Discussion Papers in Economics

TECHNOLOGY, UTILIZATION AND INFLATION: WHAT DRIVES THE NEW KEYNESIAN PHILLIPS CURVE?

By

Peter McAdam
(University of Surrey and European Central Bank)

&

Alpo Willman
(European Central Bank)

DP 09/12

Department of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web www.econ.surrey.ac.uk
ISSN: 1749-5075

Technology, Utilization and Inflation: What Drives the New Keynesian Phillips Curve?

Peter McAdam and Alpo Willman*
European Central Bank & University of Surrey, European Central Bank

July 8, 2012

Abstract

We argue that the New-Keynesian Phillips Curve literature has failed to deliver a convincing measure of real marginal costs. We start from a careful modeling of optimal price setting allowing for non-unitary factor substitution, non-neutral technical change and time-varying factor utilization rates. This ensures the resulting real marginal cost measures match volatility reductions and level changes witnessed in many US time series. The cost measure comprises conventional counter-cyclical cost elements plus pro-cyclical (and co-varying) utilization rates. Although pro-cyclical elements seem to dominate, the components of real marginal cost components are becoming less cyclical over time. Incorporating this richer driving variable produces more plausible price-stickiness estimates than otherwise and suggests a more balanced weight of backward and forward-looking inflation expectations than commonly found. Our results challenge existing views of inflation determinants and have important implications for modeling inflation in New-Keynesian models.

JEL Classification: E20, E30.

Keywords: Inflation, Real Marginal Costs, Production Function, Labor Share, Cyclicality, Utilization, Intensive Labor, Overtime Premia.

*Corresponding author.

E-mail addresses: peter.mcadam@ecb.europa.eu (P. McAdam), alpo.willman@ecb.europa.eu (A. Willman). This paper is a substantially revised and updated version of our earlier ECB working paper, No. 1369.

1 Introduction

New Keynesian Phillips Curves (NKPC) have become a popular means of analyzing inflation. They have been widely estimated (see Roberts (1995), Galí and Gertler (1999) for seminal contributions) and their merits much debated (Fuhrer (1997, 2010), Rudd and Whelan (2007), Batini (2009)). Primarily, the NKPC models inflation as a function of its expectation and some real-activity *driving variable.*

The literature has mostly focused on dynamic and expectations issues. In other words, to what degree inflation expectations are forward and/or backward looking. Less effort has been spent on how to treat the driving variable. A reflection of this may be the fact that there scarcely exists even a consensus on its cyclical nature; in the literature, the two common candidates – output gap, labor share – are respectively, pro- and counter-cyclical.

These uncertainties have empirical consequences. For instance, typical Phillips-curve "slopes" – which measure the responsiveness of inflation to the driving variable, and capture price stickiness – have been curiously flat, and often statistically insignificant.[1] This flatness contradicts micro evidence which suggesting more frequent price adjustments. An additional consideration, forcefully made by Fuhrer (2010), is that volatility patterns observed in many US time series (including inflation) since the mid-1980s appear unmatched by that in candidate driving variables.

The question of what drives inflation needs addressing for many reasons. Using a mis-specified driving variable may distort our understanding of the persistence, pressures and sources of inflation. For instance, a policy maker who views the Philips-curve slope as flat may operate very differently to one believing it steep. This has consequences for the success of stabilization policy, and the anchoring of inflation expectations.[2] Likewise, a policy maker predicating policy on the basis of a driving variable with the "wrong" pattern of cyclicality risks destabilizing policy and, in general, producing systematic forecast errors.

Against this background, our contribution is two fold:

1. We develop a more complete specification of the driving variable (real marginal costs). We allow for non-unitary capital-labor factor substitution, non-neutral technical change and disentangle technical progress from (co-varying) factor utilization rates.

2. We then ask: how costly is the use of a mis-specified driving variable? Does it affect NKPC estimates of inflation persistence and stickiness? Would a "better" measure challenge our views on the business-cycle properties of real marginal costs?

In their landmark overview, Rotemberg and Woodford (1999) reviewed means to improve real marginal cost measures: non Cobb Douglas production technology; overtime pay, labor adjustment costs and labor hoarding; variable capital utilization; and overhead labor (or fixed

[1] Updated estimates of the Galí and Gertler (1999) results have shown the driving variable to be insignificant, see Rudd and Whelan (2007) Gabriel and Martins (2009).

[2] At the extreme if modeled inflation is uncoupled from the real economy (i.e., zero slope), indeterminacy results (i.e., inflation expectations cannot be anchored).

costs). Our paper can be viewed as empirically taking up all of those issues in a unified but tractable framework.

Regarding technology choice, we estimate a Constant Elasticity of Substitution. This nests the more familiar Cobb-Douglas (CD) case. Following León-Ledesma et al. (2010a), we estimate production and technology relationships as a system with cross-equation restrictions. We also model technical progress as time-varying and "factor augmenting". These features bring production-technology relations markedly closer to the data and provide an intuitive explanation for changes in the level and growth of productivity and TFP growth over time.

This last aspect is worth emphasizing. Unlike CD, the CES form admits the possibility for non-neutral technical change; indeed modern growth literatures suggests that there is little obvious reason to believe that technical change will be neutral (or mimic balanced growth). Evidence also suggests that factor substitution is low in both the short and long run (Chirinko (2008)). Likewise, the various waves of productivity growth and swings in factor income shares in US data counsel against a simplistic modeling of economic supply.

Notwithstanding, however well production-function based real marginal costs measures are derived, they remain incomplete since they assume hired inputs are continuously in full use. Variations in utilization and capacity, though, are a well-established empirical phenomenon; since Solow (1957), we have known of the need to disentangle technical change from factor utilization rates. Accordingly, in our theoretical framework we attach convex costs to changes in factor utilization rates. These introduce an additional dependency of marginal costs on utilization. Overall utilization rates can be derived as the residual of the estimated production function, which we can then map into the individual factor utilization rates (which in turn co-vary).

The bottom line is that we arrive at a "full" real marginal cost measure comprising a weighted average of real marginal costs excluding utilization, plus utilization costs. The net cyclicality of this new measures boils down to empirics: if demand shocks dominate we might expect the driving variable to be net pro-cyclical, and vice-versa for supply shocks. The more likely outcome, though, is the coexistence of *both* types of shocks. Similarly, if some channels (such as capital deepening) are more important than others at certain times, those channels will then dominate the evolution of marginal costs. When our preferred cost measure is inserted into NKPCs as the driving variable, price stickiness becomes more consistent with micro studies, and the weight on backward and forward-looking expectations become balanced.

Regarding, other contributions in the literature, Gagnon and Khan (2005) and Gwin and VanHoose (2008) found that different measures of real marginal costs (respectively, CES production and overall industry-based cost measures), had little effect on NKPC estimates. However, in neither of those papers was there any discussion of appropriate cyclicality properties of real marginal cost measures, the incorporation of technical progress, factor utilization rates, or volatility mappings between the driving and explanatory variable. Mazumder (2010), by contrast, using US manufacturing data incorporates labor utilization into real marginal cost (albeit fitting a truncated polynomial), but finds a negative Phillips curve slope coefficient. Madeira

(2013) performs an exercise with aggregate US data using employment frictions in a New Keynesian model, and finds – like us – a closer fit to micro price stickiness estimates than standard NKPC estimations. Nekarda and Ramey (2010) examine mark-up cyclicality on US disaggregate data, although they focus less on full-capacity output specifications, volatility mappings and utilization co-movements, as here.

