Appendix B from Acemoglu and Finkelstein, "Input and Technology Choices in Regulated Industries: Evidence from the Health Care Sector"

(JPE, vol. 116, no. 5, p. 837)

A. A Neoclassical Model of Regulation

In this appendix, we present the details of the neoclassical model of regulation described in the text.

Our approach is based on four simplifying assumptions. The first is that hospitals maximize profits. Clearly, nonprofit or public hospitals have other objectives as well, but starting with the profit-maximizing case is a useful benchmark. It is also consistent with a large empirical literature that finds essentially no evidence of differential behavior across for-profit and nonprofit hospitals (see Sloan [2000] for a recent review of this literature). Second, we assume that hospitals are price takers in the input markets, facing a wage rate of w per unit of labor and a cost of capital equal to R per unit of capital. Third, we assume that hospitals are price takers for Medicare patients. Finally, and to start with, we assume that, at least at the margin, there is considerable fungibility between labor and capital inputs used for Medicare purposes and labor and capital inputs used for non-Medicare purposes; descriptions of how Medicare reimbursement operates in practice suggest that this is a realistic assumption (OTA 1984; CBO 1988). This allows us to model Medicare input reimbursement as taking a simple form in which hospital i is reimbursed for a fraction m_i of its capital and labor costs, where m_i is the "Medicare share" of this hospital. Section B in this appendix extends the framework to investigate the implications of the impact of a change in regulation regime without fungibility. Our analysis there shows that the major qualitative predictions highlighted in the model with fungibility also hold when there is limited fungibility.

1. *Environment*

Suppose that hospital i has a production function for total health services given by

$$\tilde{F}(A_i, L_i, K_i, z_i), \tag{B1}$$

where L_i and K_i are total labor and capital hired by this hospital, z_i is some other input, such as managerial effort (or doctors, who are not directly hired and paid by hospitals themselves), and A_i is a productivity term, which may differ across hospitals, for example, because of their technology choices or for other reasons. We assume that $\tilde{F}$ is increasing in all its inputs and twice continuously differentiable for positive levels of inputs.

For simplicity, we will interpret (B1) as the production function of the hospital, though, equivalently, it could be interpreted as its revenue function (with the price substituted in as a function of quantity). We also assume that z_i is fixed, and, without loss of any generality, we normalize it to $z_i = 1$ and begin with the case in which A_i is exogenous. This gives

$$F(A_i, L_i, K_i) \equiv \tilde{F}(A_i, L_i, K_i, z_i = 1), \tag{B2}$$

which we assume exhibits decreasing returns to scale in capital and labor (e.g., because the original production function $\tilde{F}$ exhibited constant returns to scale). Since $\tilde{F}$ is increasing in its inputs and twice continuously differentiable for positive inputs, so is F, and we denote the partial derivatives by F_L and F_K (and the second derivatives by F_{LL}, F_{KK}, and F_{LK}). Moreover, we make the standard Inada-type assumption that

$$\lim_{L_i \to 0} F_L(A_i, L_i, K_i) = \lim_{K_i \to 0} F_K(A_i, L_i, K_i) = \infty$$

and

$$\lim_{L_i \to \infty} F_K(A_i, L_i, K_i) = \lim_{K_i \to \infty} F_K(A_i, L_i, K_i) = 0.$$

In addition, we will often look at the cases in which $F(A_i, L_i, K_i)$ is homothetic or homogeneous in L_i and K_i or in A_i and L_i.[34]

2. Full Cost Reimbursement Regulation

Under the original regulation, which we refer to as *full cost reimbursement*, each hospital receives reimbursement for some fraction of its labor and capital used for Medicare purposes.[35] It also receives a co-payment from Medicare patients as well as revenues from non-Medicare patients (where the hospital might have some market power, which we are incorporating into the F function). With the total price per unit of health care services under the cost reimbursement regulation system denoted by $q > 0$, the maximization problem of the hospital is

$$\max_{L_i, K_i} \pi^f(i) = qF(A_i, L_i, K_i) - (1 - m_i s_L) w L_i - (1 - m_i s_K) R K_i, \tag{B3}$$

where $s_L < 1$ and $s_K < 1$ are constants capturing the relative generosity of labor and capital Medicare reimbursement and $m_i \in [0, 1]$ is the Medicare share of the hospital, which we take as given for now and endogenize in subsection 6.[36]

The first-order conditions of this maximization problem are

$$qF_L(A_i, L_i^f, K_i^f) = (1 - m_i s_L) w \tag{B4}$$

and

$$qF_K(A_i, L_i^f, K_i^f) = (1 - m_i s_K) R \tag{B5}$$

for labor and capital, respectively, where the superscript f refers to full cost reimbursement.

The Inada and the differentiability assumptions imply that these first-order conditions are necessary, and the decreasing returns (strict joint concavity) of F imply that they are sufficient. Taking the ratio of these two first-order conditions, we have

$$\frac{F_K(A_i, L_i^f, K_i^f)}{F_L(A_i, L_i^f, K_i^f)} = \frac{(1 - m_i s_K) R}{(1 - m_i s_L) w}, \tag{B6}$$

which shows that the relative input choices of the hospital will be similar to that of an unregulated firm (hospital) with the same production technology, except for the relative generosity of capital and labor reimbursements. Equation (B6) combined with the decreasing returns assumption on F implies that an increase in s_K/s_L, which corresponds to capital reimbursements becoming more generous relative to labor reimbursements, will increase K_i/L_i. Similarly, a decrease in the relative price of capital, R/w, will increase K_i/L_i. The impact of changes in m_i on K_i/L_i will depend on whether s_K is greater or less than s_L. In the former case, capital is favored relative to labor, so higher m_i will be associated with greater capital intensity.

