used to select the model, as determined by both the nesting structure and parameter restrictions, most supported by the data. This is actually a positive (if computationally costly) feature of the methodology, as the estimation of a variety of nesting structures and elasticity restrictions (> or < 1) allows readers to examine results for their

[12] Consider for example the Cobb-Douglas production function, $Q_{it} = (A_{1it}X_{1it})^{\alpha}(A_{2it}X_{2it})^{1-\alpha}$, where $X_{1it}/X_{2it} = (p_{2t}/p_{1t})\alpha/(1-\alpha)$, i.e. does not depend on A_{1it} or A_{2it}.

preferred model while also seeing the degree to which the data distinguish between modelling structures.

Turning to the details of identification, it is useful to reexpress the estimating equations in a manner that replaces the latent and unobserved growth rates and levels of factor productivity, $\mathbf{g}(A_{jit})$ and $\mathbf{ln}(A_{jit})$, with parameters & observables. Using the compact notation given in (III.5), we have

$$\text{(III.12)}\quad \mathbf{g}(A_{jit}) = \mathbf{ln}(A_{jit} / A_{jit-1}) = \mathbf{C}_{it}^{-1}\mathbf{g}(y_{jit}) - \mathbf{C}_{it}^{-1}\mathbf{B}_{it}\mathbf{g}(p_{jt} / p_{Jt})$$

$$\mathbf{ln}(A_{jit}) = \mathbf{ln}(A_{ji0}) + \sum_{s=1}^{t}[\mathbf{C}_{is}^{-1}\mathbf{g}(y_{jis}) - \mathbf{C}_{is}^{-1}\mathbf{B}_{is}\mathbf{g}(p_{js} / p_{Js})].$$

For the VAR or VEC models of (III.9a) and (III.9b), (III.5) can then be operationalized as one of two estimating equations:

$$\text{(III.13a)}\quad \text{VAR}: \mathbf{g}(y_{jit}) = \mathbf{B}_{it}\mathbf{g}(p_{jt} / p_{Jt}) + \mathbf{C}_{it}\mathbf{\Gamma}\mathbf{C}_{it-1}^{-1}[\mathbf{g}(y_{jit-1}) - \mathbf{B}_{it-1}\mathbf{g}(p_{jt-1} / p_{Jt-1})]$$
$$+ \mathbf{C}_{it}\boldsymbol{\eta}_t + \mathbf{C}_{it}\boldsymbol{\eta}_i + \mathbf{C}_{it}\boldsymbol{\varepsilon}_{it}$$

$$\text{(III.13b)}\quad \text{VEC}: \mathbf{g}(y_{jit}) = \mathbf{B}_{it}\mathbf{g}(p_{jt} / p_{Jt}) + \mathbf{C}_{it}\mathbf{\Gamma}\mathbf{C}_{it-1}^{-1}[\mathbf{g}(y_{jit-1}) - \mathbf{B}_{it-1}\mathbf{g}(p_{jt-1} / p_{Jt-1})] + \mathbf{C}_{it}\boldsymbol{\eta}_t$$
$$+ \mathbf{C}_{it}\boldsymbol{\alpha}\boldsymbol{\beta}'\sum_{s=1}^{t}[\mathbf{C}_{is}^{-1}\mathbf{g}(y_{jis}) - \mathbf{C}_{is}^{-1}\mathbf{B}_{is}\mathbf{g}(p_{js} / p_{Js})] + \mathbf{C}_{it}[\boldsymbol{\eta}_i + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{ln}(A_{ji0})] + \mathbf{C}_{it}\boldsymbol{\varepsilon}_{it}.$$

The term $[\boldsymbol{\eta}_i + \boldsymbol{\alpha}\boldsymbol{\beta}'\,\mathbf{ln}(A_{ji0})]$ shows that the initial levels of the productivity indices A_{ji0} don't matter, as their estimated effects are simply offset in the estimated J x 1 vectors of individual industry dummies $\boldsymbol{\eta}_i$, without changing the predicted values of $\mathbf{g}(y_{jit})$ in (III.13b) or $\mathbf{g}(A_{jit})$ in (III.9b). To simplify matters, we set these initial indices equal to 1, so that $\mathbf{ln}(A_{ji0}) = \mathbf{0}_J$.

Consider now, for the sake of exposition, the case where all factor income shares on the right hand side of the estimating equations are constant, $\theta_{jit} = \theta_j$ for all factors j, so that in (III.5) $\mathbf{E}_{it} = \mathbf{E}$, $\mathbf{C}_{it} = \mathbf{C}$ and $\mathbf{B}_{it} = \mathbf{B}$ for all it. The VAR model in (III.13a) is then given by:

$$\text{(III.14)}\quad \mathbf{g}(y_{jit}) = \mathbf{C\Gamma C}^{-1}\mathbf{g}(y_{jit-1}) + \mathbf{Bg}(p_{jt} / p_{Jt}) - \mathbf{C\Gamma C}^{-1}\mathbf{Bg}(p_{jt-1} / p_{Jt-1})$$
$$+ \mathbf{C}\boldsymbol{\eta}_t + \mathbf{C}\boldsymbol{\eta}_i + \mathbf{C}\boldsymbol{\varepsilon}_t \ = \ \mathbf{g}(y_{jit}) = \widetilde{\mathbf{\Gamma}}\mathbf{g}(y_{jit-1}) + \widetilde{\boldsymbol{\eta}}_t + \widetilde{\boldsymbol{\eta}}_i + \mathbf{C}\boldsymbol{\varepsilon}_{it},$$

where $\widetilde{\mathbf{\Gamma}} = \mathbf{C\Gamma C}^{-1}$, $\widetilde{\boldsymbol{\eta}}_t = \mathbf{Bg}(p_{jt} / p_{Jt}) - \mathbf{C\Gamma C}^{-1}\mathbf{Bg}(p_{jt-1} / p_{Jt-1}) + \mathbf{C}\boldsymbol{\eta}_t$, & $\widetilde{\boldsymbol{\eta}}_i = \mathbf{C}\boldsymbol{\eta}_i$. In (III.14) we see, first, that as the growth of factor prices on the right hand side is constant across i for each t, any such effects are captured in the factor x year fixed effects and hence it plays no role whatsoever in identification (i.e the residuals would be the same were it removed from the equation) and its endogeneity is irrelevant. Its sole role is, conditional on the estimates of the elasticities in **B** & **C** produced by other variation, to adjust for the impact of changing factor prices on factor allocations, thereby allowing a correct computation of underlying rates of factor augmenting technical change $\mathbf{g}(A_{jit})$. Second, we see that first order auto-correlation and industry and year means play no role in identifying the elasticities in **E** (which determines **B** and **C**) because for any non-singular value of **E** there are alternate values of $\mathbf{\Gamma}, \mathbf{\eta}_t$ & $\mathbf{\eta}_i$ such that $\widetilde{\mathbf{\Gamma}}, \widetilde{\mathbf{\eta}}_t$ & $\widetilde{\mathbf{\eta}}_i$ are unchanged.

Third, barring the error term, (III.14) is a nonstructural VAR or OLS system of seemingly unrelated equations. With normal errors the ln likelihood of such a system is maximized by setting coefficient estimates equal to their OLS values and setting the covariance estimate equal to the empirical covariance of the OLS residuals. Not surprisingly, given the points already made, with normal errors, the maximum likelihood structural model produces *exactly the same* predicted values, and then tries to match the covariance matrix of the resulting residuals as closely as possible using $\mathbf{CVC}'$. Fourth, if the covariance matrix **V** of the factor augmenting productivity shocks **ε** is unrestricted, the model is unidentified as different combinations of the elements of **V** and σ and η (in **C**) produce the same matrix $\mathbf{CVC}'$. If, however, **V** is made to be diagonal, then the values of σ and η, along with the nesting structure, will determine how well the matrix $\mathbf{CVC}'$ matches the covariance structure of the dependent variables. This is the standard assumption made in structural vector auto-regression models to achieve identification of structural parameters.

To summarize, with constant income shares as regressors and normal shocks the VAR model of this paper is basically a structural VAR, whose predicted values are exactly the same as those of the non-structural VAR and whose structural parameters (the elasticities) are identified by the assumption of orthogonal shocks, with the added twist that goodness of fit to the unrestricted non-structural covariance matrix is not only used to estimate structural parameters within models, but also to select amongst different models. The only fly in the ointment, not yet mentioned, is that with constant income shares for each nesting structure there are actually pairs of values of σ, on opposite sides of 1, matched with identical η that generate the same $\mathbf{CVC}'$ and hence match the data equally well!

Factor income shares, however, are not constant, and hence $\mathbf{C}_{it}$ and $\mathbf{E}_{it}$ are not constant. Consequently, it is possible to distinguish between equivalent model pairs by the way in which both the covariance of the dependent variables ($\mathbf{C}_{it}\mathbf{VC}'_{it}$) and the predicted values (based on $\mathbf{C}_{it}\mathbf{\Gamma C}^{-1}_{it-1}\mathbf{y}_{it-1}$, $\mathbf{B}_{it}\mathbf{g}(p_{jt}/p_{Jt})$, $\mathbf{C}_{it}\mathbf{\eta}_t$ & $\mathbf{C}_{it}\mathbf{\eta}_i$) vary with income shares. As shown later, mean income shares do most of the heavy lifting, finding that the nesting structure with a capital-labour value added aggregator (and paired values of σ on opposite sides of 1) matches the unrestricted covariance of the dependent variables quite closely despite the diagonal restriction on **V**. The response to variation in θ_{jit} then selects among the possible > or < 1 elasticity restrictions. With t-distributed errors or VEC cointegration, point estimates of the structural model do depend upon the covariance structure of errors, even with constant factor shares, and hence do not exactly match familiar non-structural OLS, VAR & VEC counterparts, so it is no longer the case that all of the identification of models & elasticities comes from matching the non-structural covariance structure with the structural $\mathbf{CVC}'$. But the intuition given above carries through, as evidenced by the fact that the point estimates of elasticities and rankings of models in those frameworks are very similar to those found using the normal VAR structural model with constant factor shares, as shown below.

With variation in factor shares and, hence, in $\mathbf{B}_{it}$, it is no longer the case that point estimates & residuals are exactly the same whether or not the growth of factor prices is included in the regression. However, with year fixed effects it is still the case that $\mathbf{g}(p_{jt}/p_{Jt})$, which varies by year but not by industry, does not impede consistency, as I prove theoretically using a simple example in the on-line appendix,[13] and plays no role in identifying elasticities. As further evidence of this, I find, as reported below, that elasticity estimates are virtually identical when the growth of factor prices is removed from the estimating equations.

Endogeneity bias may also arise from the use of factor shares as (non-linear) regressors. As changes in factor shares are influenced by the shocks $\boldsymbol{\varepsilon}_{it}$, current levels of factor shares used as regressors are ever so slightly correlated with these shocks. Monte Carlos in the on-line appendix based upon the parameter estimates below show these miniscule correlations do not inhibit standard root-N convergence (& N super convergence for cointegration parameters) to the point where mean squared error is trivially small. Moreover, as has already been noted, in practice below identification of elasticity values and hence average rates of factor augmenting productivity growth mostly comes from the way in which the modelling nesting structure interacts with mean factor shares. Endogeneity bias aside, there is also the fact that fixed effects in autoregressive models bias the estimates of autoregression and other parameters (Nickell 1981), while maximum likelihood estimation is, in any case, not unbiased. To address such concerns, I use wild and parametric bootstraps to estimate potential

[13]To be clear, I show that asymptotically the derivatives of the likelihood equal zero at the true parameter values, i.e. the estimates can be consistent. Actual consistency requires that the likelihood is globally concave and converge there, which an examination of the equations indicates can be met with various conditions on the moments of the exogenous variables and data. Rather than gain asymptotic credibility by selecting among conditions of this sort, I adopt the more practical finite sample approach of using detailed grid searches for each of the models below to confirm that the likelihood is indeed single peaked.

bias and evaluate its impact on different models.