The paper is organized as follows. Section 2 restates the NKPC framework. Section 3 discusses ways to construct real marginal costs. It argues that traditional measures are incomplete because they do not account for factor utilization rates, or, effectively, technical progress. We then define economically plausible choices for factor utilization, alongside CES production. Section 4 defines the firm's profit maximization problem. Utilization rates are then shown to be naturally co-varying. Given this, we derive real marginal costs as incorporating a "conventional" and utilization-based component. Section 5 defines our US macro data sources and transformations. Section 6 estimates the production-technology system from which we derive full real marginal costs, and then the various NKPC estimations. Finally, we conclude.

2 The NKPC

As in Galí and Gertler (1999) and subsequent literature, we assume staggered price setting under imperfect competition, where a fraction θ of firms do not change their prices in any given period. The remaining firms set prices optimally as a fixed mark-up, μ, on discounted expected real marginal costs.[3] When resetting, firms also take into account that the price may be fixed for many future periods, yielding the reset price p_t^*,

$$p_t^* = \mu + (1-\theta\beta)\,\mathbb{E}_t \sum_{k=0}^{\infty} (\theta\beta)^k \, mc_{t+k}^n \tag{1}$$

where mc^n is (the log of) nominal marginal costs, β is a discount factor, and $\mathbb{E}_t$ is the expectation operator. The overall price level is then a weighted average of lagged and reset prices, $p_t = \theta p_{t-1} + (1-\theta)\,p_t^*$. Given $mc_t^r = mc_t^n - p_t$, and constant marginal costs across firms, the familiar "New Keynesian Phillips Curve" emerges,

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \lambda\,(mc_t^r + \mu) \tag{2}$$

where $\pi_t = p_t - p_{t-1}$ is inflation and $\lambda = \frac{(1-\theta)(1-\theta\beta)}{\theta}$ represents the slope of the Phillips curve.

Additionally, it is often assumed that of the $1-\theta$ price-re-setting firms a fraction, ω, set their price according to lagged inflation. This implies a NKPC with an intrinsic expectations component,

$$\pi_t = \gamma_f \mathbb{E}_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda\,(mc_t^r + \mu) \tag{3}$$

where $\phi = \theta + \omega\,[1-\theta\,(1-\beta)]$. Parameters $\gamma_f = \frac{\theta\beta}{\phi}$, $\gamma_b = \frac{\omega}{\phi}$ and $\lambda = \frac{(1-\omega)(1-\theta)(1-\theta\beta)}{\phi}$ capture, respectively, "extrinsic", "intrinsic" and "inherited" inflation persistence.

[3] For work on time-varying markups, see, e.g., Chirinko and Fazzari (1994) and Galí et al. (2007).

3 Real Marginal Costs

Providing a richer, more unified and intuitive treatment of the real marginal cost measures, mc_t^r, is our purpose. We allow for non-unitary factor substitution and non-neutral technical change. We also disentangle technical progress from (co-varying) factor utilization rates.

This decomposes marginal costs into a "conventional" and a utilization-based component. Both have business cyclic characteristics: the former (of which labor share is a limiting case) tends to be counter cyclical, whilst the latter is pro-cyclical.

Real marginal costs are, admittedly, difficult to measure. An early approach to capturing the driving variable was to use the deviation of output from a HP filter or a linear/quadratic trend. However, often these non-structural measures entered with the "wrong" (negative) sign. Alternatively, Galí and Gertler (1999) argued in favor of proxying real marginal costs by average real unit labor costs. Under the special case of a (unitary substitution elasticity) CD production function, real marginal costs then reduce to the labor share.

The advantage of using the labor share is that it is observable, simple[4] and tended to yield the "correct" slope sign (albeit not always significant nor quantitatively important). The disadvantage is largely three fold:

1. Labor share is counter-cyclical. By contrast, theory suggests output increases not driven by technological improvements tend to raise nominal marginal costs more than prices. If so, real marginal costs should be pro-cyclical (Bils (1987), Rotemberg and Woodford (1999));[5]

2. Reflecting its Cobb-Douglas origins, the labor share based real marginal cost measure is underpinned by a counter-factual unitary elasticity of factor substitution and necessarily excludes any identifiable role for (biased or non-neutral) technical change;

3. Using labor share as a measure of real marginal costs implies that either the number of workers or their utilization rate can be adjusted costlessly at a fixed wage rate.

Over business-cycle frequencies *all* these features (unitary substitution; indeterminate technical progress; zero adjustment costs; fully utilized factors) appear unnecessarily restrictive and counter factual. And it is these weaknesses we address. Regarding points 1 and 2, we estimate an aggregate CES production-technology system embodying time-varying technical progress that augments both factors. To guard against point 3, we allow utilization rates to vary (and co-vary) over the business cycle, thus changing marginal costs in the process.

[4]It does not, for instance, even require explicit production function estimation and allows researchers to abstract from capital accumulation.

[5]Equivalently, a counter-cyclical labor share implies that the markup of (sticky) prices over marginal costs would be pro-cyclical. By contrast, the theory that suggests output increases not driven by technological improvements that tend to raise nominal marginal costs more than prices, would imply that mark-ups should be counter-cyclical.

3.1 Real Marginal Costs and the CES Production Function

Consider the CES production function:

$$Y_t = F\left(\Gamma_{K,t}\mathbb{K}_t,\, \Gamma_{N,t}\mathbb{N}_t\right) = \left[\alpha(\Gamma_{K,t}\mathbb{K}_t)^{\frac{\sigma-1}{\sigma}} + (1-\alpha)\left(\Gamma_{K,t}\mathbb{N}_t\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{4}$$

where $\alpha \in (0,1)$, and $\sigma \geq 0$ is the elasticity of substitution between "effective" capital ($\mathbb{K}$) and labor ($\mathbb{N}$). By "effective", we mean factor inputs controlling for measures of utilization: $\mathbb{K}_t = \kappa_t K_t$, and $\mathbb{N}_t = h_t N_t$, where $\kappa_t \in [0,1]$ and $h_t \geq 0$ denote the (naturally pro-cyclical) utilization rates of capital and labor, respectively.

The higher is σ, the more alike (or *substitutable)* factors of production are. If $\sigma \to \infty$, they are completely interchangeable; if $\sigma = 0$, they are locked in fixed proportions. CD arises as $\sigma = 1 \Rightarrow Y_t = A_t K_t^{\alpha} N_t^{1-\alpha}$ where A_t is the "Solow residual".[6] Despite its popularity, CD is routinely rejected by aggregate data[7], as is its prediction of constant factor income shares (see Jones (2003), McAdam and Willman (2013)).

In (4) $\Gamma_{K,t}$ and $\Gamma_{N,t}$, moreover, capture capital and labor-augmenting technical progress components. These can apply commonly to both factors equally as in $\Gamma_{K,t} = \Gamma_{N,t}$ ("Hicks neutrality"), only to labor ("Harrod neutrality"), only to capital ("Solow neutrality"), or indeed to both individually $\Gamma_{K,t} \neq \Gamma_{N,t}$ ("Factor Augmenting", or "biased" technical change).

As we know, modern business cycle models tend to impose aggregate (unitary elasticity) Cobb Douglas. This is doubly unfortunate since CD cannot separately identify labor and capital augmenting technical progress (Barro and Sala-i-Martin (2004, pp. 78-80) provide a simple proof). Many growth economists however argue that there is little empirical justification to suppose that, over business cycles, technical progress will be neutral: any kind of technical change creates winners and losers, and augments some factors more than others, e.g., Acemoglu (2002a,b). Furthermore, the various waves of productivity growth in the US (accelerating in the 1950s and 1990s, flattening out in the 1970s and 80s) underpins the need for a careful treatment.