[34] If $F(A_i, L_i, K_i)$ is homothetic in L_i and K_i, then $F_K(A_i, L_i, K_i)/F_L(A_i, L_i, K_i)$ is a function only of K_i/L_i. Alternatively, homotheticity in L_i and K_i is equivalent to $F(A_i, L_i, K_i) \equiv H_1(A_i)H_2(\phi(L_i, K_i))$, where $H_1(\cdot)$ and $H_2(\cdot)$ are increasing functions, and ϕ is increasing in both of its arguments and exhibits constant returns to scale. If $F(A_i, L_i, K_i)$ is homogeneous of degree α in L_i and K_i, then $F_K(A_i, L_i, K_i)/F_L(A_i, L_i, K_i)$ is again a function only of K_i/L_i; in addition, $F(A_i, L_i, K_i) \equiv H_1(A_i)\phi(L_i, K_i)^\alpha$, where ϕ is increasing in both of its arguments and exhibits constant returns to scale.

[35] As discussed in the text, under the pre-PPS system, Medicare-related capital and labor expenses were reimbursed in proportion to Medicare's share of patient days or charges (see Newhouse 2002, 22).

[36] The assumption that $s_L < 1$ and $s_K < 1$ ensures that, at the margin, labor and capital costs are always positive for the hospital. In fact, all we need is that $m_i s_L < 1$ and $m_i s_K < 1$. So in practice, when $m_i \leq \bar{m}$ for some $\bar{m} < 1$, we can have $s_L > 1$ and $s_K > 1$. The case in which there is true cost plus reimbursement whereby the hospital makes money by hiring more inputs is discussed in the next section.

3. Partial Cost Reimbursement Regulation

Our main interest is to compare the full cost reimbursement regulation regime described above, which is a stylized description of the regulation policy before PPS, to the partial cost reimbursement that came with PPS. As described above, under this new regime, capital continues to be reimbursed as before, but labor reimbursements cease, and instead, hospitals receive additional payments from Medicare for health services provided to Medicare patients. We model this as an increase in q to $(1+\theta m_i)q$, where $\theta > 1$ incorporates the fact that the extent to which a hospital receives the subsidy is also a function of its Medicare share.[37]

Now the maximization problem of hospital i is

$$\max_{L_i,K_i} \pi^p(i) = (1+\theta m_i)qF(A_i, L_i, K_i) - wL_i - (1-m_i s_K)RK_i. \tag{B7}$$

The first-order necessary and sufficient conditions are

$$(1+\theta m_i)qF_L(A_i, L_i^p, K_i^p) = w \tag{B8}$$

and

$$(1+\theta m_i)qF_K(A_i, L_i^p, K_i^p) = (1-m_i s_K)R, \tag{B9}$$

where the superscript p refers to partial cost reimbursement. Equations (B8) and (B9) jointly imply

$$\frac{F_K(A_i, L_i^p, K_i^p)}{F_L(A_i, L_i^p, K_i^p)} = \frac{(1-m_i s_K)R}{w}. \tag{B10}$$

Comparison of (B10) to (B6) immediately yields the following result.

Proposition 1. Suppose that $F(A_i, L_i, K_i)$ is homothetic in L_i and K_i. Then, the move from full cost reimbursement to partial cost reimbursement increases the capital-labor ratio, that is,

$$\frac{K_i^p}{L_i^p} > \frac{K_i^f}{L_i^f}. \tag{B11}$$

Moreover, this effect is stronger for hospitals with a greater Medicare share, that is,

$$\partial\left(\frac{K_i^p/L_i^p}{K_i^f/L_i^f}\right)\Big/\partial m_i > 0. \tag{B12}$$

Proof. Taking the ratio of (B10) to (B6), we obtain

$$\frac{F_K(A_i, L_i^p, K_i^p)/F_L(A_i, L_i^p, K_i^p)}{F_K(A_i, L_i^f, K_i^f)/F_L(A_i, L_i^f, K_i^f)} = 1 - m_i s_L.$$

When $F(A_i, L_i, K_i)$ is homothetic in L_i and K_i, the left-hand side is a decreasing function of $(K_i^p/L_i^p)/(K_i^f/L_i^f)$. This immediately establishes (B11), since $1 - m_i s_L < 1$, and (B12), since $1 - m_i s_L$ is decreasing in m_i. QED

This proposition is the starting point for our empirical work in the text. It shows that the move from full to partial cost reimbursement should be associated with an increase in capital-labor ratios. Moreover, equation (B12) provides an empirical strategy to investigate this effect by comparing hospitals with different Medicare shares (from the prereform period).

Next, we would like to know the impact of the change in regulation regime on the level of inputs and the total amount of health services. It is clear that the results here will depend on the generosity of the price subsidy

[37] In practice, the price subsidy under PPS is a function of Medicare (diagnosis-adjusted) admissions. Modeling it as a function of the Medicare share, m_i—which corresponds roughly to Medicare share of total output (see subsec. 6 below)—is a simplifying assumption, with no major effect on our theoretical results.