While the assumption of a diagonal **V** is the basis of identification, the degree to which the restriction is, from a practical standpoint, binding can be assessed. Below, having estimated values of σ and η within each model nesting and parameter restriction structure, I then take these as known and reestimate all other parameters of the model allowing **V** to have an unrestricted structure. I find the principal results change very little. When calculating covariance estimates or testing null hypotheses, I implement two wild bootstraps: one which imposes a diagonal covariance structure and another that retains the off-diagonal covariance found in the estimated productivity shock residuals. Again, with rare exceptions (highlighted below), results are virtually identical. Similarly, in evaluating bias I use wild & parametric bootstraps with both the empirical diagonal and off-diagonal covariance of errors. While results differ numerically, they rarely differ meaningfully, as substantive evaluations of model results are largely unchanged. These stress tests do not prove that the underlying covariance of the factor augmenting productivity shocks is diagonal. They do show, however, that there exist reasonable elasticity estimates that produce estimated factor augmenting shocks that are, practically speaking, all but diagonal. In addition to the above, in the tables below I of course also report the off-diagonal correlation of estimated shocks for the reader to see.

c. Statistical Inference and Model Selection

The statistical distribution of parameters estimated in the VEC model described in (III.9b) is complicated by the presence of regressors in the cointegrating equation that are non-stationary with infinite asymptotic variance. Existing results depend upon the specification and vary on a case by case basis (see Johansen 1995), are asymptotic, and have huge size distortions in my sample sizes, as shown in Monte Carlos in the on-line appendix. Furthermore, the model presented above is non-standard in that it includes the estimation of additional parameters in $\hat{\mathbf{E}}_{it}^{-1}$ and, in an effort to be true to the heavy tails of the data, the t-distribution. I bypass developing further case by case asymptotic theory of dubious finite sample validity by using wild bootstraps to evaluate the finite sample distribution of coefficient estimates and test statistics. These wild bootstraps allow for departures from the assumption of a diagonal covariance error structure, as well as misspecification of the functional form of the error distribution. Monte Carlos in the on-line appendix find that these procedures, when used with t-distribution estimation which limits the influence of outliers, yield finite sample null rejection probabilities that are very much closer to nominal value than those provided by asymptotic theory.

To be more specific, having estimated the parameters of a multi-equation model using maximum likelihood techniques, I multiply the estimated factor augmenting shocks by -1 or +1 with a 50/50 probability, use these to create new predicted values in (III.13), and reestimate the model. This procedure is repeated 200 times and the distribution of coefficients or likelihood ratios used to evaluate the same statistics for the original sample. I allow for two error structures. In the first, I impose the assumption of a diagonal covariance used in the identification of $\hat{\mathbf{E}}_{it}^{-1}$ by multiplying each shock by an individually drawn ±1 random variable. In the second, I multiply the entire J x 1 vector of factor augmenting shocks for an industry x year observation by a common ±1 random variable. This preserves any off-diagonal correlation. When juxtaposed against each other, the two methods allow the reader to evaluate the practical importance of the assumption of a diagonal covariance structure in the neighbourhood of the parameter point estimates. The use of estimated shocks, moreover, allows the error distribution to deviate from the normal or t- specified in the likelihood. Thus, reported statistical tests are under the null that the estimated model is generally true, but possibly misspecified in constraining the covariance of the errors to be zero and restricting these to the normal or t-distribution.

Johansen (1995) proposes testing the existence and degree of cointegration using a sequence of tests that compare the likelihood of the rank M < J cointegration model with that of the full rank J cointegration model. In each case the test is the null of cointegration rank equal to M against the alternative of cointegration rank greater than M, and the test procedure begins by testing M = 0 (VAR), continuing up through the integers until it fails to reject a particular level of M. If the test rejects M = J - 1, then cointegration is taken to be of full rank J, which is equivalent to saying that the data are, in fact, trend and not difference stationary. I perform this sequence of tests by, in each case, comparing the test statistic for the sample with the distribution produced by a wild bootstrap where the data generating process is based upon the point estimates of the tested null and the shocks or vectors of shocks are multiplied by independent ±1 variables in the manner described above. This helps identify cases where the alternative hypothesis "looks better" than the restricted model not because the restrictions of the model are unwarranted, but because it misspecifies the error process. One might not, for example, want to reject the VAR model in favour of the VEC framework just because the VAR error process is incorrectly specified when there is actually no cointegration of any sort.

As noted above, choosing among different multi-factor nesting structures and elasticity restrictions (> or < 1) involves comparing non-nested models. In such circumstances, economists often use the Akaike information criterion (AIC), motivated by the Kullback-Leibler information criterion (KLIC), which is the difference between the expectation of the ln-likelihood for the true model minus that of the misspecified model. The AIC is non-statistical, selecting the model with the lowest value of -2*lnL +2*k*, where *k* is the number of estimated parameters, no matter how close the nearest competitor. Vuong (1989), however, provides a more statistically grounded procedure.

Vuong (1989) considers the case of choosing between two, possibly both incorrect, models whose pseudo-parameters converge to fixed values as the sample size increases. With ln L_{ij} denoting the ln-likelihood of observation *i* under model *j* and *n* the number of observations, Vuong shows that the test statistic

$$\text{(III.15)} \sum_{i=1}^{n} L_{i1} - L_{i0} \bigg/ \sqrt{\sum_{i=1}^{n} (L_{i1} - L_{i0})^2}$$

asymptotically has a standard normal distribution under the null that the two models have the same KLIC and almost surely converges to +∞ or -∞ as models 1 or 0, respectively, have the lower KLIC. For competing non-nested models with the same number of parameters, as is always the case below, models selected on Vuong's criterion always meet the AIC, but Vuong's approach also identifies cases where differences based upon the AIC are not statistically meaningful.

Simulating the finite sample distribution of Vuong's test statistic is difficult, as the null does not require that either model be correct. I calculate the distribution using the wild bootstrap data generating processes described above based upon the parameter estimates and shocks of the lower ranked model and use it to construct a "p-value" of the test statistic. This p-value indicates the probability a Vuong difference greater than or equal to that observed in the sample would arise if the only (possible) misspecification in model 0 is with regards to the functional form of the error process and (when the bootstrap does not impose a diagonal covariance) the covariance of the shocks. I find that models that are ranked low by differences in ln likelihoods or Vuong's test statistic are, when evaluated using the wild bootstrap, sometimes observationally equivalent to the very best. While Vuong's asymptotic theory chooses whichever model has the lowest KLIC divergence, using the wild bootstrap to evaluate his test statistic tries to avoid decisions that are driven by misspecifications of the error process rather than fundamental misspecifications of the range of the elasticity of substitution (< or > 1) or the factor nesting structure.

IV. Results

This section estimates the model described above using the BEA KLEMS industry-level capital (K), labour (L) and intermediate input (I) data for 1987-2021. As noted earlier, I assume that gross output for industry i at time t is characterized by the nested production function $Q_{it} = F^i[G^i(A_{1it}X_{1it}, A_{2it}X_{2it}), A_{3it}X_{3it}]$, with F^i and G^i constant returns to scale aggregators with elasticities of substitution within the range of the data of σ between inputs 1 and 2 and η between input 3 and the aggregator of 1 and 2, respectively. Convex parameter spaces are created by restricting σ and η to be greater or less than 1. In total, there are 12 non-nested models of the data generating process depending upon which of K, L or I is factor 3 and the elasticity restrictions. To identify which model is being discussed, I use the notation $((K,L)^{\sigma>1},I)^{\eta<1}$, wherein the parentheses indicate the nested factor groupings, the superscripted Greek letters the associated local elasticities of substitution within those groupings, and the inequalities relative to 1 the parameter restrictions.

Table 5 begins by providing insight into what identifies the highest likelihood model by, following the discussion in section III, providing results for the VAR normal framework with factor shares set artificially to their mean values, so that the horse race is completely reduced to one of matching the residual covariance of the data, given in the top line of the table. As noted in section II, in the KLEMS data the residual volatility of the growth of intermediate input is an order of magnitude greater than that of other factors and total factor productivity. This induces a high residual volatility in $g(I/L)$. The challenge faced by the models is to explain this volatility without inducing residual error into the capital labour ratio and total factor productivity. The $((K,L),I)$ nesting isolates intermediate input from capital and labour, allowing the model to match the volatility of intermediate input without affecting the volatility of $g(K/L)$, as seen in Table 5. The other frameworks, in trying to match the volatility of $g(I/L)$, end up polluting and exaggerating the volatility of $g(K/L)$, as well as its covariance with $g(I/L)$ and $g(TFP)$.

Table 5 shows that despite the restriction of a diagonal covariance matrix for factor augmenting shocks, with mean income shares as regressors the $((K,L),I)$ nesting can very closely match the unconstrained residual covariance structure of the data given in the top row. However, with mean income shares, for every η there are pairs of σ on opposite sides of 1 with identical likelihoods, as noted earlier and seen in the table.[14] Selection across these variants depends upon variation in factor income

[14]When the paired value with $\sigma < 1$ hits the non-negativity constraint, the paired model has a lower likelihood and modifies the value of η, as occurs with the $((I,L),K)$ and $((K,I),L)$ nesting structures in the table.