All these aspects fuse together fruitfully when we construct real marginal costs: the ratio of the real wage, w_t, to the marginal product of labor. Given (4), this becomes,[8]

$$MC_t^r \equiv \frac{W_t/P_t}{F_{\mathbb{N},t}} = \begin{cases} \frac{w_t}{(1-\alpha)}\left(\frac{N_t}{Y_t}\right)^{1/\sigma}\Gamma_{N,t}^{(1-\sigma)/\sigma} h_t^{1/\sigma} & if \quad \sigma \neq 1 \qquad (\mathbf{a}) \\ \frac{1}{(1-\alpha)}\frac{w_t N_t}{Y_t} h_t & if \quad \sigma = 1 \qquad (\mathbf{b}) \end{cases} \tag{5}$$

Equation (5b) reveals the proportionality between real marginal costs and the labor income share under CD. With non-unitary substitution (5a), however, there is an additional role for (trend) labor augmenting technical progress (and an indirect effect for capital augmentation). The size of the substitution elasticity (i.e., unitary/non-unitary) only affects the impact with

[6] Implicitly $A_t = \Gamma_t \kappa_t^{\alpha} h_t^{1-\alpha}$ (Γ_t denotes non-cyclical (trend) technical progress).

[7] Klump et al. (2007) and Chirinko (2008) suggest $0.4 - 0.6$ as a benchmark aggregate elasticity range for the US.

[8] Alternatively, marginal cost can be expressed in term of ratio of the user cost and marginal product of capital; an optimizing firm would naturally equalize marginal costs across all factor margins (see below).

which the various channels are transmitted into marginal costs, thus both measures (5(a) and 5(b)) will have similar cyclicality properties (namely, counter-cyclical[9]). Although their cyclicality is common, their moments will differ with important consequences for tracking performance. Both measures are also affected by the labor utilization margin, h_t, which, by contrast, is necessarily pro-cyclical. As the following sections demonstrate, the role of time-varying labor utilization becomes non-trivial if changes in the labor utilization rate are associated with convex costs that makes the wage rate w_t dependent on h_t.

Since (5b) is a limiting case of (5a), we concentrate on the latter. How do we implement it in practice? First, obviously, we require estimates of σ, α, and $\Gamma_{N,t}$ and $\Gamma_{K,t}$. To achieve that, we estimate aggregate production-technology relationships allowing for an unrestricted substitution elasticity, and non-neutral factor augmenting technical progress. In doing so, we make the identifying assumption that growth in technical change is smoothly evolving but non-constant.[10] Following Klump et al. (2007) and others, we model technical progress components as a flexible Box-Cox transformation. This provides an informative (albeit reduced form) means to capture smoothly-evolving technical progress.[11]

Next, to complete the computation of expression (5), we require some measure of aggregate labor utilization and how labor costs (wages) vary in response to work above (and below) "normal" working hours (given that aggregate wages include a straight time and non-straight time rate). We demonstrate that although latent at the aggregate level, labor and capital utilization rates can be uncovered from the overall utilization rate (i.e., the production-function residual, see Appendix C) which is by definition a function of the individual utilization rates: $u_t = u[h_t, \kappa_t]$. It is to these aspects that we now turn.

3.2 Utilization measures

The prerequisite for variation in factor utilization rates is that a firm cannot costlessly change its factor composition. Without adjustment costs, inputs would operate continuously at maximal intensity. They create a short-run trade-off between changes in installed inputs and the intensities at which they are used.

Individual utilization rates are in general unobserved. Measures of labor utilization are avail-

[9]The labor share, $\frac{wN}{Y}$, is counter cyclical because observable labor productivity, Y/N, is pro-cyclical commonly thought due to labor hoarding, whereas the real wage is largely a-cyclical.

[10]Basu et al. (2006) estimated the contribution of factor utilization to the Solow residual and found that the 'purified' TFP followed a random walk with no serial correlation in the residual, implying practically a-cyclical TFP.

[11]Box-Cox is a flexible functional form that nests various possible dynamic growth paths of a variable around its average using a single curvature parameter, λ: $\Gamma_t^j = e^{\gamma_t^j}$ where,

$$\gamma_t^j = \frac{\gamma_j}{\lambda_j}\left(\widetilde{t}^{\lambda_j} - 1\right),\ j = K, N$$

where $\widetilde{t} = t/t_0$ (0 denotes average or normalization point), and parameter γ_j is the growth rate of technical progress normalized at the sample average; how the growth rate of technical progress varies around that average value is given by $\{\gamma_t^j\}$. Parameter $\lambda = 1$ yields the (textbook) linear specification; $\lambda = 0$ a log-linear specification; and $\lambda < 0$ a hyperbolic one for technical progress. $\lambda > 1$, describe explosive patterns of growth in technical progress.

able only for US Manufacturing (which currently only accounts for around 10% of employment and GDP). Capital utilization series, moreover, are generally unavailable.[12] Accordingly, we examine some candidate functions that are tractable and economically plausible.

3.2.1 Effective Labor, $h_t N_t$.

Typically, around two-thirds of the variation in total hired hours originates from employment; the rest from changes in hours per worker, e.g., Hart (2004). The relatively small proportion of the variation of paid hours per worker reflects the fact that labor contracts are typically framed in terms of "normal" working hours and that there are extensive labor adjustment (hiring-&-firing) costs. Therefore, it is difficult for firms to reduce hired hours per worker below that norm and often impossible to raise them without increasing marginal costs. Accordingly, it may be optimal for firms to allow the *intensity* at which hired labor is utilized to vary in response to demand pressures. Hired hours may therefore underestimate the true cyclical variation of utilized labor.

Following the indivisible labor literature (e.g., Kinoshita (1987), Trejo (1991), Rogerson (1988)), we assume contracts are defined in terms of fixed (or normal) working hours per employee, i.e., in terms of the "straight-time" wage rate. Hours in excess of normal hours may attract a premium.[13] That part is standard. But we also assume firms have locally limited possibilities to decrease paid hours (and thus costs) when de facto worked hours fall below normal ones. Total wage costs per employee can therefore be presented as a convex function of the deviation of the labor utilization rate h_t from "normal" hours $\overline{h}$ (set to 1 for convenience).[14]

Using a variant of the "fixed-wage" model of Trejo (1991) for overtime pay, the following function gives a local approximation of this relation in the neighborhood of effective hours equalling normal hours,[15]

$$W_t = \overline{W}_t \left[h_t + \frac{a}{2} (h_t - 1)^2 \right] \tag{6}$$

where W_t is the total nominal wage per worker and $\overline{W}_t$ is the nominal straight-time wage.[16] Parameter $a \geq 0$ measures the degree of convexity in the total wage. If there is no overtime or no convexity in the wage schedule ($h_t = 1$ and/or $a = 0$) wages remain at their straight-time rate, $W_t = \overline{W}_t$. When overtime and a convex premium exists, $\frac{a}{2}(h_t - 1)^2$ measures the escalation of wage costs. Alternatively, when labor utilization falls *below* the norm, $h_t < 1$, pressure on wages

[12] Sometimes proxies such as electricity usage are used for particular sectors.

[13] The share of private US industry jobs with overtime provisions is around 80%, and higher in some occupational groups (machinery operation; transport; administrative services), Barkume (2007). Overtime is defined by the Code of Federal Regulations as payments when hours worked exceed that required by the employee's contract or extra payments associated with special workdays: weekends, holidays.

[14] Whilst the overtime pay schedule of a single worker takes a kinked form, this is not so at a firm level and even less on higher aggregation levels, if there are simultaneously employees working at less than full intensity and those working overtime at full intensity (see Bils (1987)).