(price cap) $\theta > 0$. We can obtain more insights by focusing on the case in which the price cap, θ, is sufficiently low. As discussed in the text, this case is consistent with the existing work on PPS.

Let us consider the extreme case with $\theta = 0$ (clearly, by continuity, the same results apply when θ is sufficiently small around zero). In this case, we can analyze the effect of the change in the cost reimbursement regime as comparative statics of s_L; a reduction in s_L from positive to zero is equivalent to a change in regulation regime from full cost reimbursement to partial cost reimbursement.

Proposition 2. Suppose that $\theta = 0$, and let $L_i(s_L)$ and $K_i(s_L)$ be the optimal choices for hospital i at labor subsidy rate s_L. Then

$$\frac{dL_i(s_L)}{ds_L} = \frac{-m_i F_{KK}}{F_{LL}F_{KK} - (F_{LK})^2} > 0. \tag{B13}$$

Moreover, let $F(A_i, L_i, K_i)$ be homogeneous of degree $\alpha < 1$ in L_i and K_i; that is, $F(A_i, L_i, K_i) = H_1(A_i)\phi(L_i, K_i)^{\alpha}$, with $\phi(\cdot, \cdot)$ exhibiting constant returns to scale. Let the (local) elasticity of substitution between capital and labor of the $\phi(\cdot, \cdot)$ function be σ_ϕ. Then

$$\frac{dK_i(s_L)}{ds_L} \gtreqless 0 \quad \text{if and only if} \quad \frac{1}{1-\alpha} \gtreqless \sigma_\phi. \tag{B14}$$

Proof. To prove this proposition, totally differentiate the first-order conditions (B4) and (B5) with respect to L_i, K_i, and s_L, and write the resulting system as

$$\begin{pmatrix} F_{LL} & F_{LK} \\ F_{LK} & F_{KK} \end{pmatrix} \begin{pmatrix} dL \\ dK \end{pmatrix} = \begin{pmatrix} -m_i \\ 0 \end{pmatrix} ds_L.$$

Applying Cramer's rule immediately gives (B13), and the fact that $F_{LL}F_{KK} - (F_{LK})^2 > 0$ and $F_{KK} < 0$ follows from the concavity of F, thus establishing the fact that $dL_i(s_L)/ds_L > 0$ as stated in (B13). Similarly, from Cramer's rule,

$$\frac{dK_i(s_L)}{ds_L} = \frac{m_i F_{LK}}{F_{LL}F_{KK} - (F_{LK})^2}. \tag{B15}$$

Therefore, this will be positive when $F_{LK} > 0$ and negative when $F_{LK} < 0$. When F is homogeneous of degree α, that is, $F(A_i, L_i, K_i) = H_1(A_i)\phi(L_i, K_i)^{\alpha}$, it is easy to verify that

$$F_{LK} \propto (\alpha - 1)\phi_L\phi_K + \phi_{LK}\phi.$$

Recall that when ϕ exhibits constant returns to scale, the elasticity of substitution is given by

$$\sigma_\phi \equiv \frac{\phi_L\phi_K}{\phi_{LK}\phi}.$$

This implies that $F_{LK} < 0$ if and only if $1/(1-\alpha) < \sigma_\phi$ and is positive if and only if $1/(1-\alpha) > \sigma_\phi$, thus establishing (B14). QED

This proposition shows that when the price cap is not very generous, the firm will respond to the switch from full to partial cost reimbursement by reducing its labor input, that is, $dL_i(s_L)/ds_L > 0$.

The noteworthy result is that even in this case, capital inputs may increase; that is, $dK_i(s_L)/ds_L \leq 0$ is possible. Whether they do so or not depends on the amount of "decreasing returns" to labor and capital, which is measured by the α parameter and the elasticity of substitution, σ_ϕ. If $\sigma_\phi < 1$, so that labor and capital are gross complements in the ϕ function, capital will always decline as well. Similarly, if $\alpha = 1$, so that there are constant returns to scale to capital and labor jointly, again, capital will always decline.[38] However, if $\alpha < 1$ and there is

[38] This is obvious in proposition 2 because of constant returns to scale, i.e., $\alpha = 1$. Alternatively, with constant returns to scale in labor and capital, the Euler theorem implies that $F_{LK} > 0$, so (B14) immediately yields $dK_i(s_L)/ds_L > 0$.

sufficient substitution between labor and capital, that is, $\sigma_\phi > 1$, the firm can (partially) make up for the decline in its labor demand by increasing its capital inputs.

4. Technology Choices

The overall amount of capital inputs used by the hospital is a combination of capital embodying new technologies and other types of capital, such as structures (e.g., buildings). These different types of capital may respond differentially to the change in regulation. To study how technology will respond to the regulation regime, we now model technology choices.

Suppose that technology is always embodied in capital, and it can be measured by a real number, that is, $A_i \in \mathbb{R}_+$, as specified by the production functions in (B1) or (B2). In particular, let us posit that there is a large number of (perfectly substitutable) technologies, each indexed by $x \in [0, \infty)$. Technology x requires a capital outlay of $\kappa(x)$.[39] We rank technologies such that $\kappa(x)$ is increasing. Furthermore, to simplify the analysis, let us assume that $\kappa(\cdot)$ is continuously differentiable. Since the productivity of the hospital depends only on how many of these technologies are adopted, that is, only on A_i, it will adopt low-x technologies before high-x technologies; that is, there will exist a cutoff level x_i^* such that hospital i adopts all technologies $x \leq x_i^*$ and, moreover, clearly $x_i^* \equiv A_i$. Hence the capital cost of technology for hospital i when it adopts technology A_i is

$$K_{a,i} \equiv \int_0^{A_i} \kappa(x)dx, \tag{B16}$$

which is in addition to its capital costs for structures. Note from (B16) that the marginal cost of adopting technology A_i is $\kappa(A_i)$; moreover, since $\kappa(x)$ is increasing, this marginal cost is increasing in A_i. Other differences in productivity across hospitals are ignored for simplicity.