Table 5: Residual Covariance Matrix of Normal VAR Models against that of the Data

	ln L	σ	η	variance (1) *K/L*	variance (2) *I/L*	variance (3) *TFP*	covariance (1),(2)	covariance (1),(3)	covariance (2),(3)	ln L	σ	η
data:				.0028	.0126	.0015	.0007	-.0003	-.0013			
models:	setting factor shares equal to mean values									actual factor shares		
1: $((K,L)^{\sigma<1}I)^{\eta<1}$	8259	.573	.000	.0028	.0093	.0023	.0006	-.0002	-.0013	7915	.430	.000
2: $((K,L)^{\sigma<1}I)^{\eta>1}$	8385	.589	2.46	.0028	.0130	.0015	.0012	-.0001	-.0013	7863	.425	2.59
3: $((K,L)^{\sigma>1}I)^{\eta<1}$	8259	1.46	.000	.0028	.0093	.0023	.0006	-.0002	-.0013	7646	1.58	.000
4: $((K,L)^{\sigma>1}I)^{\eta>1}$	8385	1.42	2.46	.0028	.0130	.0015	.0012	-.0001	-.0013	7860	1.56	2.53
5: $((I,L)^{\sigma<1}K)^{\eta<1}$	7756	.000	.413	.0068	.0101	.0025	.0042	-.0014	-.0026	7638	.000	.234
6: $((I,L)^{\sigma<1}K)^{\eta>1}$	7433	.000	1.55	.0091	.0106	.0021	.0057	.0013	.0009	7123	.000	1.77
7: $((I,L)^{\sigma>1}K)^{\eta<1}$	7853	2.04	.293	.0075	.0126	.0018	.0054	-.0014	-.0025	7078	2.18	.000
8: $((I,L)^{\sigma>1}K)^{\eta>1}$	7489	2.37	1.64	.0104	.0126	.0016	.0072	.0010	.0005	7401	2.45	1.69
9: $((K,I)^{\sigma<1}L)^{\eta<1}$	7665	.000	.124	.0077	.0086	.0022	.0025	.0006	-.0017	7606	.000	.120
10: $((K,I)^{\sigma<1}L)^{\eta>1}$	7561	.132	1.68	.0106	.0080	.0021	.0023	-.0028	-.0009	6447	.000	2.07
11: $((K,I)^{\sigma>1}L)^{\eta<1}$	7737	1.71	.033	.0082	.0091	.0017	.0017	.0010	-.0015	7199	1.86	.000
12: $((K,I)^{\sigma>1}L)^{\eta>1}$	7561	2.48	1.68	.0106	.0080	.0021	.0023	-.0028	-.0009	7192	2.27	1.93

Notes: Residual covariance matrix of the data and structural models is net of first order vector autoregression lags and industry and year fixed effects, in the latter case for underlying factor augmenting technical change.

shares. Here the data favour more extreme elasticities of substitution than are used to match mean moments, as estimated elasticities almost always fall and rise when restricted to be < 1 or > 1, respectively. With actual shares model 1 (σ<1,η<1) has a higher likelihood than other variants of the ((*K*,*L*),*I*) nesting structure, although model 2 (σ<1, η>1), which has the highest likelihood with mean shares, is a close second

Table 6 presents the ln-likelihoods using actual factor shares of the vector autoregression (VAR) and vector error correction with rank one or two cointegration (VEC1 & VEC2) models, as in (III.9a) and (III.9b) earlier above, assuming the factor augmenting shocks are multivariate normal or *t* with diagonal covariance matrix.[15] The listing of models follows that of Table 5, with those with the highest and second highest likelihoods across all specifications conveniently designated as models 1 and 2, respectively. In all models the likelihood with the *t* distribution is consistently 1700 to 2400 ln points higher than with the normal and point estimates of the *t* degrees of freedom (shown for the t-distribution VEC1 model) are close to 2, i.e. the errors are very non-normal with fat tails and not much more than second moments. However, point estimates of the elasticities σ and η vary little between the normal, t, VAR, VEC1 and VEC2 specifications. Given the data, these completely determine

[15] As the models are non-linear, in addition to maximizing using Newton's method I concentrate the likelihoods as functions of σ and η, or these + the t- degrees of freedom, and conduct detailed grid searches (10k or 16k points), finding that the likelihood is single peaked for each of the 72 models in Table 5. Grid searches do not guarantee uniqueness of the max, especially as concentrating the likelihood for the t- models involves solving non-linear systems, but this is at least encouraging.

Table 6: Results by Nesting Structure x Quadrant Model

(A) ln likelihoods & bootstrap p-values of cointegration rank tests H_0: rank = r vs H_1: rank > r

Model	normal VAR	normal VEC1	normal VEC2	t - distribution VAR	t - distribution VEC1	t - distribution VEC2	normal distribution r = 0	normal distribution r = 1	normal distribution r = 2	*t* - distribution r = 0	*t* - distribution r = 1	*t* - distribution r = 2
1: $((K,L)^{\sigma<1}I)^{\eta<1}$	7915	7983	8020	9781	9837	9867	.015/.050	.395/.260	.885/.875	.000/.000	.070/.050	.635/.550
2: $((K, L)^{\sigma<1}I)^{\eta>1}$	7863	7925	7968	9701	9760	9786	.005/.055	.320/.275	.915/.925	.000/.000	.125/.115	.640/.565
3: $((K,L)^{\sigma>1}I)^{\eta<1}$	7646	7696	7732	9553	9589	9619	.080/.165	.250/.145	.860/.845	.000/.010	.020/.020	.583/.510
4: $((K,L)^{\sigma>1}I)^{\eta>1}$	7860	7920	7964	9604	9657	9687	.010/.040	.310/.270	.980/.970	.000/.010	.165/.175	.950/.889
5: $((I,L)^{\sigma<1}K)^{\eta<1}$	7638	7730	7769	9571	9642	9673	.000/.000	.425/.055	.900/.650	.000/.000	.076/.010	.920/.645
6: $((I,L)^{\sigma<1}K)^{\eta>1}$	7123	7213	7242	9151	9212	9237	.000/.010	.710/.120	.955/.150	.000/.000	.115/.045	.860/.442
7: $((I,L)^{\sigma>1}K)^{\eta<1}$	7078	7169	7225	9184	9263	9292	.000/.000	.070/.005	.925/.425	.000/.000	.186/.135	.912/.596
8: $((I,L)^{\sigma>1}K)^{\eta>1}$	7401	7481	7527	9243	9302	9336	.000/.000	.215/.015	1.00/.595	.000/.005	.035/.040	.975/.700
9: $((K,I)^{\sigma<1}L)^{\eta<1}$	7606	7698	7735	9562	9631	9662	.000/.000	.450/.180	.945/.665	.000/.000	.075/.025	.930/.563
10: $((K,I)^{\sigma<1}L)^{\eta>1}$	6447	6526	6584	8865	8914	8944	.025/.085	.220/.040	1.00/.435	.000/.010	.075/.141	.950/.558
11: $((K,I)^{\sigma>1}L)^{\eta<1}$	7199	7275	7328	9347	9391	9428	.020/.080	.140/.010	.975/.670	.000/.010	.040/.072	.934/.688
12: $((K,I)^{\sigma>1}L)^{\eta>1}$	7192	7270	7327	9109	9159	9198	.000/.025	.120/.035	1.00/.685	.000/.005	.020/.040	.990/.720

(B) elasticities and factor augmenting productivity growth estimates

	normal VAR σ	normal VAR η	normal VEC1 σ	normal VEC1 η	normal VEC2 σ	normal VEC2 η	t - VAR σ	t - VAR η	t-distribution VEC1 σ	t-distribution VEC1 η	t-distribution VEC1 *dof*	t-distribution VEC1 $g(A_K)$	t-distribution VEC1 $g(A_L)$	t-distribution VEC1 $g(A_I)$	t - VEC2 σ	t - VEC2 η
1:	.430	.000	.419	.000	.418	.000	.436	.000	.425 (.013/.008)	.000 (.003/.000)	2.05	-.010	.017	-.000	.423	.000
2:	.425	2.59	.419	2.61	.420	2.59	.475	2.45	.469 (.015/.008)	2.46 (.013/.018)	2.06	-.013	. 016	. 002	.467	2.45
3:	1.58	.000	1.59	.000	1.59	.000	1.55	.000	1.56 (.007/.007)	.000 (.000/.000)	1.94	.013	.001	-.000	1.56	.000
4:	1.56	2.53	1.56	2.53	1.56	2.53	1.53	2.44	1.52 (.008/.007)	2.43 (.011/.018)	2.11	.012	-.001	.002	1.53	2.44
5:	.000	.234	.000	.229	.000	.232	.000	.196	.000 (.009/.000)	.194 (.028/.011)	1.93	-.006	.014	-.000	.000	.191
6:	.000	1.77	.000	1.78	.000	1.77	.000	1.68	.000 (.000/.000)	1.68 (.009/.010)	1.79	.008	.009	-.004	.000	1.68
7:	2.18	.000	2.17	.000	2.18	.000	2.10	.076	2.10 (.010/.013)	.062 (.100/.013)	1.85	-.005	.007	.005	2.10	.067
8:	2.45	1.69	2.43	1.69	2.45	1.70	2.33	1.66	2.32 (.011/.014)	1.65 (.010/.012)	1.88	.008	.003	.001	2.34	1.66
9:	.000	.120	.000	.113	.000	.112	.000	.140	.000 (.000/.000)	.134 (.027/.009)	1.93	-.005	.014	-.001	.000	.129
10:	.000	2.07	.000	2.05	.000	2.06	.015	1.95	.001 (.000/.016)	1.94 (.010/.012)	1.73	-.000	.002	.005	.000	1.95
11:	1.86	.000	1.86	.000	1.86	.000	1.83	.000	1.82 (.010/.009)	.000 (.039/.002)	1.88	.001	.014	-.004	1.83	.000
12:	2.27	1.93	2.25	1.93	2.27	1.93	2.20	1.89	2.19 (.011/.015)	1.89 (.010/.011)	1.88	.005	.002	.003	2.20	1.90

Notes: VECr = VEC with rank *r* cointegration. *dof* = t-degrees of freedom. Wild bootstrap p-values in (A) and standard errors n parentheses in panel (B) are with diagonal covariance *dgp*/with unrestricted covariance *dgp*, as described in the text. $g(A_i)$ = average rates weighting industry rates by shares of factor income.

rates of factor augmenting productivity growth through (III.12) above, which consequently do not differ much across specifications. Average economy-wide rates are given for the t-VEC1 model. In 7 of 12 models substantially positive factor augmenting productivity growth in one factor is offset by substantially negative productivity growth in another. In top ranked models 1 and 2, the tradeoff is concentrated in labour and capital, as average intermediate input productivity growth is near zero.

The right hand side of panel (A) in Table 6 reports p-values for Johansen's sequential testing procedure described in section III, which proceeds up through the integers until one is unable to reject the null of rank r cointegration at the test's nominal level. As explained earlier, as specifics differ from Johansen's test and Monte Carlos in the on-line appendix anyway show enormous finite sample size distortions when using asymptotic distributions in panels of this size, p-values are based on the wild bootstrap with 200 draws from its distribution.[16] The first bootstrap number in each p-value pair multiplies each factor augmenting shock by an independent ±1, imposing the assumption of a diagonal covariance matrix used in estimation, while the second number multiplies each industry x time triplet of factor augmenting shocks by a common ±1, retaining the non-zero covariance of the shocks present in the residuals. Monte Carlos in the on-line appendix find that the wild bootstrap performs erratically when the likelihood is normal, but is relatively reliable and consistent when the errors are fat tailed t-distributed and the likelihood evaluated using the t-distribution (which underweights outliers). Using both diagonal and unrestricted (correlated) shocks with both diagonal and unrestricted bootstrap procedures, i.e. with each bootstrap alternately correctly or incorrectly specifying the underlying data generating process, I find that with the t-distribution a nominal cutoff of .01 (i.e. a p-value ≤ .01) consistently ensures no more than a .051 empirical null rejection probability in all eventualities.

With the .01 cutoff in mind, we see in Table 6 that with t-likelihoods the two wild bootstrap error data generating processes produce the same decisions at the .05 empirical level as, despite a few borderline .01 p-values, they always reject the rank 0

[16]When sampling a distribution an exact p-value (relative to the distribution) is given by $p = (G + u(T+1))/(N+1)$, where N represents the number of draws, G the number of greater outcomes, T the number of tied outcomes & u is uniformly distributed on (0,1) (see Jockel 1986 & the online appendix of Young 2019). Given the large number of potential outcomes in my large sample, ties are not an issue, while in many tests G is 0. To make the latter clearer, throughout the paper I calculate p-values as G/N, so that such p-values are reported as .000, rather than using a u to add a random number between .000 and .005 to every p-value. Rare likelihoods that do not converge are dropped (i.e. N is sometimes less than 200), so not every p-value is a multiple of .005.