[15] Shapiro (1986) and Bils (1987) used quite similar overtime premium specifications, but unlike us they make no allowance for cost changes when $h_t < \overline{h}$.

[16] As in Trejo (1991), the straight time wage rate for the firm is given and, conditional on the wage-cost schedule (6), effective hours are completely demand determined.

(relative to the straight-time rate) tends to be weak and head count, at least initially, tends to be unchanged (e.g., Burnside and Eichenbaum (1996)).

Figure 1 (starting from $h_t = \overline{h} = 1, \overline{W} = 100$) illustrates schedule (6) for three a values and a h continuum.[17] Note, if extensive labor adjustment costs are zero, all adjustment can be done via this margin and, independently from the size of a, $\mathbb{N}_t = N_t$ for all t.[18]

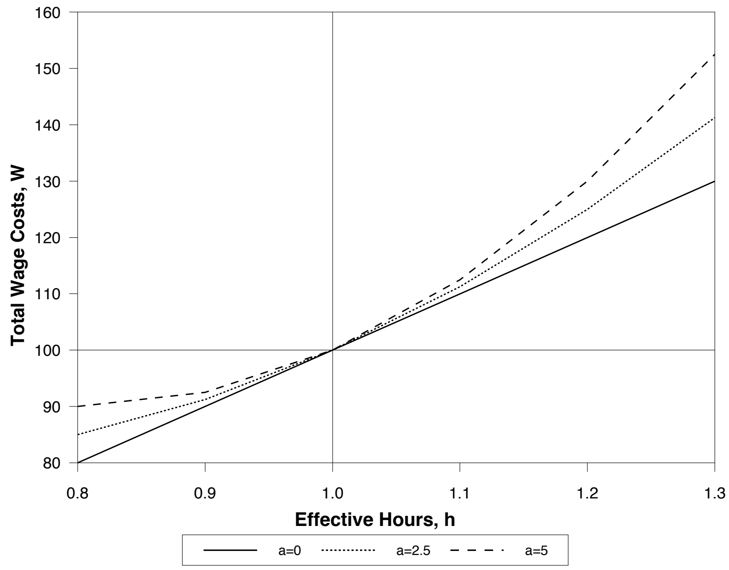

Figure 1: Utilization Rates under Effective Hours

3.2.2 Effective Capital, $\kappa_t K_t$.

As with labor, we assume that increases in the capital utilization rate, κ_t increases marginal costs, at an increasing rate (towards infinity when the utilization rate approaches unity): $\Phi'(\kappa_t)$, $\Phi''(\kappa_t) > 0$. Moreover, it is likely that utilization costs co-vary over the cycle; capital equipment used at above-normal capacity will have implications for labor usage, and vice versa.

Next, we turn to the firm's maximization problem. From that we can derive the firm's optimal combination of factor utilization rates and how they co-move. We further derive a relationship between the observed capacity utilization rate and individual factor utilization rates.

4 The Firm's Maximization Problem

Assume that firms operate in a monopolistically competitive environment and produce one good of variety f with the following production function:

$$F(\Gamma_{K,t}\kappa_{ft}K_{ft}, \Gamma_{N,t}h_{ft}N_{ft}) - \chi_{ft} \tag{7}$$

[17] A similar choice of functional form for labor costs is the Linex function, Varian (1974). However, we found our NKPC estimations were very similar upon using this functional form. We thank Frank Smets for this suggestion.

[18] For a more rigorous analysis of labor adjustment costs, see Cooper and Willis (2009).

where χ represents fixed costs in production. Assume further that demand for good f is given by,

$$Y_{ft} = \left(\frac{P_{ft}}{P_t} \right)^{-\varepsilon} Y_t \tag{8}$$

Let $\Omega_j(\cdot)$ refers to a factor adjustment cost function. Total labor costs for the firm comprise real wages times employment plus adjustment costs (the latter priced at the straight-time real wage):

$$\frac{W_t}{P_t} N_{ft} + \frac{\overline{W}_t}{P_t} \Omega_N \left(N_{ft}, N_{ft-1} \right) \tag{9}$$

where W_t is determined by (6). The total costs of capital evolve according to,

$$I_{ft} + \Phi\left(\kappa_{ft}\right) K_{ft} + \Omega_K \left(K_{ft}, K_{ft-1}, K_{ft-2} \right) \tag{10}$$

where investment $I_{ft} = K_{ft} + \delta K_{ft-1}$, with depreciation rate $\delta \in (0, 1)$. In our case, note, the firm's optimizing conditions constitute an unusually rich framework: incorporating factor-augmenting technical progress, dual utilization margins, dual adjustment costs, a mark-up and a fixed cost. A full treatment, including proofs, is given in Appendix A.

However, technicalities aside, it is intuitive that an optimizing firm would equalize the marginal cost of raising output across all factor margins. In our case, this amounts to the symmetric expression,

$$\frac{\Phi'\left(\kappa_{ft}\right)}{F_{K_{ft}}} \widetilde{\kappa}_{ft} = \frac{W'_{h_{ft}}}{F_{N_{ft}}} h_t \tag{11}$$

where $\widetilde{\kappa}_{ft} = \frac{\kappa_{ft}}{\kappa_{ss}}$ ($\kappa_{ss} \leq 1$ denotes equilibrium utilization), with the derivative $W'_{h_{ft}} = \frac{\partial W_{ft}}{\partial h_{ft}} \equiv \overline{W}_t \left(1 + a \left(h_{ft} - 1\right)\right)$.

Equation (11) defines the relationship between capital and labor utilization rates. A closed-form is obtained after applying a first-order Taylor approximation to (11):

$$\log \widetilde{\kappa}_{ft} = \rho_t^{\widetilde{\kappa}, h} \log h_{ft} \tag{12}$$

where,

$$\rho_t^{\widetilde{\kappa}, h} \left(a, \sigma, \kappa_{ss}, \Phi'\right) = \left(\frac{1}{\sigma} + a \right) / \left(\frac{1}{\sigma} + \frac{\kappa_{ss}}{\Phi'\left(\kappa_{ss}\right)} \right) \tag{13}$$

captures the co-movement between capital and labor utilization rates.[19] If $\rho_t^{\widetilde{\kappa}, h} = 0$, variations in labor utilization have no impact on the deviation of capital utilization from its steady state. This would imply that capital utilization is costless to vary, reflecting a flat cost profile. Alternatively, if capital utilization costs are steep (i.e., $\Phi'\left(\kappa_{ss}\right)$ is "high") then $\rho_t^{\widetilde{\kappa}, h}$ will tend to be high.

Looking at (13), confirms that the degree to which factor utilization rates co-move is a function of the wage-curvature parameter, a, as well as the substitution elasticity, σ. But their co-movement also depends on the equilibrium capital utilization rate. Thus, although equation

[19] If production is Leontief, utilization rates move in step for all a: $\sigma \to 0$, $\rho_t^{\widetilde{\kappa}, h} \to 1$; if linear then, $\sigma \to \infty$, $\rho_t^{\widetilde{\kappa}, h} \to \frac{a}{\kappa_{ss}/\Phi'(\kappa_{ss})}$; if CD, then $\sigma \to 1$, $\rho_t^{\widetilde{\kappa}, h} \to \frac{1+a}{1+\frac{\kappa_{ss}}{\Phi'(\kappa_{ss})}}$.