Since we now allow for the adoption of new technologies embodied in capital, the remaining capital is interpreted as “structures” capital and denoted by $K_{s,i}$. Hence, we write

$$F(A_i, L_i, K_{s,i}) = \psi(A_i, L_i)^\beta K_{s,i}^\eta, \tag{B17}$$

where $\eta \in [0, 1 - \beta)$ and ψ exhibits constant returns to scale, which imposes homogeneity of degree $\beta < 1$ between A_i and L_i. The rest of the setup is unchanged.

Once again, since, for arbitrary θ’s, total output (health services) and inputs can increase or decrease, we focus on the case of $\theta = 0$. We have the following proposition.

Proposition 3. Suppose that $\theta = 0$ and the production function is given by (B17), with $\psi(\cdot, \cdot)$ exhibiting constant returns to scale. Let $L_i(s_L)$, $A_i(s_L)$, $K_{a,i}(s_L)$, and $K_{s,i}(s_L)$ be the optimal choices for hospital i at labor subsidy rate s_L. Let ε_ψ be the (local) elasticity of substitution between L_i and A_i in the function $\psi(\cdot, \cdot)$. Then we have

$$\frac{dL_i(s_L)}{ds_L} > 0 \quad \text{and} \quad \frac{dK_{s,i}(s_L)}{ds_L} > 0, \tag{B18}$$

and

$$\frac{\partial K_{a,i}(s_L)}{\partial s_L} \gtreqless 0 \quad \text{and} \quad \frac{\partial A_i(s_L)}{\partial s_L} \gtreqless 0 \quad \text{if and only if} \quad \frac{1-\eta}{1-\beta-\eta} \gtreqless \varepsilon_\psi. \tag{B19}$$

Proof. From the form in (B17), the first-order necessary and sufficient conditions (under full cost reimbursement) are

[39] In practice, new technologies may differ in their productivity and may also require both capital and labor inputs for their adoption and operation. In the latter case, changes in the relative prices of capital and labor will also affect which technologies are more likely to be adopted. We do not model these issues explicitly both to simplify the analysis and also because we cannot measure the relative capital intensity of technologies in our empirical work.

$$q\beta\psi_L(A_i, L_i)\psi(A_i, L_i)^{\beta-1}H_s(K_{s,i}) = (1 - m_i s_L)w,$$
$$q\beta\psi_A(A_i, L_i)\psi(A_i, L_i)^{\beta-1}H_s(K_{s,i}) = (1 - m_i s_K)R\kappa_i(A_i),$$
$$q\psi(A_i, L_i)^{\beta}H_s'(K_{s,i}) = (1 - m_i s_K)R.$$

Taking logs and totally differentiating with respect to A_i, L_i, $K_{s,i}$, and s_L, we obtain the following system of equations:

$$\begin{pmatrix} \frac{\psi_{LL}(A_i, L_i)}{\psi_L(A_i, L_i)} - (1-\beta)\frac{\psi_L(A_i, L_i)}{\psi(A_i, L_i)} & \frac{\psi_{AL}(A_i, L_i)}{\psi_L(A_i, L_i)} - (1-\beta)\frac{\psi_A(A_i, L_i)}{\psi(A_i, L_i)} & \frac{\eta}{K_{s,i}} \\ \frac{\psi_{AL}(A_i, L_i)}{\psi_A(A_i, L_i)} - (1-\beta)\frac{\psi_L(A_i, L_i)}{\psi(A_i, L_i)} & -\frac{\kappa_i'(A_i)}{\kappa_i(A_i)}\frac{\psi_{AA}(A_i, L_i)}{\psi_A(A_i, L_i)} - (1-\beta)\frac{\psi_A(A_i, L_i)}{\psi(A_i, L_i)} & \frac{\eta}{K_{s,i}} \\ \beta\frac{\psi_L(A_i, L_i)}{\psi(A_i, L_i)} & \beta\frac{\psi_A(A_i, L_i)}{\psi(A_i, L_i)} & \frac{\eta - 1}{K_{s,i}} \end{pmatrix} \begin{pmatrix} dL_i \\ dA_i \\ dK_{s,i} \end{pmatrix} = \begin{pmatrix} \frac{-m_i}{1 - m_i s_L} \\ 0 \\ 0 \end{pmatrix} ds_L.$$

Applying Cramer's rule again and using the fact that (B17) is strictly concave, we immediately obtain $dL_i(s_L)/ds_L > 0$ and $dK_{s,i}(s_L)/ds_L > 0$ and that $dA_i(s_L)/ds_L$ is proportional to

$$(1 - \eta)\psi_{AL}\psi + (1 - \beta)(\eta - 1)\psi_A\psi_L + \beta\eta\psi_A\psi_L.$$