VAR model in favour of rank > 0 cointegration and accept the rank 1 cointegration VEC1 model in preference to higher rank cointegration. This is especially clear for top ranked models 1 and 2, but even models 3, 4, 8, 10 & 12, which do not find strongly negative productivity growth for any factor in the lower panel of the table, have t-distribution results which favour rejecting the VAR framework in favour of VEC1 at the .05 empirical level. With universally large p-values for the rank 2 cointegration test, no model rejects rank 2 cointegration in favour of rank 3, which would imply the absence of unit roots.

Table 7 gives estimates of degrees of freedom, cointegration parameters **β** & **α** and off-diagonal shock correlations for the 12 forms of the t-distribution VEC1 model. Panel (A) presents results using the baseline procedure of assuming a diagonal covariance matrix for the factor augmenting shocks so that σ and η are identified and estimated, as reported in Table 6, while panel (B) takes the values of σ and η as known at the values found in panel (A) and estimates the remaining parameters allowing for off-diagonal correlations between the shocks. While off-diagonal shock correlations are assumed to be zero in the likelihood of panel (A), their empirical values can be calculated using the residuals for either procedure,[17] as is done in the table. Comparing panels (A) and (B) we see that for top ranked models 1 and 2, where the correlations between residuals are small, given the point estimates of σ and η the assumption of a diagonal covariance matrix has very little impact on the cointegration parameter estimates. In lower-likelihood ranked models with larger off-diagonal correlations, the impact can be greater, although mostly when the **β** cointegration parameters are estimated with imprecision using the assumption of a diagonal covariance matrix. Outside of 7 instances in models 3, 10 & 11, all point estimates of **β** and **α** parameters in panel (A) lie in the 95 percent confidence interval of those estimated in panel (B), and outside of 8 instances in the same three models, all point estimates of the same in panel (B) lie in the 95 percent confidence interval of those estimated in panel (A). Thus, conditional on the estimates of σ and η, in most cases point estimates of the cointegration parameters under different assumptions regarding the diagonality of shocks are statistically indistinguishable.

The signs and magnitudes of the parameters **β** defining the cointegrating frontier $\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) = c_{it}$ and **α** governing the adjustment process to deviations from that frontier have no independent meaning as they would all change if instead of normalizing β_K to 1, as is donein Table 7, we were instead to normalize it to, e.g, -½.

[17]I follow *t*-distribution maximum likelihood estimation and calculate the covariance matrix of the residuals using the weights described earlier above.

Table 7: t - VEC1 Parameter Estimates & Shock Correlation

	β_L	β_I	α_K	α_L	α_I	ρ_{KL}	ρ_{KI}	ρ_{LI}
(A) estimating σ and η (likelihood assumes diagonal covariance matrix of shocks)								
1:	1.23 (.219)	-.994 (.219)	-.026 (.006)	-.014 (.004)	.013 (.004)	.147 (.028)	-.074 (.026)	-.146 (.026)
2:	.533 (.230)	2.73 (.944)	-.009 (.005)	-.006 (.003)	-.014 (.003)	.175 (.024)	-.006 (.026)	.039 (.025)
3:	.957 (.290)	-1.75 (.389)	-.007 (.005)	-.011 (.004)	.018 (.005)	.230 (.022)	-.108 (.023)	-.096 (.029)
4:	1.28 (10.9)	-9.84 (84.5)	-.002 (.003)	.001 (.002)	.006 (.004)	.206 (.020)	.100 (.024)	.066 (.021)
5:	1.66 (.284)	-1.02 (.221)	-.024 (.005)	-.014 (.003)	.015 (.004)	.398 (.026)	-.194 (.027)	-.104 (.026)
6:	-8.76 (3.38)	4.49 (2.12)	-.007 (.004)	.006 (.003)	-.002 (.001)	-.350 (.025)	.344 (.027)	-.034 (.028)
7:	.071 (.127)	1.76 (.319)	-.022 (.005)	-.002 (.004)	-.021 (.003)	-.208 (.022)	.293 (.025)	.183 (.021)
8:	8.92 (337)	-20.9 (717)	-.003 (.003)	-.000 (.001)	.003 (.003)	.477 (.024)	-.211 (.025)	.086 (.021)
9:	1.24 (.220)	-.827 (.184)	-.022 (.005)	-.020 (.004)	.019 (.005)	.411 (.027)	-.053 (.027)	-.230 (.025)
10:	-.147 (.114)	.674 (.129)	-.050 (.008)	.004 (.005)	-.011 (.004)	-.340 (.026)	.168 (.028)	.412 (.021)
11:	-2.53 (.538)	-2.73 (.753)	-.012 (.007)	.008 (.003)	.005 (.002)	-.310 (.027)	.191 (.027)	.237 (.026)
12:	.898 (25.3)	-17.5 (196)	-.000 (.001)	-.001 (.001)	.003 (.006)	.566 (.023)	.043 (.028)	-.174 (.019)
(B taking σ and η as given at values estimated above (unrestricted covariance matrix of shocks)								
1:	1.06 (.241)	-1.00 (.252)	-.027 (.006)	-.013 (.004)	.013 (.004)	.214 (.008)	-.113 (.007)	-.206 (.008)
2:	.495 (.261)	3.58 (1.45)	-.007 (.004)	-.004 (.003)	-.013 (.003)	.248 (.007)	-.004 (.009)	.061 (.008)
3:	1.30 (18.1)	-6.90 (100)	.002 (.001)	-.002 (.002)	.007 (.005)	.325 (.007)	-.174 (.008)	-.153 (.008)
4:	1.11 (.486)	-6.54 (2.15)	-.004 (.003)	.001 (.002)	.010 (.003)	.294 (.008)	.157 (.009)	.103 (.009)
5:	1.43 (.452)	-1.14 (.348)	-.025 (.006)	-.013 (.003)	.016 (.004)	.547 (.006)	-.273 (.007)	-.175 (.008)
6:	-8.68 (102)	3.42 (60.4)	-.007 (.005)	.006 (.005)	-.002 (.002)	-.472 (.007)	.463 (.007)	-.083 (.008)
7:	.153 (.185)	1.96 (.570)	-.020 (.006)	-.002 (.005)	-.020 (.004)	-.261 (.008)	.378 (.007)	.241 (.009)
8:	7.66 (58.0)	-26.4 (228)	-.002 (.004)	-.000 (.001)	.002 (.004)	.622 (.006)	-.260 (.008)	.070 (.009)
9:	1.13 (.360)	-1.19 (.303)	-.021 (.006)	-.017 (.005)	.020 (.005)	.558 (.005)	-.115 (.007)	-.315 (.008)
10:	-.129 (.133)	.333 (.148)	-.056 (.009)	.008 (.007)	-.008 (.004)	-.424 (.008)	.178 (.008)	.515 (.006)
11:	-1.44 (.245)	-.429 (.266)	-.041 (.010)	.013 (.004)	.003 (.004)	-.391 (.008)	.250 (.009)	.297 (.008)
12:	.445 (2.37)	-5.41 (17.8)	-.003 (.003)	-.003 (.003)	.012 (.007)	.724 (.004)	.026 (.009)	-.212 (.009)

Notes: Standard errors (in parentheses) calculated using the wild bootstrap with diagonal covariance (A) and unconstrained covariance (B), following the assumptions of the estimation procedures. Standard errors for panel (A) calculated using a wild bootstrap with unconstrained covariance are similar and are given in the on-line appendix.

Only relative signs and magnitudes matter. When factors share the same sign in $\boldsymbol{\beta}$, then from $\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) = c_{it}$ we see that movements along the frontier involve tradeoffs between measured factor augmenting productivities, whereas when they differ in sign movements along the frontier involve complementary changes in factor augmenting productivity. As seen in the table, with the exception of models 6, 10 & 11 with lower likelihoods (Table 6), the cointegration frontiers generally involve tradeoffs between measured capital and labour productivity, with $\beta_L > 0$. Model 1 (and ⅔ of all models) shows complementarities between capital and intermediate inputs with $\beta_I < 0$, but second ranked model 2 actually finds tradeoffs between all three factors with β_K, β_L & β_I all > 0.

Table 8 turns to a general consideration of bias. As argued above, given the use of year fixed effects, endogenous factor prices play no meaningful role in identification. To illustrate this, when the 72 VAR & VEC, normal & t- models of Table 6 are reestimated with the growth of factor prices removed from the estimating

Table 8: Wild & Parametric Bootstrap Estimates of Bias in t - VEC1 Model
(mean difference between point estimates and parameters of *dgp* in 200 samples)

	σ	η	β_L	β_I	α_K	α_L	α_I	σ	η	β_L	β_I	α_K	α_L	α_I
	(A) wild imposing diagonal covariance							(B) wild with unrestricted covariance						
1:	-.064	.000	.012	-.065	-.009	-.006	.006	.038	-.000	.107	-.206	-.008	-.006	.006
2:	-.099	.080	-.048	.354	-.004	-.001	-.003	.043	.002	.085	.727	-.004	-.001	-.002
3:	.117	.000	.015	-.092	-.005	-.005	.009	-.058	-.000	.051	-.328	-.006	-.004	.005
4:	.119	.082	-.571	7.25	-.002	-.000	.005	-.062	-.072	1.58	-11.18	-.001	-.000	.001
5:	.000	-.049	-.076	-.063	-.007	-.005	.005	.000	.031	.270	-.261	-.007	-.005	.005
6:	.000	.021	.723	.095	-.002	.004	-.001	.000	.000	-.357	.985	-.005	.005	-.002
7:	.139	-.049	-.123	.107	-.006	-.002	-.004	-.066	-.001	-.058	.101	-.008	-.000	-.007
8:	.120	.053	9.534	-18.41	-.002	-.000	.003	-.085	-.057	21.72	-40.88	-.002	-.001	.002
9:	.000	-.011	.003	-.092	-.007	-.004	.006	.000	.011	.174	-.182	-.006	-.008	.007
10:	-.001	.046	-.079	-.065	-.019	-.000	-.002	.075	-.075	-.079	.045	-.021	.006	-.004
11:	.046	.002	.454	.448	-.010	.004	.003	-.047	.001	-.162	.289	-.008	.004	.003
12:	.078	.103	-1.244	22.63	-.001	-.000	.005	-.072	-.083	1.18	-.415	-.001	-.001	.001
	(C) parametric imposing diagonal covariance							(D) parametric with unrestricted covariance						
1:	-.001	.011	.029	-.046	-.003	-.003	.002	.055	.000	.013	-.097	-.004	-.002	.002
2:	.000	-.001	.018	.320	-.001	-.000	-.001	.069	-.025	.060	.552	-.001	.000	.000
3:	-.002	.009	.012	-.176	-.002	-.001	.003	-.082	.000	.041	-.320	-.001	-.001	.001
4:	.000	-.004	.140	-.506	-.001	-.000	.001	-.086	-.108	.201	-2.98	-.000	-.000	-.000
5:	.009	.001	.032	-.039	-.003	-.002	.002	.000	.062	.193	-.102	-.003	-.001	.001
6:	.010	.000	.498	-.082	-.001	.002	-.000	.013	-.027	-.109	.340	-.001	.001	-.000
7:	-.000	.002	-.019	.027	-.003	-.000	-.002	-.166	.053	.076	.002	-.002	.002	-.003
8:	-.001	-.000	-1.41	3.78	-.001	-.000	.001	-.148	-.079	-3.39	9.31	-.000	-.000	.000
9:	.011	.006	.039	-.030	-.003	-.002	.002	.003	.018	.037	-.027	-.002	-.003	.002
10:	.008	-.001	-.017	.004	-.006	.000	-.001	.209	-.135	.030	.154	-.006	.003	.002
11:	-.003	.008	-.072	.025	-.003	.001	.001	-.120	.023	-.830	-.484	-.002	-.000	.000
12:	.001	.001	.267	-3.05	-.000	-.000	.001	-.120	-.120	.180	13.86	-.000	-.000	-.000