(12) suggests strict proportionality between h and $\widetilde{\kappa}$, this is not necessarily so since $\rho_t^{\widetilde{\kappa},h}$ is conditional on κ_{ss} which in turn depends on the real interest rate (and hence monetary policy) regime.[20]

Finally, we can derive a relationship between the overall capacity utilization rate, u_t, as the (factor-income-share) weighted average of individual rates (see Appendix C),

$$u_t - 1 \approx (1+\mu)\left[\alpha\left(\widetilde{\kappa}_t - 1\right) + (1-\alpha)\left(h_t - 1\right)\right] \tag{14}$$

This in turn implies the following relationship between labor utilization and total capacity utilization,

$$\log h_t = \frac{1}{1+\mu}\Theta_t \log u_t \tag{15}$$

where $\Theta_t = \left\{1 + \alpha\left(\rho_t^{\widetilde{\kappa},h} - 1\right)\right\}^{-1}$ is the elasticity of effective hours with respect to utilization and where $(1+\mu)$ is the increasing returns to scale factor associated with the fixed cost. If $\rho_t^{\widetilde{\kappa},h} < 1 - \frac{\mu}{\alpha(1+\mu)}$ on average then labor utilization is more volatile than overall utilization, and vice versa.

Overall utilization, u_t, whilst latent, can be uncovered in our framework as the residual of the estimated production function. Series h_t, however, can only be identified by making an assumption about $\rho_t^{\widetilde{\kappa},h}$(which we treat as a constant within a given range, see below).

4.1 The "Full" Measure of the Firm's Real Marginal Costs

Marginal costs then decompose into two components: a "conventional" component (labeled simply cmc_t^r) and an "utilization" component which captures costs associated to factor utilization (see Appendix A). Straightforwardly, the NKPC then becomes (and equivalently for the intrinsic inflation extension, (3)),

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \lambda\left(\underbrace{w_t/f_{N,t}}_{cmc_t^r} + \varphi^u \log u_t + \mu\right) \tag{16}$$

where w_t is the real wage and where it can be shown that ,

$$\varphi^u = \frac{a + \sigma^{-1} - 1}{1 + \alpha\left(\rho_t^{\widetilde{\kappa},h} - 1\right)} \tag{17}$$

Overall utilization, u_t can be recovered from the estimated production system. The drawback (see equation (17)) is that whilst one estimates φ^u, ideally one wishes to uncover both a and $\rho_t^{\widetilde{\kappa},h}$ (i.e., wage curvature and average utilization co-movement). But these are not mutually identifiable; identification of one requires prior information on the other. We take a pragmatic

[20] It can be demonstrated that, $\frac{\partial \kappa_{ss}}{\partial r} = \frac{1}{(\Phi'' - \Phi')}\frac{(1-\delta)}{(1+r)^2}$. Thus, if utilization costs are sufficiently convex, i.e., $\Phi'' > \Phi'$, an increase in the real user cost moves capital utilization towards its technical upper bound. This is intuitive: the firm reacts to higher capital costs by raising utilization, thereby reducing the need to invest in the more expensive capital stock.

approach and back out a, assuming a $[0, 1.5]$ range of $\rho_t^{\widetilde{\kappa},h}$ values.

5 Data

We use quarterly seasonally adjusted series for the US from 1953q1 to 2011q4. Our principal source was the NIPA Tables (National Income and Product Accounts) for production and income. NIPA provides Gross Domestic Product. Our capital stock series is private non residential based on accumulated NIPA investment series. For full consistency with the definition of this investment series, we therefore take public sector accounts and residential services out of GDP.

The output deflator is obtained as a ratio of nominal to constant price output. Employment is defined as a sum of self-employed persons and private sector full-time equivalent employees. As this NIPA employment series is annual, total private non-farm employees of the Bureau of Labor Statistics (Table B-1) was used to interpolate. Labor income is defined as the product of compensation to employees and labor income of self-employed workers. In evaluating the latter, compensation-per-employee is used as a shadow price of labor of self-employed workers as in Blanchard (1997) and Gollin (2002). Real capital income was calculated as a residual of the value of production excluding the aggregate mark-up and labor income, $\frac{Y}{1+\mu} - wN$, where we assume that mark-up $\mu = 0.05$ in line with several studies, although results were not sensitive to reasonable variations around that value.[21] Appendix B gives the full description on data sources and transformations.

The next section estimates the 3-equation production-technology system from which we derive conventional and utilization based measures of real marginal costs. Thereafter, we use those measures in NKPC and NKPC with intrinsic persistence, and with and without the "full" driving variable.

6 Estimations

6.1 Production and Technology

Table 1 shows results for the full-utilization CES production-technology system estimation (where $h = \widetilde{\kappa} = 1$):[22]

$$\log Y_t = \log F\left(\Gamma_t^K K_t, \Gamma_t^N N_t\right) + \varepsilon_t^Y \tag{18a}$$

$$F_K = (1+\mu)\, r_t^K + \varepsilon_t^{F_K} \tag{18b}$$

$$\log P_t = \log(1+\mu) + \log(W_t/F_N) + \varepsilon_t^P \tag{18c}$$

[21] Rotemberg and Woodford (1999) argue for very small pure profits in the US; Basu and Fernald (1997) argue that pure profits might be a couple of percent of GDP at most. As a result, most macro literature assumes that there is a markup but also a fixed cost of production.

[22] In estimation, we use a Generalized Nonlinear Least Squares (GNLLS) system estimator which is equivalent to a nonlinear SUR model allowing for cross-equation error correlation. The León-Ledesma et al. (2010b) Monte Carlo study demonstrates this estimator is able (in contrast to single-equation estimators) to identify unbiasedly both the substitution elasticity and factor augmenting technical progress parameters.

where, following (5), $F_{N|h=1} = (1-\alpha)\left(\Gamma_t^N\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{Y_t}{N_t}\right)^{\frac{1}{\sigma}}$, and equivalently $F_{K|\widetilde{\kappa}=1} = \alpha\left(\Gamma_t^K\right)^{\frac{\sigma-1}{\sigma}}\left(\frac{Y_t}{K_t}\right)^{\frac{1}{\sigma}}$. Equations (18a-18c) are consistent with the firm's maximization framework (see Appendix C) and represent the aggregated full-capacity production function and its first order profit maximization conditions with respect to capital and labor.

The table reports the substitution elasticity, σ; growth in labor and capital-augmenting technical change, γ_N, γ_K; the Box-Cox curvature parameters, λ_N, λ_K; and the p-values for the bootstrapped residual stationarity tests on ε_t^Y, $\varepsilon_t^{F_K}$, ε_t^P (all of which are stationary).[23] Although our specification implies time-varying growth in technical change, the table shows values at their sample (or normalized) average (e.g., León-Ledesma et al. (2010a)).

TABLE 1
ESTIMATES OF PRODUCTION-TECHNOLOGY SYSTEM

α	γ_N	$\gamma_{N,t>92}$	γ_K	$\gamma_{K,t>92}$	σ
0.264	0.013	0.013	0.004	-0.013	0.579
–	(0.001)	(0.002)	(0.000)	(0.002)	(0.002)
Restrictions & Tests					
		$\sigma = 1$ [0.000]			
$Hicks: \gamma_N = \gamma_K$ [0.000]		$Harrod: \gamma_N = 0$ [0.000]		$Solow: \gamma_K = 0$ [0.000]	
ADF_{ε^P} [0.024]		$ADF_{\varepsilon^{F_K}}$ [0.011]		ADF_{ε^Y} [0.022]	

Notes: Sample: 1953q1-2011q4. In terms of the Box-Cox curvature parameters, we found λ_N=0.156, λ_{N1}=1 and $\lambda_K = \lambda_{K1} \approx 0$. Robust standard errors reported are in brackets, probability values in squared brackets.