Again using the definition of the elasticity of substitution with constant returns to scale, that is, $\varepsilon_\psi \equiv \psi_A\psi_L/\psi_{AL}\psi$, and the fact that $K_{a,i}$ is a monotonic transformation of A_i yields (B19). QED

This proposition generalizes proposition 2 to an environment with labor, capital, and technology choices and is the starting point of our empirical analysis of technology choices. It indicates that the same kind of comparison between the elasticity of substitution and returns to scale also guides whether or not technology adoption will be encouraged by the change in the regulation regime. In this case, the comparison is between the elasticity of substitution between technology (or capital embodying the new technology) and labor, ε_ψ, and a composite term capturing both decreasing returns to labor and technology and to the structures capital. In particular, when $\eta = 0$, the condition in (B19) is equivalent to that in (B14); but when $\eta > 0$, this condition would be harder to satisfy for a given level of β because structures capital also adjusts, leaving less room for technology adjustment (though naturally in practice a higher η would correspond to a lower β).[40] Nevertheless, the qualitative insights are similar to those in proposition 2 and indicate that with sufficient decreasing returns and a sufficiently large degree of substitution between technology and labor, an increase in labor costs associated with the switch to partial cost reimbursement will induce technology adoption in the affected hospitals.

The important implication for our empirical work is that even if the price cap under the partial regulation regime is not very generous, so that overall labor inputs decline, technology-labor substitution may induce further technology adoption. Naturally, technology and capital expenditures on technology are more likely to increase when θ is positive; that is, with $\theta > 0$, they may increase even when $\varepsilon_\psi < (1 - \eta)/(1 - \beta - \eta)$. Nevertheless proposition 3 gives a useful benchmark and highlights the importance of substitutability between labor and technology (or capital).[41]

Another interesting implication of proposition 3 is that we could have a configuration in which expenditures on technology (and overall technology adoption) increase with the switch from full cost reimbursements to PPS, whereas total capital expenditures may decrease or remain unchanged, because they also include the component on structures expenditure. This is relevant for interpreting the empirical results below.

[40] In practice, the condition $\varepsilon_\psi \geq (1 - \eta)/(1 - \beta - \eta)$ in (B19) may not be too restrictive since, in addition to the structures capital, doctors' labor is excluded from the ψ function. Thus if we think of doctors as included in the factor z in terms of the original production function $\tilde{F}$, the parameter β would correspond to the share of technology (equipment capital) and nurse and custodian labor, whereas η is the share of structures capital. In addition, when $\theta > 0$, ε_ψ being greater than a lower threshold than $(1 - \eta)/(1 - \beta - \eta)$ is sufficient for the demand for capital to increase.

[41] In the health services sector, there is a natural substitution between technology and labor, which takes place by varying the length of stay in a hospital. Use of more high-tech equipment may save on labor by allowing patients to leave earlier, which amounts to substituting technology for labor. We investigate this issue empirically below.

5. Skill Composition of Employment

Finally, in our empirical work we also look at changes in the composition of the workforce, in particular, of nurses. To do this, the production function can be generalized to

$$F(A_i,\ U_i,\ S_i,\ K_i), \tag{B20}$$

where U_i denotes unskilled labor (nurses) and S_i denotes skilled labor (nurses). An increase in the capital-labor ratio and technology adoption will increase the ratio of skilled to unskilled labor as long as technology and/or capital is more complementary to skilled than to unskilled labor. To state the result here in the simplest possible form, suppose that A_i is fixed, so that the main effect of the change in regulation will work through an increase in the capital stock overall (including equipment as well as structures capital). We have the following proposition.

PROPOSITION 4. Suppose that $F(A_i, U_i, S_i, K_i)$ is homothetic in U_i, S_i, and K_i, and denote the (local) elasticity of substitution between U_i and K_i by σ_U and the elasticity of substitution between S_i and K_i by σ_S. Then

$$\frac{S_i^p}{U_i^p} \gtreqless \frac{S_i^f}{U_i^f} \quad \text{if and only if} \quad \sigma_S \lesseqgtr \sigma_U.$$

Moreover, the gap between S_i^p/U_i^p and S_i^f/U_i^f is greater when m_i is greater.

The proof is omitted.

This proposition therefore shows that when capital is more complementary to skilled than to unskilled labor, the removal of the implicit subsidy to labor involved in the change from full cost reimbursement to partial cost reimbursement will increase the skill composition of hospitals. A similar proposition could be stated for the case in which the main margin of adjustment is technology (embodied in capital), which would correspond to technology-skill complementarity rather than capital-skill complementarity.

6. Choice of Medicare Share

We now briefly discuss how the Medicare share of hospital i, m_i, can be endogenized. Suppose that the hospital produces two distinct "products," Medicare health services and non-Medicare health services (the latter may also include outpatient Medicare, which is reimbursed differently). Let the production functions for these two products be

$$F_m(A_{m,i},\ L_{m,i},\ K_{m,i}) \quad \text{and} \quad F_n(A_{n,i},\ L_{n,i},\ K_{n,i}),$$

with respective prices q_m and q_n and exogenous technology terms $A_{m,i}$ and $A_{n,i}$, and let

$$m_i = \frac{F_m(A_{m,i},\ L_{m,i},\ K_{m,i})}{F_m(A_{m,i},\ L_{m,i},\ K_{m,i}) + F_n(A_{n,i},\ L_{n,i},\ K_{n,i})} \tag{B21}$$

be the Medicare share of total output. Alternatively, we could have defined m_i as the Medicare share of total operating expenses, $m_i = L_{m,i}/(L_{m,i} + L_{n,i})$, or the Medicare share of capital expenses, $m_i = K_{m,i}/(K_{m,i} + K_{n,i})$, in both cases with identical results.