Notes: *dgps* based upon point estimates of models and wild bootstrap ±1 transformations of estimated residuals or parametric bootstrap residuals drawn from the t-distribution with degrees of freedom estimated in the models.

equations, the mean of the 144 elasticity estimates changes only slightly, from 1.0533 to 1.0541, with a correlation of .9999 between the two sets of estimates. There remains, however, the issue of the bias introduced by individual fixed effects in short panel data (Nickell 1981), the bias due to the failure to account for the off-diagonal correlation of shocks (notwithstanding the sample specific results of Table 7), and the more general possibility of bias in maximum likelihood estimation. To this end, Table 8 reports the bias found in 200 sample draws from *dgps* based upon the point estimates of the models with either wild bootstrap disturbances formed by ±1 transformations of estimated residuals or parameteric bootstrap disturbances drawn from the t-distribution with the degrees of freedom estimated in the models. The covariance matrix of shocks is either constrained to be diagonal or is unrestricted, following the covariance structure found in the estimated residuals. The parametric bootstrap eliminates bias due to misspecification of the error distribution and provides a narrower estimate of bias due to short panel fixed effects and maximum likelihood estimation alone, which might be of interest to some readers.

As seen in Table 8, while bias generally appears small, particularly in highest ranked model 1, there are some spectacular outliers, most notably the estimates of the cointegrating parameters β_L and β_I in models 4, 8 & 12 and, to a lesser extent, 6 & 11. In these cases, however, the parameters are estimated with a great deal of imprecision in the first place (Table 6), making such differences both statistically less meaningful and less reliably estimated with the computationally affordable 200 sample draws. More generally, the absolute value of bias tends to be greater when the disturbances are correlated rather than diagonal and are wild transformations of the empirical residuals rather than parametrically drawn from the exact distribution specified in the likelihood.[18] Despite their varying signs & magnitudes, section V further below finds that, with the exception of model 4, adjustment for these estimates of bias does not substantively affect conclusions regarding the contribution of movements along the cointegration frontier to aggregate productivity growth.

Table 9 considers selection across the 12 models using Vuong's test statistic and bootstrapped p-values. Panel (A) compares model 1 against the remaining models, while panel (B) compares model 2 against the remaining models. Based upon the asymptotic standard normal distribution of Vuong's test statistic, all of the t-distributed models select in favour of model 1 in one-sided tests at the .05 level. The table also provides wild bootstrapped "p-values", which calculate the probability of a Vuong statistic greater than that found in the sample using the point estimates and estimated shocks of the alternative (to model 1 or 2) model given on each row as the *dgp*. Two p-values are given, using either a diagonal or empirical covariance of the shocks. While both p-values are almost always 0 (i.e. in the bootstrap simulations no instance of a test statistic greater than that reported was found using the model of each row as the *dgp*), this is the one place where there is a substantive difference between the results given by the two *dgps*. Allowing for a non-diagonal covariance between the shocks, we see that misspecification of the model (i.e. the assumption of independent shocks) makes it highly probable that model 1 would be selected in preference to model 5 or 9 (where also $\sigma < 1$ & $\eta < 1$) when the latter model is the true *dgp*. However, despite their different nesting structure, models 5 and 9 have almost identical estimates of the cointegration parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ rates (Table 7) and similar patterns of negative capital, positive labour and zero intermediates factor augmenting

[18]The sum of the 84 absolute biases in each panel are: (A) 64, (B) 82, (C) 11, & (D) 35.

Table 9: Vuong's Test Statistic and Wild Bootstrap P-Values

	test statistic				wild bootstrap p-values			
	normal distribution		*t* - distribution		normal distribution		*t* - distribution	
	VAR	VEC1	VAR	VEC1	VAR	VEC1	VAR	VEC1
(A) model 1 vs remaining models as nulls								
1:								
2:	0.5	0.6	1.8	1.7	.000/.000	.000/.000	.000/.000	.000/.000
3:	2.1	2.3	5.5	5.7	.000/.000	.000/.000	.000/.000	.000/.000
4:	0.6	0.7	3.9	3.9	.000/.000	.000/.000	.000/.000	.000/.000
5:	4.5	4.2	9.1	8.7	.000/.765	.000/.700	.000/.415	.000/.265
6:	6.2	6.2	12.9	12.6	.000/.000	.000/.000	.000/.000	.000/.000
7:	7.0	6.6	13.5	12.4	.000/.000	.000/.000	.000/.000	.000/.000
8:	4.5	4.4	10.7	10.5	.000/.000	.000/.000	.000/.000	.000/.000
9:	5.0	4.8	9.6	9.1	.005/.760	.000/.730	.000/.670	.000/.605
10:	7.6	7.3	16.5	16.4	.000/.000	.000/.000	.000/.000	.000/.000
11:	3.7	3.7	8.9	9.0	.000/.000	.000/.000	.000/.000	.000/.000
12:	5.0	5.0	12.3	12.3	.000/.000	.000/.000	.000/.000	.000/.000
(B) model 2 vs remaining models as nulls								
1:	-0.5	-0.6	-1.8	-1.7	.000/.000	.000/.000	.000/.000	000/.000
2:								
3:	2.4	2.4	3.6	3.9	.000/.000	.000/.000	.000/.000	.000/.000
4:	0.0	0.0	2.2	2.3	.000/.000	.000/.000	.000/.000	.000/.000
5:	1.9	1.6	2.5	2.2	.000/.000	.000/.000	.000/.000	.000/.000
6:	5.1	5.1	9.2	8.6	.000/.000	.000/.000	.000/.000	.000/.000
7:	6.1	5.9	10.1	9.7	.000/.000	.000/.000	.000/.000	.000/.000
8:	3.8	3.7	9.2	8.6	.000/.010	.000/.000	.000/.000	.000/.000
9:	2.2	2.0	2.7	2.4	.000/.000	.000/.000	.000/.000	.000/.000
10:	7.1	6.9	13.9	14.0	.000/.000	.000/.000	.000/.000	.000/.000
11:	3.2	3.2	6.3	6.4	.000/.000	.000/.000	.000/.000	.000/.000
12:	4.5	4.6	11.4	11.2	.000/.010	.000/.000	.000/.000	.000/.000

Notes: Wild bootstraps p-values based upon diagonal covariance/unrestricted covariance use the data generating process of the alternative model listed in each row.

productivity growth (Table 6) as model 1. Moreover, this ambiguity does not arise when the models are compared to second ranked model 2, where the p-value of getting a Vuong difference greater than or equal to that found in the sample is zero under the null of the row model *dgp* regardless of the covariance structure of shocks.

The top rows of panels (A) and (B) in Table 9 highlight the difference between the nulls underlying Vuong's test and the bootstrap tests given in the table. In panel (A) we see that if model 2 is the true *dgp* the bootstrapped probability of the recorded Vuong statistic in favour of model 1 is 0, leading to the conclusion that model 2 is not the true *dgp*. In panel (B) we see that the bootstrapped probability of a Vuong statistic greater than or equal to the negative values found against model 2 when model 1 is the true *dgp* is also 0. That is, the sample test statistic should be much more negative (favour model 1 more strongly) and we can conclude that model 1 is not the true *dgp* either. The asymptotically normal distribution of Vuong's test statistic gives the probability of the estimated difference between models 1 and 2 if they have the same Kullback-Leibler distance from the true underlying data generating process, without requiring either model to be correct. Here in a one-sided test at the .05 asymptotic level and using t-likelihoods we conclude that model 1's Kullback-Leibler divergence from the true *dgp* is less than that of model 2. In combination, these results remind us that while model 1 may be the best approximation to the true *dgp*, none of the models presented here are likely to be precisely true.

V. Shifts of and Movements along the Cointegration Frontier

The t - distribution VEC1 cointegrating coefficients $\boldsymbol{\beta}$ implicitly define a frontier $\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) = c_{it}$ linking the non-stationary $\mathbf{ln}(A_{jit})$ to the trend stationary c_{it}. This naturally invites a decomposition of changes in $\mathbf{ln}(A_{jit})$ into components stemming from movements along a given frontier, that is for a given value of c_{it}, and those associated with a shift of that frontier, i.e. changes or trends in c_{it}. To provide a clarifying example, this section first revisits the model of task based technical change noted in the introduction, before presenting techniques that provide, or at least bound, such decompositions.

Consider the three factor nested CES production function where, with α_{jit} denoting the measure of tasks performed by factor j and a_{jit} factor augmenting productivity, and assuming that factor inputs are divided evenly across their tasks, we have

$$\text{(V.1)}\ \ Q_{it} = \left\{ \left[\alpha_{1it}\left(\frac{a_{1it}X_{1it}}{\alpha_{1it}}\right)^{\frac{\sigma-1}{\sigma}} + \alpha_{2it}\left(\frac{a_{2it}X_{2it}}{\alpha_{2it}}\right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}\frac{\eta-1}{\eta}} + \alpha_{3it}\left(\frac{a_{3it}X_{3it}}{\alpha_{3it}}\right)^{\frac{\eta-1}{\eta}} \right\}^{\frac{\eta}{\eta-1}}$$

$$= \left\{ \left[(\alpha_{1it}^{1/(\sigma-1)}a_{1it}X_{1it})^{\frac{\sigma-1}{\sigma}} + (\alpha_{2it}^{1/(\sigma-1)}a_{2it}X_{2it})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}\frac{\eta-1}{\eta}} + (\alpha_{3it}^{1/(\eta-1)}a_{3it}X_{3it})^{\frac{\eta-1}{\eta}} \right\}^{\frac{\eta}{\eta-1}},$$

so that

$$\text{(V.2)}\ \ Q_{it} = F(G(A_{1it}X_{1it}, A_{2it}X_{2it}), A_{3it}X_{3it})$$

$$\text{with } A_{1it} = \alpha_{1it}^{1/(\sigma-1)}a_{1it},\ A_{2it} = \alpha_{2it}^{1/(\sigma-1)}a_{2it}\ \&\ A_{3it} = \alpha_{3it}^{1/(\eta-1)}a_{3it}.$$

Task based technical change might be viewed as positing a restriction linking the number of tasks (including those yet to be invented) available to each input. Expressing this restriction in terms of elasticity based monotonic transformations, we might have

$$\text{(V.3)}\ \ T(\alpha_{1it}^{1/(\sigma-1)}, \alpha_{2it}^{1/(\sigma-1)}, \alpha_{3it}^{1/(\eta-1)}) = 0$$

or log-linearizing

$$\text{(V.4)}\quad \beta_1 \frac{1}{\sigma-1}\ln(\alpha_{1it}) + \beta_2 \frac{1}{\sigma-1}\ln(\alpha_{2it}) + \beta_3 \frac{1}{\eta-1}\ln(\alpha_{3it}) \approx c,$$

so that using (V.2) we see that

$$\text{(V.5)}\quad \boldsymbol{\beta}'\mathbf{ln}(A_{jit}) \approx c - \boldsymbol{\beta}'\mathbf{ln}(a_{jit}) = c_{it}.$$

Changes in the number of tasks α_{jit} performed by each factor that leave the left-hand side of (V.4) unchanged are movements along the task frontier T and, similarly, movements along the cointegration frontier, leaving the value of $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ and c_{it} unchanged. I refer to these below as "linked" productivity changes which generate "movements along the frontier". Changes in underlying true factor augmenting productivity a_{jit} appear as changing values of $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$. I refer to these below as "unlinked" or "shifts of the frontier".