The estimated substitution elasticity (at 0.58) is significantly different from unity at 1%, and consistent with consensus aggregate values (e.g., Chirinko (2008)), as is the finding that biased technical progress is net labor saving, $\gamma_N - \gamma_K > 0$. Furthermore, all forms of conventional technical neutrality are decisively rejected.

Moreover, given the suggestions that there was (e.g., Oliner and Sichel (2000), Hansen (2001), Benati (2007)) a structural break in US labor productivity (and TFP growth) in the 1990s, we locate a break in factor-augmenting technical progress from 1992q1 onwards.[24] The evolution of (log) TFP and its decomposition into capital and labor augmenting components is shown in **Figure 2**.

[23] Given that we do not know the distribution of the statistic under the no-cointegration null, we use bootstrapped p-values following Park (2003) and Chang and Park (2003). The bootstrapped ADF-statistics are compatible with residual stationarity but that does not mean that residuals are not cross-correlated. The correlation of the residuals of the production function and the labor first-order condition is high but well below unity. This high correlation suggests that utilization measures *are* an important determinant of the marginal costs, while the deviation from a unitary correlation corroborates non-trivial price-setting frictions. Appendix C elaborates on these aspects.

[24] We dated the break point by optimizing the system log determinant across quarterly break increments from the start until the end of the sample; our detected break point accords very well with those suggested in the literature (see Benati (2007) and the references therein).

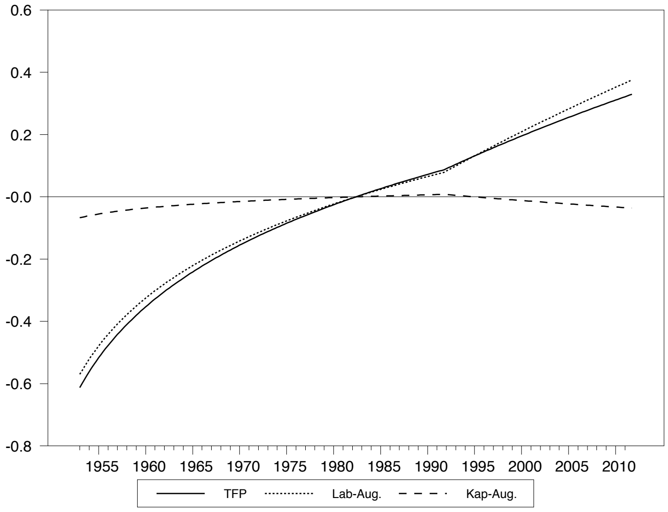

Figure 2: Log Level TFP and Decomposition into Capital and Labor Augmenting Components

These allow us to interpret US productivity growth; the 1950s were periods of exceptionally high TFP growth with a subsequent strong deceleration consistent with observed US labor productivity and growth patterns. In the early 1990s, we see a renewed acceleration of TFP growth, led by strong labor-augmenting technical change but decelerating capital-augmenting technical change. As an aside, this pattern – rising labor augmentation and falling capital augmentation – accords with the predictions of models of "biased" technical change.[25] That technical progress took off so sharply in the early 1990s accounts for a reduction in the level of real marginal costs, but, as we shall see, its volatility also reduced considerably (from the mid-1980s).

Figure 3 shows conventional real marginal costs for the CES and CD case. Both are stationary with similar (i.e., counter-cyclical) cycles. A striking difference is that (over the full sample at least) the CES variant is substantially more volatile. Another – even more striking – difference is that the CES driving variable undergoes a substantial and sustained volatility reduction from the mid-1980s onwards, consistent with observed real volatility patterns in the US economy around this time, e.g., McConnell and Perez-Quiros (2000). But the CD-based measure exhibits no such volatility change.

To illustrate, over the *Pre-Moderation* (1953q1-1983q4), *Great Moderation* (1984q1-2007q4) and *Great Recession* onwards (2007q4 onwards), the variance of the CD/labor share real marginal

[25] In such models scarcity, reflected by relative factor prices, generates incentives to invest in factor-saving innovations, i.e., firms reduce the need for scarce factors and increase the use of abundant ones. Asymptotically all technical progress must be labor augmenting. Unlike labor, capital may be accumulated limitlessly; thus labor represents the constraining factor, and, to avoid an explosion of wage income, firms bias technical improvements accordingly. Acemoglu (2002a) however suggested that while technical progress necessarily is labor-augmenting along the BGP, it may become capital-biased in transition. Given that the 1990s were periods of low unemployment (almost 2 percentage points lower than in the 1980s), it is consistent that the improvement in technical progress in the 1990s took the labor augmenting form. McAdam and Willman (2011) more fully discuss these aspects of the results.

costs was flat, around 0.3 throughout. By contrast, in the CES case it fell from 1.2 to 0.2 then back up to 0.5 (thus the Great Recession was thus 2.5 times as volatile as the Great Moderation but still only around half as volatile as the Pre-Moderation).

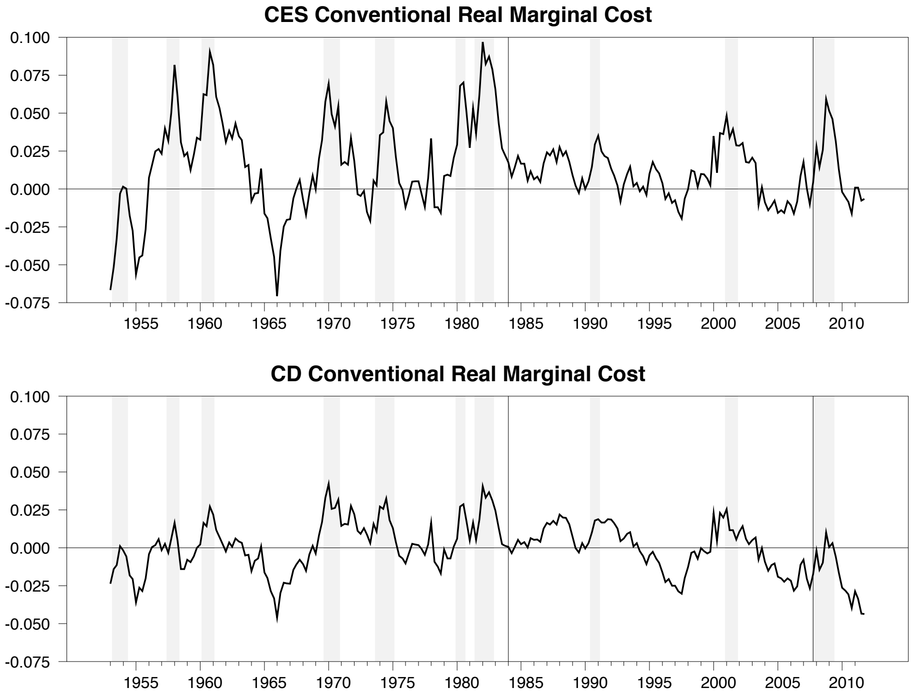

Figure 3: Conventional Real Marginal Costs (CES and CD cases)

NOTES: Shaded areas represent recessions as identified by the NBER. This figure plots the conventional measure of real marginal cost for the CES and CD cases, as indicated by equation set (5). The vertical lines denote the Great Moderation and Great Recession periods.