The maximization problem of the hospital under full cost reimbursement is

$$\begin{aligned} \max_{L_{m,i},K_{m,i},L_{n,i},K_{n,i},m_i} \quad \pi^m(i) = {} & q_m F_m(A_{m,i},\ L_{m,i},\ K_{m,i}) + q_n F_n(A_{n,i},\ L_{n,i},\ K_{n,i}) \\ & - (1 - m_i s_L) w (L_{m,i} + L_{n,i}) - (1 - m_i s_K) R (K_{m,i} + K_{n,i}) \end{aligned} \tag{B22}$$

subject to (B21).

This maximization problem can be broken into two parts. First, maximize $q_m F_m(A_{m,i}, L_{m,i}, K_{m,i}) + q_n F_n(A_{n,i}, L_{n,i}, K_{n,i})$ with respect to $L_{m,i}$, $K_{m,i}$, $L_{n,i}$, and $K_{n,i}$ for given m_i and subject to (B21) and to the constraints that $L_i = L_{m,i} + L_{n,i}$ and $K_i = K_{m,i} + K_{n,i}$. Define the value of the solution to this problem as $F(L_i, K_i, m_i)$, which depends only on the total amount of labor, $L_i = L_{m,i} + L_{n,i}$, and the total amount of capital, $K_i = K_{m,i} + K_{n,i}$.

Once this first step of maximization is carried out, the solution to the maximization under full cost reimbursement in (B22) can be obtained from

$$\max_{L_i, K_i, m_i} \tilde{\pi}^m(i) = F(L_i, K_i, m_i) - (1 - m_i s_L)wL_i - (1 - m_i s_K)RK_i.$$

Similarly, with the same assumptions as in the analysis so far, the maximization problem under the partial cost reimbursement regulation regime (with $\theta = 0$) can be written as

$$\max_{L_i, K_i, m_i} \tilde{\pi}^p(i) = F(L_i, K_i, m_i) - wL_i - (1 - m_i s_K)RK_i.$$

This implies that the analysis presented so far can be carried out as before, with the only addition that now m_i is also a choice variable. The following proposition generalizes proposition 1 to this case.

Proposition 5. Let the Medicare shares with full and partial cost reimbursement be, respectively, m_i^f and m_i^p; then as long as

$$\frac{m_i^f - m_i^p}{m_i^f(1 - m_i^p)} < \frac{s_L}{s_K}, \tag{B23}$$

the move from the full to the partial cost reimbursement regulation increases the capital-labor ratio, that is,

$$\frac{K_i^p}{L_i^p} > \frac{K_i^f}{L_i^f}. \tag{B24}$$

Proof. The first-order conditions with respect to capital and labor imply

$$\frac{F_K(A_i, L_i^p, K_i^p)/F_L(A_i, L_i^p, K_i^p)}{F_K(A_i, L_i^f, K_i^f)/F_L(A_i, L_i^f, K_i^f)} = \frac{(1 - m_i^f s_L)(1 - m_i^p s_K)}{1 - m_i^f s_K}.$$

The right-hand side of this equation being less than one is sufficient for (B24), which is in turn guaranteed by assumption (B23). QED

Notice that (B23) is automatically satisfied if $m_i^f \leq m_i^p$, and we obtain the same results as in our analysis so far. It is also straightforward to see, however, that when $\theta = 0$ (or is very small), profit maximization will lead to $m_i^f > m_i^p$. Nevertheless, as long as this decline in the Medicare share is not so large as to violate (B23), the basic insights from our analysis continue to apply (note, in particular, that [B23] is not very restrictive even with $m_i^f > m_i^p$).

Finally, the analysis leading up to proposition 5 did not introduce any explicit differences across firms in the composition of demand for health care that they face. Presumably, differences in Medicare shares reflect differences across hospitals in terms of the composition of their local demand. When such differences are introduced, proposition 5 can also be generalized in a straightforward manner so as to obtain the equivalents of the other results presented so far. For example, suppose that differences in the composition of local demand across hospitals can be parameterized by some hospital-specific time-invariant variable μ_i. Then, before PPS, (initial) differences in Medicare share, m_i, will be driven by differences in μ_i. In particular, high-μ_i hospitals will, in equilibrium, choose a higher Medicare share, m_i, both before and after the introduction of PPS. Then, under some mild regularity conditions, it can be shown that following the introduction of the PPS, m_i will change but will continue to reflect differences in μ_i. Consequently, the effect of PPS will be greater on hospitals with greater μ_i, and empirically, this will correspond to more pronounced effects on hospitals with a greater initial Medicare share.

B. Cost Plus Reimbursement without Fungibility

The analysis in the the baseline model of regulation presented in the previous section was simplified by the fact that we allowed the hospital to substitute labor (and capital) between the Medicare and non-Medicare products and focused on the case in which $s_L m_i < 1$ and $s_K m_i < 1$. The combination of these two assumptions implied that the hospital always faced positive marginal costs of hiring more labor, capital, and technology.