Returning to empirics, to measure shifts of the frontier Johansen (1995) advocates projecting estimated time trends in an error correction model on the adjustment factors $\boldsymbol{\alpha}$. Following his suggestion, for our model, where estimated time trends are fixed effects, we might decompose as follows

$$\text{(V.6)}\quad \mathbf{g}(A_{jit}) = \boldsymbol{\Gamma}\mathbf{g}(A_{ijt-1}) + \boldsymbol{\alpha}[\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) + \hat{\gamma}_t + \hat{\gamma}_i] + \boldsymbol{\eta}_t^{residual} + \boldsymbol{\eta}_i^{residual} + \boldsymbol{\varepsilon}_{it},$$
$$\text{where for } l = t \text{ or } i:\ \hat{\gamma}_l = (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\boldsymbol{\eta}_l \ \&\ \boldsymbol{\eta}_l^{residual} = \boldsymbol{\eta}_l - \boldsymbol{\alpha}\hat{\gamma}_l,$$

after which, as the projections of the year fixed effects on $\boldsymbol{\alpha}$ are quite volatile, a summary Johansen-type measure of the trends in c_{it} can be arrived at by projecting the negative of these on a constant and time

$$\text{(V.7)}\quad -\hat{\gamma}_t = a_J + g_J t.$$

The logic of Johansen's suggestion is that mean effects that are "balanced" across $\boldsymbol{\alpha}$ do not generate any subsequent adjustment and hence can be interpreted as changes consistent with a shift of the frontier.

The problem with this approach is that it is univariate, i.e. assumes that there are no components in $\boldsymbol{\eta}$ which are correlated with $\boldsymbol{\alpha}$ but not associated with a trend in c_{it} and whose exclusion, therefore, biases the estimated values. Since both $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$ estimate fundamental directions of productivity growth, this assumption is problematic. In practice, as shown below, in my data Johansen's approach generates estimated trends which are multiples of observed movements of $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$, suggestive of omitted variable bias with common innovative forces determining $\boldsymbol{\alpha}$ and $\boldsymbol{\eta}$.

An alternative "proof is in the pudding" approach, intimated by the last sentence, is to simply look at trends in $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ at the industry level or $\boldsymbol{\beta}'\mathbf{ln}(A_{jt})$ at the economy-wide level, i.e. run the regressions:

(V.8) $\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) = a_{Oi} + g_O(\text{industry})t$ & $\boldsymbol{\beta}'\mathbf{ln}(A_{jt}) = a_O + g_O(\text{economy})t$,

where a_O and a_{Oi} are constants and industry fixed effects and the g_O estimates of observed trends. g_O(industry) corresponds most closely to g_J, as the latter is based on unweighted estimates of $\boldsymbol{\eta}_t$, but I also present results g_O(economy) for ln(A_{jt}) calculated as the factor income share weighted sum of industry growth rates, as this corresponds more closely to other measures examined further below. If c_{it} is trending at some common rate g_O, over long periods of time $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ should be trending at that rate as well.

Table 10 reports the estimated growth rates of c_{it} which, with β_K set equal to 1, are normalized in units of capital augmenting productivity growth. As seen in panel (A), the Johansen-type growth rates are generally larger, and in fact often 1 or 2 orders of magnitude larger, than observed empirical trends in $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$. While the $\boldsymbol{\alpha}$ adjustment process of the cointegration model allows for some divergence between the growth rates of c_{it} and $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ over periods of time, the gaps implied by the differing growth estimates in Table 10 are utterly implausible. For model 1, for example, the 4.1 percent gap between g_J and g_O(industry) suggests an average industry change in $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ - c_{it} of -1.39 from the beginning to the end of the sample. With $\boldsymbol{\alpha}$ for that model equal to (-.026,-.014,.013) & $\boldsymbol{\beta}$ to (1,1.23,-.994) (Table 7), this would imply upward pressure on $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ of .078 per annum by 2021, an enormous acceleration relative to the .005 mean growth observed during the sample period. The factor income share weighted economy wide index $\boldsymbol{\beta}'\mathbf{ln}(A_{jt})$ for model 1 has a fairly steady mean growth rate of .011, which has no obvious tendency to accelerate, as seen in Figure I. These results favour an omitted variable bias interpretation of the g_J estimates.

Estimates of g_O in panel (B) of Table 10 taking σ and η as known and allowing an unconstrained covariance of shocks are for the most part similar to those based upon estimated σ and η and diagonal shocks. In the case of model 1, they indicate slightly lower mean growth rates of .002 and .008 at the industry and economy-wide level. The g_O trends for second ranked model 2, however, are all but 0 in both estimation frameworks and levels of aggregation, with that model's g_O(economy) index of panel (A) graphed in Figure I as well. As shown in panel (C), adjustments of the parameters of panel (A) for bias using the four different bias estimates of Table 8 have, for the most part, little substantive impact on estimated trends, or the lack thereof. The exceptions are models 4, 8 & 12 where, as seen earlier in Tables 7 & 8, both standard errors and bias estimates are large. These models have the largest estimates (in absolute value) of g_O trends in panel (A) and, depending upon which

Table 10: Time Trends of Cointegration Frontiers by Estimation Method & Model

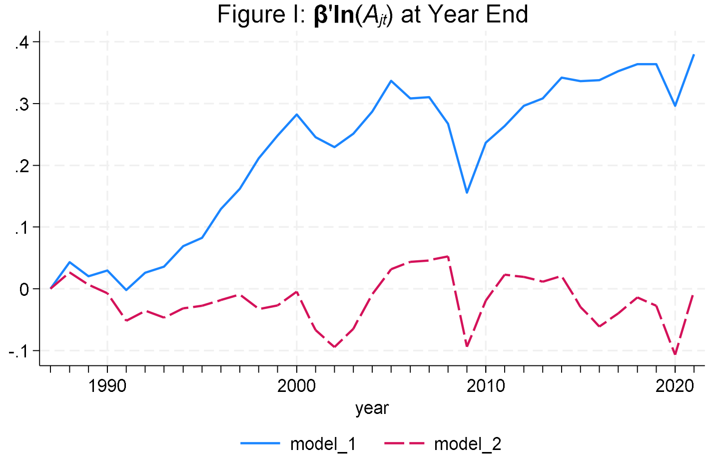
	(A) estimating σ and η (assuming diagonal covariance of shocks)			(B) σ and η known = estimated values (allowing covariance of shocks)		
	g_J	g_O(industry)	g_O(economy)	g_J	g_O(industry)	g_O(economy)
1:	.046 (.001)	.005 (.000)	.011 (.000)	.043 (.000)	.002 (.000)	.008 (.000)
2:	.015 (.003)	-.002 (.000)	-.000 (.000)	.021 (.000)	-.001 (.000)	.001 (.000)
3:	.038 (.007)	.018 (.000)	.016 (.000)	.282 (.001)	.016 (.000)	.017 (.000)
4:	-.222 (.017)	-.002 (.000)	-.010 (.000)	-.197 (.000)	.005 (.000)	-.003 (.000)
5:	.044 (.002)	.012 (.000)	.017 (.000)	.038 (.000)	.008 (.000)	.013 (.000)
6:	-.456 (.025)	-.083 (.003)	-.090 (.003)	-.444 (.003)	-.078 (.001)	-.084 (.001)
7:	.001 (.001)	.002 (.000)	.004 (.000)	.002 (.000)	.004 (.000)	.006 (.000)
8:	-.730 (.058)	.038 (.004)	.004 (.002)	-.944 (.000)	.028 (.000)	-.003 (.000)
9:	.030 (.000)	.009 (.000)	.013 (.000)	.027 (.000)	.007 (.000)	.012 (.000)
10:	.004 (.000)	.001 (.000)	.004 (.000)	.001 (.000)	-.001 (.000)	.002 (.000)
11:	-.225 (.007)	-.020 (.001)	-.023 (.001)	-.088 (.000)	-.013 (.000)	-.017 (.000)
12:	-.127 (.017)	-.040 (.014)	-.048 (.013)	-.065 (.000)	-.006 (.000)	-.010 (.000)

(C) by adjustment of (A) for bootstrap estimate of bias

	diagonal wild		unconstrained wild		diagonal parametric		unconstrained parametric	
	g_O(ind)	g_O(econ)	g_O(ind)	g_O(econ)	g_O(ind)	g_O(econ)	g_O(ind)	g_O(econ)
1:	.004	.010	.004	.009	.004	.010	.005	.011
2:	-.005	-.002	-.004	-.002	-.003	-.001	-.003	-.001
3:	.021	.017	.017	.015	.018	.015	.017	.015
4:	-.016	-.029	.018	.016	-.001	-.009	.003	-.003
5:	.012	.018	.009	.013	.011	.016	.010	.014
6:	-.088	-.095	-.076	-.082	-.088	-.094	-.081	-.088
7:	.003	.004	.002	.004	.002	.004	.001	.003
8:	.010	.008	-.032	.005	.041	.003	.048	.006
9:	.009	.013	.006	.010	.008	.012	.008	.012
10:	.002	.005	.001	.003	.001	.004	-.001	.002
11:	-.024	-.027	-.017	-.020	-.019	-.022	-.012	-.014
12:	-.115	-.135	-.042	-.045	-.032	-.039	-.076	-.084

Notes: Standard errors in () are estimated using the delta method and the wild bootstrap distribution of estimated parameters, retaining the empirical covariance of shocks in panel (B) while imposing a diagonal covariance in panel (A) (standard error estimates retaining the empirical covariance of shocks are the same or smaller). In instances where bias estimates suggest a slightly negative σ or η, they are set equal to 0.

Figure I: **β'ln(*A*$_{jt}$)** at Year End



bias estimate is chosen, such trends can be made to disappear or reverse in sign.

Regarding movements along the frontier, an appeal to cost minimizing behaviour provides a means of identifying these, or at least bounding their contribution. As noted in the introduction, firms will only implement new production techniques that generate linked changes in factor productivities (movements along the frontier) if these lower costs, i.e. raise total factor productivity. Similarly, firms will only implement new production techniques that do not link individual factor productivities (shifts of the frontier) if they generate positive factor augmenting productivity growth for the individual factors. Put differently, barring large sustained macro-level shocks, neither a substantial negative contribution of linked changes in productivity to total factor productivity growth nor substantial negative growth of unlinked individual factor productivities is plausible.