What is behind this difference in volatility patterns? Recalling equation (5), the variances of conventional real marginal costs can be decomposed in the following stationary manner:

$$var_{w-g_N} + \frac{1}{\sigma^2} var_{y-n-g_N} - \frac{2}{\sigma} cov_{w-g_N, y-n-g_N} \tag{19}$$

where var_{w-g_N}, var_{y-n-g_N} are the variance of the log deviation of the real wage rate and labor productivity respectively from labor-augmenting technical change. The substantially higher relative weight given to the latter in the CES measure (i.e., $1/0.58^2 \approx 3$) ensures that precisely that component which decreased the most attracts the higher weight.[26]

Thus, the CES-based real conventional marginal cost measure replicates relatively better the observed volatility reduction otherwise witnessed in many US time series. This is an important observation since, by contrast, some discussion suggested that reduction in US inflation volatility in recent years had not been matched by that in candidate driving variables, e.g. Fuhrer (2010).

[26] This need not contradict other popular explanations for the reductions in inflation volatility such as improved monetary policy or fewer shocks, since these will impact the evolution of factor prices and factor accumulation.

6.2 NKPC Estimations

Table 2 presents the NKPC results over 1953q1-2011q4. It shows the Calvo parameter, θ, and solves for the composite slope parameter, λ, and price stickiness duration, $D = 1/(1-\theta)$. We show results with the conventional (CES) driving variable and with the "full" driving variable (the latter thus adding parameter φ^u as in (16)). From $\widehat{\varphi^u}$, we can back out the wage-curvature parameter a (exploiting equation (17)), assuming $\rho_t^{\widetilde{\kappa},h} \in [0, 1.5]$.

We use the non-linear GMM-CUE (continuously updated) estimator[27] and, for brevity, restrict discounting to $\beta = 0.99$ (this is theoretically well-founded and empirically innocuous). Regarding instruments, we tried to be parsimonious, although paralleling Galí and Gertler (1999).[28] We used: first lag of inflation; first and second lag of conventional marginal cost; first lag of utilization; second, third and fourth lag of interest rate spread (the difference of the 5-year and 3-month Treasury Bond yields, denoted $\widetilde{r}_t$); third and fourth lag of hourly compensation growth rates (π_t^w); fourth lag of oil price inflation (π_t^o). Probability-value for the Hansen's J statistic of the over-identifying restrictions are in squared brackets.

All parameters are significant and correctly signed. Where the conventional driving variable is used (the first row), we uncover relatively high (and counter-factual) price stickiness (9 quarters). We also find seemingly modest (though statistically significant) values of the slope parameter (0.015). These results accord with those of much of the literature (in fact we improve upon matters since λ is significant).

However, when the full marginal cost measure is used, duration estimates reduce to around 6 quarters, since $\widehat{\theta}$ falls to 0.82. The point estimate of φ^h at 0.75 is significant (and insignificantly different from 1), but only weakly uncovers a wage curvature parameter range.[29] Given the reduction in θ , the slope more than doubles. Finally, that $\widehat{\varphi^u} < 1$, suggests, at least to a first approximation, that the full measure of real marginal costs is counter-cyclical.

[27] Among Generalized Empirical Likelihood (GEL) estimators, CUE estimation has good finite-sample properties and is more efficient, Anatolyev (2005). Two-step GMM methods, by contrast, can display poor small-sample properties, e.g., Hansen et al. (1996) and are not invariant to the specification of the moment conditions (see Gabriel and Martins (2009)).

[28] The (still common) practice of using too many instruments and too general corrections for serial correlation seriously impairs the power of the J test in finite samples and biases GMM results in favor of least squares, Newey and Windmeijer (2009). Galí and Gertler (1999) used over 20 instruments and a 12 lag Newey-West estimate of the covariance matrix.

[29] The statutory overtime premium is 50% for covered employees although it has been suggested that the effective rate is around 25-30% (Trejo (1991), Hamermesh (2006)).

TABLE 2
ESTIMATES OF THE NKPC

θ	φ^u	λ	D	$a^{1)}$	J
0.890	–	0.015	9.1	–	[0.560]
(0.016)		(0.004)	(1.3)		
0.821	$0.749^{2)}$	0.041	5.6	-0.2::0.1	[0.588]
(0.032)	(0.187)	(0.018)	(1.1)		

NOTES: Sample: 1953q1-2011q4. 1) Predicated on $\rho_t^{\widetilde{\kappa},h} \in [0, 1.5]$, the wage curvature parameter is backed out from equation (17) as $a = \widehat{\varphi^u}\left[1 + \alpha\left(\rho_t^{\widetilde{\kappa},h} - 1\right)\right] - \widehat{\sigma}^{-1} + 1$, where from Table 1, $\alpha = 0.26$. 2) Wald tests: $\underset{[0.19]}{\varphi^u = 1}$

Instruments used: $Z_t = \pi_{t-1},\ cmc_{t-1}^r,\ cmc_{t-2}^r,\ u_{t-1},\ \widetilde{r}_{t-2},\ \widetilde{r}_{t-3},\ \widetilde{r}_{t-4},\ \pi_{t-3}^w,\ \pi_{t-4}^w,\ \pi_{t-4}^o$. We used a HAC weighting matrix, a Quadratic (Andrews) kernel, bandwidth: 4.

6.3 NKPC with Intrinsic Persistence

When the conventional driving variable is used (**Table 3**), we find a high share of forward-looking price setting, $\gamma_f = 0.64$, but relatively modest (though significant) slope estimates and long fixed price duration estimates ($\approx$ 6 quarters). The share of backward-looking price re-setters is around 45%. This constellation is also a common finding in the literature, e.g., Galí and Gertler (1999), table 2; Tsoukis et al. (2011) (although with our proviso of a significant slope).

When we introduce full marginal costs, a more balanced weighting of backward and forward-looking price setting emerges (around 0.5 each), and at 3 quarters, sticky-price estimates become more aligned with micro price-setting studies (e.g., Bils and Klenow (2004)). Although, the slope coefficient, λ, is largely unchanged, the reduction in duration comes from a combination of low price stickiness but relatively high inertial price resetting schemes.

Given that $\widehat{\varphi^u} = 1.29$, this would imply premium curvature parameters $(0.2 - 0.7)$ within the range of statutory overtime rates, and is suggestive of a net pro-cyclical driving variable.[30]

[30] Although not our primary concern, we also considered incremental forecasting power. We ran three ARDL models over the full sample period,

$$\pi_t = \alpha_0 + \alpha_i \sum_{j=1}^{2} \pi_{t-j} + \beta_i \sum_{j=1}^{2} cmc_{t-j} + \gamma_i \sum_{j=1}^{2} fmc_{t-j}$$

where $fmc_t = cmc_t + \varphi^u \log u_t + \mu$, with the nested models, $\mathbb{M}_1 : \beta_i = \gamma_i = 0 \forall i$, $\mathbb{M}_2 : \gamma_i = 0 \forall i$, $\mathbb{M}_3 : \beta_i = 0 \forall i$. Having estimated them, we compared their fit with RMSE criterion. Setting $RMSE^{\mathbb{M}_1} = 1$ we found $RMSE^{\mathbb{M}_2} = 0.98294$ and $RMSE^{\mathbb{M}_2} = 0.93804$. Whilst the addition of conventional marginal costs (cmc) has some (around 1.3%) improvement in forecasting power over the full sample, that of full real marginal (fmc) gave around a 7% improvement.

TABLE 3
ESTIMATES OF THE NKPC WITH INTRINSIC PERSISTENCE

θ	ω	φ^u	γ_f	γ_b	λ	D	a	J
0.824	0.456	–	0.639	0.357	0.014	5.7	–	[0.504]
(0.023)	(0.038)		(0.006)	(0.015)	(0.005)	(0.7)		
0.692	0.799	1.294	0.461	0.538	0.013	3.2	0.2::0.7	[0.157]
(0.145)	(0.123)	(0.303)	(0.052)	(0.087)	(0.007)	(1.5)		

NOTES: See notes to Table 2.