An alternative model would be one in which there is cost plus reimbursement, in the sense that for every dollar spent on capital or labor, the hospital receives more than one dollar back, that is, $s_L > 1$ and $s_K > 1$, and there is no fungibility. In this case, the model developed in this appendix so far needs to be modified since it would imply that the hospital would like to choose infinite amounts of capital and labor (unless F_L and F_K become negative). This not only would be unrealistic but also would run into regulatory constraints. This section briefly discusses how the analysis is modified once these regulatory constraints are incorporated. In particular, Medicare stipulates that hospitals can charge for "reasonable and customary" costs for Medicare services. We interpret this as implying that the amount of reimbursement required by the hospitals has to be less than a fraction of the average productivity of each factor that is being reimbursed under Medicare.

Let us simply focus on the Medicare services provided by the hospital and ignore technology choices (which, as before, can be incorporated in a straightforward manner). Moreover, assume throughout that $s_L > 1$ and $s_K > 1$. This implies that the profits of the hospital i are

$$\pi^f(i) = qF(L_i, K_i) + s_L w\tilde{L}_i + s_K R\tilde{K}_i - wL_i - RK_i, \tag{B25}$$

where L_i and K_i are the total amounts of capital and labor hired by the hospital, and $\tilde{L}_i$ and $\tilde{K}_i$ are the total amounts of labor and capital for which the hospital requests reimbursement from Medicare. Although we have assumed that there is no fungibility, in the sense that the hospital cannot demand reimbursement for labor and capital used for other purposes, it can always use additional labor and capital for Medicare-related activities even if it does not ask for reimbursement. We will see that this might be useful depending on how tight the reimbursement constraints imposed by Medicare are.

In particular, we model these constraints as follows:

$$s_L w\tilde{L}_i \le B_L F(L_i, K_i) \tag{B26}$$

and

$$s_K R\tilde{K}_i \le B_K F(L_i, K_i). \tag{B27}$$

Simply put, these constraints require the reimbursement received from Medicare for labor and capital not to exceed a certain fraction of the health services provided to Medicare patients. To clarify this interpretation, for example, (B26) can be expressed as $s_L w/B_L \le F(L_i, K_i)/\tilde{L}_i$, which shows that this constraint equivalently requires the average product of labor (used for reimbursement) not to exceed a certain threshold.

All the other assumptions from the main model, in particular, that F is increasing, strictly concave, and twice continuously differentiable in both of its arguments, still apply. The constraints (B26) and (B27) also explain why we had to allow for the hospital to be able to choose more labor and capital than the amounts for which it demands reimbursement from Medicare. In particular, imagine that B_L is very small (in the limit, $B_L \to 0$). If we had imposed that $\tilde{L}_i = L_i$ and labor were an essential factor of production, then the hospital would have to shut down; but with our formulation, and in reality, it can function profitably by choosing $\tilde{L}_i < L_i$. This discussion also shows that if the reimbursement constraints (B26) and (B27) are not too binding, the solution will typically have $\tilde{L}_i = L_i$ and $\tilde{K}_i = K_i$.

Consequently, under full cost (plus) reimbursement, the firm chooses $\tilde{L}_i$, L_i, $\tilde{K}_i$, and K_i to maximize (B25) subject to (B26), (B27), and the natural constraints arising from nonfungibility that $\tilde{L}_i \le L_i$ and $\tilde{K}_i \le K_i$ (so that the amounts of labor and capital reimbursed are less than the total amount of labor and capital used in Medicare-related activities).

Lemma 1. Profit maximization implies that with full cost reimbursement, both (B26) and (B27) will be binding.

Proof. Suppose not and that, for example, (B26) is slack. Since F is increasing in L_i and $s_L > 1$, the hospital can set $\tilde{L}_i = L_i$ and increase L_i until (B26) binds; this will increase the value of profits in (B25), yielding a contradiction. The same argument applies to (B27), proving the lemma. QED

This lemma enables us to substitute for constraints (B26) and (B27) and write the maximization problem under full cost reimbursement regulation as follows:

$$\max_{\tilde{L}_i, L_i, \tilde{K}_i, K_i} (q + B_L + B_K)F(L_i, K_i) - wL_i - RK_i \tag{B28}$$

subject to $\tilde{L}_i \leq L_i$ and $\tilde{K}_i \leq K_i$. Intuitively, if the hospital will hire more labor or capital than what it demands reimbursement for, the marginal cost of this labor and capital will be given by the factor market prices, w and R, and the amounts $B_L F(L_i, K_i)$ and $B_K F(L_i, K_i)$ will be perceived by the hospital as lump-sum transfers. Alternatively, the firm will hire exactly $\tilde{L}_i$ and $\tilde{K}_i$.

The first-order conditions of this problem are

$$(q + B_L + B_K)F_L(L_i^f, K_i^f) \geq w \quad \text{and} \quad \tilde{L}_i^f \leq L_i^f \tag{B29}$$

and

$$(q + B_L + B_K)F_K(L_i^f, K_i^f) \geq R \quad \text{and} \quad \tilde{K}_i^f \leq K_i^f, \tag{B30}$$

both holding with complementary slackness and f denoting full cost reimbursement.