The logic given above suggests the following calculation: given estimated mean factor augmenting growth rates $g(A_K)$, $g(A_L)$, & $g(A_I)$, select movements $g^*(A_K)$, $g^*(A_L)$, & $g^*(A_I)$ along the cointegration frontier that maximize cointegrated *TFP* growth subject to inequality constraints on individual, un-linked, rates of productivity growth, i.e.

$$\text{(V.9)}\quad \text{Max } \theta_K g^*(A_K) + \theta_L g^*(A_L) + \theta_I g^*(A_I)$$
$$\text{subject to: } \beta_K g^*(A_K) + \beta_L g^*(A_L) + \beta_I g^*(A_I) = 0,$$
$$g(A_K) - g^*(A_K) \geq -\delta,\ \ g(A_L) - g^*(A_L) \geq -\delta,\ \ \&\ g(A_I) - g^*(A_I) \geq -\delta,$$

where $-\delta$ is the maximum negative rate of residual un-linked individual factor productivity growth allowed. This calculation provides a logical check on the estimated cointegration parameters $\boldsymbol{\beta}$, since when the maximum is negative, or a large δ is required to make it positive, the tradeoffs implied by the coefficients cannot reasonably be interpreted as profit maximizing. Simultaneously, it provides an upper bound on the maximum contribution of movements along the frontier to aggregate productivity growth and, as will be seen, with additional constraints some sense of the forms that unlinked productivity growth might have taken as well.

The calculation in (V.9) should not and can not be implemented at the industry level. First, the $\boldsymbol{\beta}$ coefficients are calculated off of the entire sample, not attuned to heterogeneity across industries, and hence incapable of producing reasonable results on an industry by industry basis. Estimates of $\boldsymbol{\beta}$ at the industry level are hopelessly inaccurate, as standard error estimates for economy-wide average $\boldsymbol{\beta}$ are already quite large (Table 7). Second, average 1987-2021 TFP growth in 20 of the 61 US KLEMS industries is, in any case, negative,[19] implying that the maximand can only be greater than 0 if $-\delta$ for individual factors is less than 0. For these reasons, instead of solving (V.9) at the industry level I calculate the maximand for the aggregate economy, reporting actions consistent with economy-wide estimates of **β** that at the economy-wide level are on average consistent with cost reducing positive productivity growth.

TFP growth in the aggregate economy is the Domar weighted sum of individual industry gross output productivity growth, which can be re-expressed as:

$$\text{(V.10)} \quad g(TFP_t) = \sum_{i=1}^{N} \frac{P_{it}Q_{it}}{GDP} g(TFP_{it}) = \sum_{i=1}^{N} \frac{P_{it}Q_{it}}{GDP} \sum_{j=1}^{J} \theta_{jit} g(A_{jit}) = \sum_{i=1}^{N} \sum_{j=1}^{J} g(A_{jit}) \frac{P_{jt}X_{jit}}{GDP}$$

$$= \sum_{j=1}^{J} \frac{P_{jt}X_{jt}}{GDP} g(A_{jt}), \text{ where } g(A_{jt}) = \sum_{i=1}^{N} \frac{P_{jt}X_{jit}}{P_{jt}X_{jt}} g(A_{jit})$$

and where X_{jit} & X_{jt} denote industry & economy wide use of factor j, P_{it} output prices and P_{jt} input prices. The economy-wide $g(A_{jt})$, estimates of which are reported earlier in Table 6 & Figure I and used in the estimation of g_O(economy) in Table 9, are the factor income share weighted sum of individual industry factor augmenting growth rates. The Domar factor income shares, $P_{jt}X_{jt}/GDP$, on average sum to about 1.8. Average annual values of both are used as the $g(A_j)$ and θ_j in maximizing (V.9).

Table 11 reports the maximand of (V.9) for different values of the lower bound on un-linked productivity growth $-\delta$. In panel (A) using the baseline estimates where σ and η are identified assuming diagonal shocks, with the non-negativity constraint $-\delta = 0$, the maximum for models 3, 5 & 6 is negative, while that for model 2 is undefined as no values of $g^*(A_j)$ satisfy all four constraints in (V.9). However, with a miniscule -.0001 ln lower bound, cumulatively equivalent to a tiny negative .34 of one percent decline over the 1987-2021 period, model 2's estimates can satisfy the constraints with, in fact, a maximum average TFP growth rate from movements along the frontier equal to .0056, just shy of the .0058 annual TFP growth experienced in the aggregate private sector economy during this period. This is not surprising, as the trends in c_{it} for model 2 are 0 (Table 10), so that all of factor augmenting growth is taking place on a single level curve of $\boldsymbol{\beta}'\mathbf{ln}(A_{jt})$. With a still small -.001 lower bound, all models can satisfy the constraints with positive TFP growth, indicating that the point estimates of cointegration parameters are consistent with profit maximizing behaviour, while with an overgenerous -.01 lowerbound, equivalent to a maximum 34 percent decline over the sample period, maximum TFP growth along the cointegration

---

[19] This may reflect measurement error or real shocks. Baqaee et al (2024), for example, show that supply shocks in the form of upstream firm deaths translate into real cost increases for downstream firms , i.e. meausred productivity losses.

Table 11: Maximum Possible TFP Growth from Movements Along Cointegration Frontier by Lower Bound on Residual Factor Augmenting Growth & Model

	(A) estimating σ and η (diagonal covariance of shocks)				(B) taking σ and η as given (unconstrained covariance of shocks)			
-δ:	0	-.0001	-.001	-.01	0	-.0001	-.001	-.01
1:	.0010	.0012	.0024	.0146	.0022	.0024	.0036	.0165
2:	...	.0056	.0061	.0107	.0053	.0054	.0060	.0117
3:	-.0004	-.0002	.0014	.0174	-.0014	-.0010	.0024	.0208
4:	.0052	.0053	.0063	.0171	.0056	.0058	.0069	.0184
5:	-.0004	-.0003	.0009	.0124	.0000	.0001	.0013	.0133
6:	-.0004	-.0002	.0012	.0160	-.0001	.0000	.0014	.0156
7:	.0044	.0045	.0051	.0114	.0038	.0038	.0044	.0097
8:	.0046	.0048	.0072	.0204	.0050	.0053	.0072	.0187
9:	.0003	.0004	.0016	.0132	.0008	.0010	.0023	.0156
10:	.0045	.0046	.0057	.0168	.0051	.0053	.0065	.0188
11:	.0001	.0002	.0009	.0087	-.0013	-.0011	.0002	.0137
12:	.0037	.0038	.0048	.0147	.0044	.0045	.0056	.0167

(C) by adjustment of (A) for boostrap estimate of bias

	diagonal wild		unconstrained wild		diagonal parametric		unconstrained parametric	
-δ:	0	-.001	0	-.001	0	-.001	0	-.001
1:	.0012	.0025	.0017	.0031	.0012	.0026	.0011	.0024
2:	...	.0065	...	.0068	...	.0063	...	.0063
3:	-.0008	.0009	-.0002	.0015	-.0004	.0013	-.0001	.0015
4:	.0047	.0058	.0003	.0014	.0052	.0064	.0056	.0068
5:	-.0004	.0008	.0002	.0014	-.0004	.0009	-.0003	.0010
6:	-.0003	.0013	-.0002	.0014	-.0004	.0012	-.0004	.0012
7:	.0041	.0048	.0045	.0052	.0044	.0051	.0047	.0055
8:	.0032	.0059	.0050	.0072	.0047	.0073	.0045	.0072
9:	.0003	.0016	.0014	.0027	.0006	.0018	.0006	.0019
10:	.0041	.0053	.0046	.0058	.0045	.0057	.0051	.0064
11:	.0001	.0009	.0004	.0012	.0002	.0010	.0004	.0013
12:	.0033	.0044	.0037	.0048	.0038	.0049	.0038	.0048

Notes: Standard error estimates, based upon the delta method using the wild bootstrap covariance matrix of coefficient estimates, in panel (B) are all .0000 and in panel(A) are mostly .0000, with a few .0001 or .0002, using the wild bootstrap based upon diagonal covariance or unrestricted covariance. In instances where bias estimates suggest a slightly negative σ or η, they are set equal to 0.

frontier in all models is well above the aggregate for the private sector economy. Stress tests of the headline results in Table 11 have very modest effects on results. Panel (B), which takes σ and η as equal to the values in panel (A) but otherwise uses an unrestricted covariance matrix of shocks in estimating the $\boldsymbol{\beta}$ parameters, by and large finds similar patterns across models. The most notable change is that the maximum contribution of movements along the frontier to TFP growth in model 1 with -δ = 0 rises to .022, or a little over ⅓ of aggregate TFP growth. Adjustments in panel (C) of the parameter estimates in panel (A) using various estimates of bias have

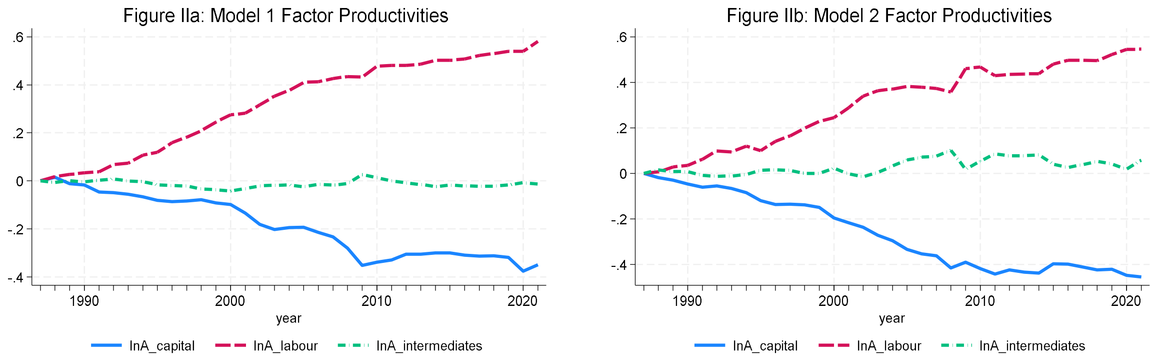
Table 12: Model 1 1987-2021 $g^*(TFP)$ Maximizing Movements Along Cointegration Frontier by Type of Un-linked Factor Augmenting Productivity Growth

	movements along frontier				un-linked productivity growth			
	$g^*(A_K)$	$g^*(A_L)$	$g^*(A_I)$	$g^*(TFP)$	$g(A_K)$ - $g^*(A_K)$	$g(A_L)$ - $g^*(A_L)$	$g(A_I)$ - $g^*(A_I)$	$g(TFP)$ - $g^*(TFP)$
Solow neutral	-.0212	.0169	-.0004	.0010	.0112	-.0000	-.0000	.0047
Harrod neutral	-.0101	.0079	-.0004	.0002	-.0000	.0091	-.0000	.0056
Hicks neutral	-.0151	.0119	-.0004	.0006	.0050	.0050	-.0000	.0052
all factor neutral	-.0191	.0079	-.0094	-.0112	.0090	.0090	.0090	.0170

Notes: g^*(TFP) & $g(TFP)$ are calculated by multiplying the annual factor augmenting growth rates by annual Domar factor weights $P_{jt}X_{jt}/GDP_t$ which sum to 1 for capital and labour but 1.8 including intermediates. $g^*(A_j)$ is set at the average level for all years. Because $g(A_j)$-$g^*(A_j)$ is larger in years when θ_j is larger, the contribution of Hicks neutral productivity growth to TFP growth is greater than the average rate of Hicks neutral productivity growth.

virtually no effect outside of model 4, where the contribution can be made equal to virtually all or none of aggregate TFP growth, depending upon the adjustment. For other models, and most notably models 6, 8, 11 & 12 where bias estimates were large (Table 8), the effect is substantively minimal, as models consistently indicate either a maximand equal to most of or only a small fraction of aggregate TFP growth. The maximand in (V.9) depends on constraints determined by the σ,η, & $\boldsymbol{\beta}$ parameters together and outside of model 4 adjustments for bias largely cancel out.

Table 12 concludes by zeroing in on model 1, with the highest likelihood, showing the patterns of factor augmenting productivity growth consistent with positive productivity growth from linked cointegrated movements and greater than or equal to 0 ($-\delta$=0) residual non-linked productivity growth. In the top line of the table we find that the maximand of (V.9) is actually one with Solow neutral unlinked factor augmenting growth. The second line of the table solves (V.9) with the additional restriction that un-linked productivity growth is Harrod neutral, i.e. solely labour augmenting. The maximand is still positive, but truly minimal, showing only .02 of one percent annual productivity gains from cointegrated changes. The third line considers a Hicks neutral restriction, with equal rates of capital and labour unlinked productivity growth, and finds the maximand is positive and equal to .0006 per annum. Finally, the fourth line imposes the constraint that all three unlinked productivity growth rates be equal and finds the maximand, at -.0112, to be negative and roughly twice the .058 magnitude of aggregate realized productivity growth. This possibility is clearly ruled out. These calculations, and those in Table 11, are by construction all consistent with the actual outward shift of $\boldsymbol{\beta}'\mathbf{ln}(A_{jt})$, as (V.9) imposes the constraint $\beta_K g^*(A_K) + \beta_L g^*(A_L) + \beta_I g^*(A_I) = 0$. Consequently, similar Table 12 calculations for model 2 are pointless as the rates of unlinked productivity growth in



all scenarios are negligible, with movements along the frontier accounting for almost all of TFP growth.

To summarize, model 1 finds that the majority of gross (positive or negative) movements of factor productivity growth are associated with movements along the cointegration frontier, but the net gain from these was modest, accounting for at most ⅙, or when stress tested perhaps as much as ⅓, of aggregate productivity growth, with unlinked productivity growth accounting for the remainder and generating a 1.1 percent annual outward shift (in units of capital productivity) of the cointegration frontier. Second ranked model 2, however, finds that movements along the frontier essentially accounted for all of both factor augmenting productivity and TFP growth.

VI. Conclusion

Figure II graphs the cumulative ln factor augmenting productivity indices for the VAR t-distribution versions of top ranked models 1 and 2 earlier above. These indices show a secular decline in capital productivity since 1987 that is only halted or slowed by the financial crisis of 2008. Through the lens of early growth models (e.g. Solow 1962), this decline is a sign of growing investment inefficiency, an inefficiency stopped by post-financial crisis capital stringency. Given the length (2 decades) and depth (40 percent) of the decline, this interpretation seems highly dubious.

An alternative explanation of Figure II is that the decline in capital productivity is intrinsically linked to the rise in labour productivity in the same figure, both being part and parcel of a change in production relationships which, when aggregated together, improves total factor productivity and hence is implemented by cost minimizing firms. Statistically, this is confirmed by the rejection of the VAR model in favour of a rank 1 cointegration VEC framework. Theoretically, a model of task reallocation can motivate these linkages and rationalize the observed trends as depicting the growing allocation of tasks to capital at the expense of those performed by labour since, with elasticities of substitution less than 1, a factor's productivity declines (increases) as it is spread across more (fewer) tasks. Within this framework, the post-2008 flattening of *both* the capital and labour productivity curves is a sign of the enduring negative impact of the financial crisis, which has slowed down the rate at which firms are making productivity and cost-improving tradeoffs in production relationships.

The implications of cointegration for the modelling and study of productivity potentially go much deeper than the simple linking of productivities. Cointegration binds the non-stationary elements of ln factor productivity to a stationary, possibly trend stationary, variable, $\boldsymbol{\beta}'\mathbf{ln}(A_{jit}) = c_{it}$. Thus, while the forces moving the linked values of $\mathbf{ln}(A_{jit})$ are non-stationary, generating non-stationary TFP and factor input ratios, the forces changing the linking restriction are stationary. In the analysis above, model 2 shows no trend in $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$, and movements along the cointegration frontier account for all of productivity growth & hence, if one will allow, all technical change. However, in the case of model 1, with the highest likelihood, $\boldsymbol{\beta}'\mathbf{ln}(A_{jit})$ is trending upward at about 1 percent per annum, with movements along the cointegration frontier accounting for at most ⅙ of economy-wide total factor productivity growth, and consequently a strong indication of positive trends in c_{it} and non-cointegration linked productivity growth. This suggests the existence of two processes of innovation with different statistical properties, one based upon tradeoffs between factor productivities with unit roots and another which relaxes those tradeoffs and is trend stationary. Models such as task based innovation which motivate observed constraints between the elements of $\mathbf{ln}(A_{jit})$ that identify cointegration coefficients in the data, must in turn be informed or restricted by the finding of cointegration, which suggests that any forces relaxing those constraints have quite different stochastic properties.

Bibliography

Acemoglu, Daron and David Autor (2011). "Skills, tasks and technologies: Implications for employment and earnings." In Orley Ashenfelter and David Card, eds, Handbook of Labor Economics, Vol. 4b. Netherlands: North-Holland.

Acemoglu, Daron and Pascual Restrepo (2018). "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American Economic Review* 108: 1488-1542.

Acemoglu, Daron and Pascual Restrepo (2020). "Robots and jobs: Evidence from US labor markets." *Journal of Political Economy* 128: 2188-2244.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones (2019). "Artificial intelligence and economic growth." In Agrawal, Gans and Goldfarb, eds, The

Economics of Artificial Intelligence: An Agenda. Chicago: University of Chicago Press.

Autor, David H., Frank Levy and Richard J. Murnane (2003). "The skill content of recent technological change: An empirical exploration." *Quarterly Journal of Economics* 116: 1279-1333.

Autor, David H. and David Dorn (2013). "The growth of low-skill service jobs and the polarization of the US labor market." *American Economic Review* 103: 1553-1597.

Baqaee, David, Ariel Burstein, Cédric Duprez and Emmanuel Farhi (2024). "Consumer Surplus from Suppliers: How Bit is it and Does it Matter for Growth?" Working paper.

Basu, Susanto, John G. Fernald and Miles S. Kimball (2006). "Are technology improvements contractionary?" *American Economic Review* 96: 1418-1448.

Bils, Mark, Baris Kaymak and Kai-Jie Wu (forthcoming). "Labor Substitutability among Schooling Groups." *American Economic Journal: Macroeconomics*.

Byrne, David M., John G. Fernald and Marshall B. Reinsdorf (2016). "Does the United States have a productivity slowdown or a measurement problem?" *Brookings Papers on Economic Activity*, Spring 2016: 109-182.

Diamond, Peter, Daniel McFadden and Miguel Rodriguez (1978). "Measurement of the elasticity of factor substitution and bias of technical change." In M. Fuss and Daniel McFadden (editors), Production Economics 2. Amsterdam: North-Holland.

Dickey, D. A., and W. A. Fuller (1979). "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association* 74: 427–431.

Duffy, John, Chris Papageorgiou, and Fidel Perez-Sebastian (2004). "Capital skill complementarity? Evidence from a panel of countries." *Review of Economics and Statistics* 86: 327-344.

Gechert, Sebastian, Tomas Havranek, Zuzana Irsova, and Dominika Kolcunova (2022). "Measuring capital-labor substitution: The importance of method choices and publication bias." *Review of Economic Dynamics* 45: 55-82.

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

Hadri, K. (2000). "Testing for stationarity in heterogeneous panel data." *Econometrics Journal* 3: 148–161.

Harris, R. D. F., and E. Tzavalis (1999). "Inference for unit roots in dynamic panels where the time dimension is fixed." *Journal of Econometrics* 91: 201–226.

Harrod, Henry Roy Forbes (1948). Towards a Dynamic Economics. London: Macmillan.

Hicks, John Richard (1932). The Theory of Wages. London: Macmillan.

Hsieh, Chang-tai, and Peter J. Klenow (2009). "Misallocation and manufacturing TFP in China and India." *Quarterly Journal of Economics* 124: 1403-1448.

Hsieh, Chang-tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow (2019). "The allocation of talent and U.S. economic growth." *Econometrica* 87: 1439-1474.

Im, K. S., M. H. Pesaran, and Y. Shin (2003). "Testing for unit roots in heterogeneous panels." *Journal of Econometrics* 115: 53–74.

Jockel, Karl-Heinz (1986). "Finite Sample Properties and Asymptotic Efficiency of Monte Carlo Tests." *The Annals of Statistics* 14: 336-347

Johansen, Soren (1995). Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press.

Kaldor, Nicholas (1961). "Chapter 10: Capital Accumulation and Economic Growth". In Lutz, Friedrich; Hague, Douglas (eds.). *Capital Accumulation and Economic Growth*. MacMillan and Co. Ltd. pp. 177–222

Karabarbounis, Loukas and Brent Neiman (2013). "The global decline of the labor share." *Quarterly Journal of Economics* 129: 61-103.

Klump, Rainer, Peter McAdam and Alpo Willman (2007). "Factor substitution and factor-augmenting technical progress in the United States: A normalized supply-side system approach." *Review of Economics and Statistics* 89: 183-192.

Klump, Rainer, Peter McAdam and Alpo Willman (2012). "The normalized CES production function: Theory and empirics." *Journal of Economic Surveys* 26: 769-799.

Knoblach, Michael, Martin Roessler and Patrick Zwerschke (2020). "The elasticity of substitution between capital and labour in the US economy: A meta-regression analysis." *Oxford Bulletin of Economics and Statistics* 82: 62-82.

Koh, Dongya, Raül Santaeulàlia-Llopis, and Yu Zheng (2020). "Labor share decline and intellectual property products capital." *Econometrica* 88: 2609-2628.

León-Ledesma, M.A., P. McAdam and A. Willman (2010). "Identifying the elasticity of substitution with biased technical change." *American Economic Review* 100: 1330-1357.

León-Ledesma, M.A., P. McAdam and A. Willman (2015). "Production technology estimates and balanced growth." *Oxford Bulletin of Economics and Statistics* 77: 40-65.

Muck, Jakub (2017). "Elasticity of substitition between labor and capital: robust evidence from developed economies." Narodowy Bank Polski, working paper 271.

Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49: 1417-1426.

Sato, Kazuo (1977). "A note on factor substitution and efficiency." *Review of Economics and Statistics* 59: 360-366.

Solow, Robert M (1962). "Technical progress, capital formation, and economic growth." *American Economic Review* 52: 76-86.

Syverson, Chad (2017). "Challenges to mismeasurement explanations for the US productivity slowdown." *Journal of Economic Perspectives* 31: 165-186

Spitz-Oener, Alexandra (2006). "Technical change, job tasks, and rising educational demands: Looking outside the wage structure." *Journal of Labor Economics* 24: 235-270.

Stehrer, Robert, Alexandra Bykova, Kirsten Jäger, Oliver Reiter and Monika Schwarzhappel (2019). "Industry Level Growth and Productivity Data with Special Focus on Intangible Assets." Working paper, Vienna Institute for International Economic Studies, 2019.

Vuong, Quang H. (1989). "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses." *Econometrica* 57: 307-333.

Young, Alwyn (2019). On-Line Appendix I for "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics* 134: 557-598.

Zeira, Joseph (1998). "Workers, Machines and Economic Growth." *Quarterly Journal of Economics* 113: 1091-1117.