Lemma 1 has another important implication for our analysis. If the solution to the maximization problem of the hospital involves $\tilde{L}_i^f = L_i^f$ and $\tilde{K}_i^f = K_i^f$, then (B26) and (B27) define two equations in two unknowns, $\tilde{L}_i^f$ and $\tilde{K}_i^f$; moreover, decreasing returns to capital and labor imply that there exists a unique tuple (L^*, K^*) satisfying these two equations. Therefore, if we have the second inequalities in (B29) and (B30) hold as equality, we must have $\tilde{L}_i^f = L_i^f = L^*$ and $\tilde{K}_i^f = K_i^f = K^*$. The above discussion then suggests that as long as (B26) and (B28) are not very restrictive (i.e., are sufficiently generous), we will be in a situation in which the firm hires the levels of labor and capital that will exactly satisfy these two constraints, (L^*, K^*).

Next let us turn to the partial cost reimbursement regime, where there is no reimbursement for labor, so the constraint (B26) and s_L are removed, and the firm now receives $q + B_P$ per unit of Medicare health services, where $B_P \geq 0$. The maximization problem then becomes

$$\pi^f(i) = (q + B_P)F(L_i, K_i) + s_K R\tilde{K}_i - wL_i - RK_i \tag{B31}$$

subject to (B27) and $\tilde{K}_i \leq K_i$. We then immediately have the following result, which parallels lemma 1.

Lemma 2. Profit maximization implies that with partial cost reimbursement, (B27) will be binding.

The proof is omitted.

Consequently, the maximization problem of the firm can be written as

$$\max_{\tilde{L}_i, L_i, \tilde{K}_i, K_i} (q + B_P + B_K)F(L_i, K_i) - wL_i - RK_i$$

subject to $\tilde{K}_i \leq K_i$. In this case, we have the following first-order conditions:

$$(q + B_P + B_K)F_L(L_i^p, K_i^p) = w$$

and

$$(q + B_P + B_K)F_K(L_i^p, K_i^p) \geq R \quad \text{and} \quad \tilde{K}_i^p \leq K_i^p,$$

with the second condition holding with complementary slackness.

The difficulty in the analysis in this case stems from the fact that either (B26) or (B27) could be very tight, with correspondingly large Lagrange multipliers. For example, this would be the case when $B_L \rightarrow 0$, so that there was effectively no reimbursement of labor because of the tightness of the "reasonable and customary" constraint. Nevertheless, the following proposition can be established.

Proposition 6. Suppose that under full cost reimbursement $L_i^f = L^*$ and $K_i^f = K^*$. Consider a change to partial cost reimbursement with $B_P < B_L$. Then we have

$$L_i^p < L_i^f. \tag{B32}$$

Moreover, if F is homogeneous of degree $\beta < 1$ in L_i and K_i, then

$$\frac{K_i^f}{L_i^f} < \frac{K_i^p}{L_i^p}. \tag{B33}$$

Proof. The first-order conditions for (B28) imply that $(q + B_L + B_K)F_L(L^*, K^*) \geq w$ and $(q + B_L + B_K)F_K(L^*, K^*) \geq R$, whereas the first-order conditions for (B31) imply $(q + B_P + B_K)F_L(L_i^p, K_i^p) = w$ and $(q + B_P + B_K)F_K(L_i^p, K_i^p) \geq R$. To obtain a contradiction, suppose that $L_i^p \geq L_i^f$. Lemma 2 implies that (B27) holds as an equality. Since $L_i^p \geq L_i^f = L^*$, (B27) then implies $K_i^p \geq K_i^f = K^*$. First, suppose that $K_i^p = K^*$. Then diminishing returns to labor implies that $(q + B_L + B_K)F_L(L^*, K^*) \geq w$ is inconsistent with $(q + B_P + B_K)F_L(L_i^p, K^*) = w$, $L_i^p \geq L^*$, and $B_P < B_L$, yielding a contradiction. Second, suppose that $K_i^p > K^*$. Then (B30) implies $(q + B_P + B_K)F_K(L_i^p, K_i^p) = R$. Then $B_P < B_L$ implies that

$$(q + B_P + B_K)F_L(L_i^p, K_i^p) > (q + B_L + B_K)F_L(L^*, K^*),$$

$$(q + B_P + B_K)F_K(L_i^p, K_i^p) > (q + B_L + B_K)F_K(L^*, K^*),$$

which is inconsistent with $K_i^p > K^*$ and $L_i^p \geq L^*$ given decreasing returns, yielding another contradiction and establishing that we must have $L_i^p < L^*$, that is, (B32).

To obtain (B33), first note that if $K_i^p \geq \tilde{K}_i^f = K^*$, given (B32), (B33) would apply immediately. Therefore, we only have to show that it also holds when $K_i^p < \tilde{K}_i^f = K^*$. Suppose this is the case. Then, use lemma 2 and the homogeneity assumption on F to reexpress (B27) as

$$s_K R \tilde{K}_i^p (K_i^p)^\beta \leq B_K F\left(\frac{L_i^p}{K_i^p}, 1\right).$$

Since $\tilde{K}_i^p \leq K_i^p < \tilde{K}_i^f = K^*$, it must be that $L_i^p / K_i^p < L_i^f / K_i^f$, establishing (B33). QED

This proposition generalizes the results from our basic analysis with fungibility in the previous section to the case without fungibility, though the results are weaker since they hold under some additional conditions. Most important, the main results apply as long as the full cost reimbursement is sufficiently generous to start with so as to ensure $L_i^f = \tilde{L}_i^f = L^*$ and $K_i^f = \tilde{K}_i^f = K^*$, and partial cost reimbursement is less generous than full cost reimbursement as captured by the condition that $B_P < B_L$. Both of these appear as plausible conditions in the context of the PPS reform.