6.4 Tracking Inflation: A Graphical Analysis of Real Marginal Costs

The upper panels of **Figure 3** plot the conventional real marginal costs and utilization component. Reflecting the cyclicality differences, the correlation between components is mostly negative (although well below unity). This accords with our discussion on the residuals of our production-technology system (18a)-(18c). Under frictionless price setting this correlation would be -1, because in that case the only reason why the mark-up over conventional nominal marginal costs do not exactly equal the price level is the unobserved variation of intensities at which recruited employment is utilized. Frictions in price setting decreases this correlation below unity.

The bottom panel plots the full measure of CES-based real marginal costs (i.e., $fmc_t = cmc_t + \varphi^u \log u_t + \mu$) and annualized quarterly inflation.[31] To help interpret matters, consider re-writing the intrinsic NKPC case as,

$$\pi_t - \gamma_f \mathbb{E}_t \pi_{t+1} - \gamma_b \pi_{t-1} = \lambda \underbrace{\left(\underbrace{cmc_t^n + \mu - p_t}_{gap_1} + \underbrace{\varphi^u \log u_t}_{gap_2} \right)}_{fmc_t} \tag{20}$$

The rhs of (20) presents the full real marginal cost as a sum of two gaps. Gap_1 is the deviation of conventional nominal marginal costs cmc_t^n plus the markup from the price level; gap_2 is the weighted utilization component (itself, by definition, a gap concept).

If the sum of these gaps is positive there will be upward price pressures – i.e., inflation (and possibly also expected inflation) will be accelerating. If around zero, then inflation will be stable (though not necessarily zero). From Table 3 we see that since $\widehat{\gamma_f} + \widehat{\gamma_b} \approx 1$ both sides of (20) can be zero at any level of stable inflation.

The full measure fits the profile of inflation well since deviations of full real marginal costs from zero coincide well with the periods of accelerating, decelerating or stable inflation. For instance, from the early 1960s to mid1970s, we see a long period of positively trending fmc_t, with attendant accelerating inflation. Contributions wise, it is essentially demand-push turning into cost-push inflation in the recession of 1969q4 1970q4.

[31]For presentational purposes the latter is HP-filtered to remove high frequency components.

Likewise, the inflation spike of the 1973q4-1975q1 recession was cost-push based. Simultaneously the exceptionally low utilization rate resulted in the steep drop of fmc_t evoking strongly decelerated inflation. Thereafter, although both real marginal cost components of around zero, fmc_t (as their sum) rises clearly above zero coinciding with a period of accelerating inflation.

In the Great Moderation, fmc_t is not only low but also less volatile than before. This is driven by volatility reduction in *both* components. In turn, these are properties inherited by inflation itself.

Finally, in the Great Recession we see $fmc_t < 0$. However, although overall inflation has remained modest we see only a modest temporary deceleration (apparently less than would be suggested by the decline in fmc_t). A closer examination of the components show that the conventional component is close to zero (consistent with a stable inflation). The utilization component, although also recovering strongly towards the end of our sample, remains subdued.[32]

As regards the cyclical properties of the real marginal costs, the full measure turns from counter- (when measured conventionally) to pro-cyclical reflecting the dominance of (pro-cyclical) utilization rates. We also observe that the decrease of cyclicality of fmc_t is less visible than in its conventional and utilization components. In many cases (e.g., recessionary cycles 1957q3-1958q2, 1973q4-1975q1 and 1981q3-1982q4) where our preferred measure lags the cycle, unlike its components. We also estimated by recursive least squares the full, conventional and utilization component of marginal costs on a constant and the CBO output-gap measure.[33] We found that the individual components were becoming absolutely less cyclical, although interestingly they compensate each other at the Full measure.

[32] Possibly, our utilization measure overestimate the actual amount of labor hoarding in the economy. This may reflect difficulties to correctly identify the contributions of technology and factor utilizations on the markedly declined growth of production just in the end of estimation period and before its proper recovery to a more normal path.

[33] Results and recursive coefficient graphics are available on request.

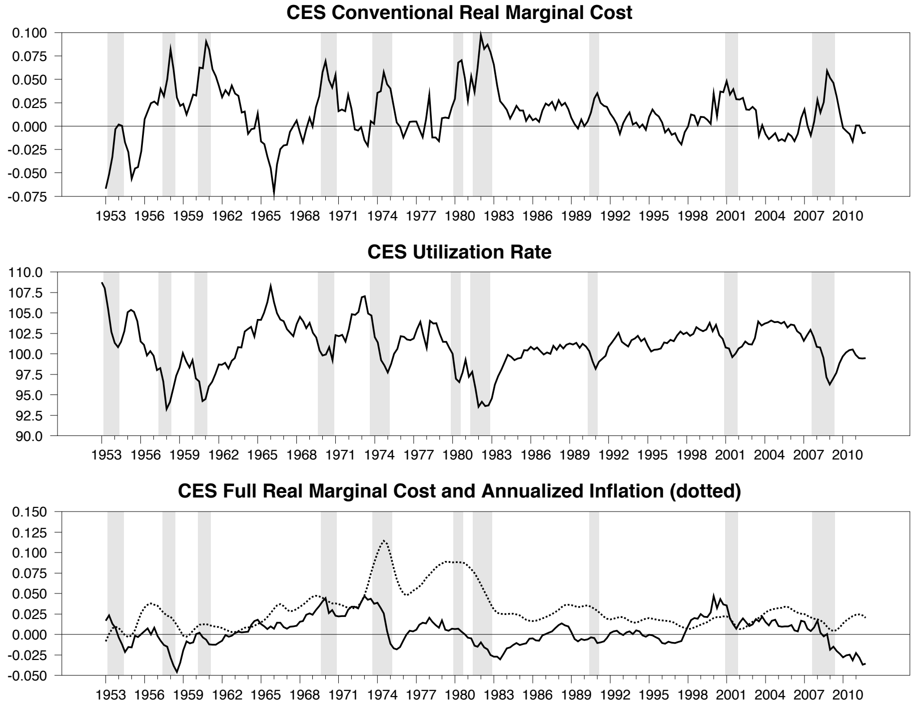

Figure 4: Components of Real Marginal Costs; the Full Measure and Filtered Inflation

NOTES: Shaded areas represent recessions as identified by the NBER. This figure plots the full and conventional measure of real marginal cost, as in equation (16), where the conventional case implying $\varphi^u = 0$. In the middle panel, for presentational purposes, the utilization rate is indexed to 100 at 1980q1. The final panel shows fmc_t as against filtered inflation.

7 Conclusions

What drives inflation matters for understanding inflation and for setting policy. We argued that real marginal costs measures routinely used in the literature (and routinely adapted into policy models) are flawed. Following Rotemberg and Woodford (1999), we constructed richer measures and reappraised the sensitivity of inflation specifications. We conclude:

1. The use of CD (and hence labor share as the driving variable) is unsuitable given its empirical rejection. Its use also suppresses aspects key to understanding cost margins: non-neutral technical change and non-unitary factor substitutability. By contrast, CES-based real marginal costs provide a more intuitive lens through which to track marginal costs. They are also able to match the recent volatility and level-shift patterns witnessed over time in many US time series.

2. Both conventional measures (CD or CES), though, are counter-cyclical – an inconvenient aspect of which, is that anti-inflation policies would operate pro-cyclically. Conventional measures do not account for variations in factor utilization rates. To disentangle technical progress from utilization, we modeled the former as a factor-augmenting and smoothly-evolving process. And we introduced a parametric form of "effective labor hours" to

capture overtime/hoarding costs. Allowance was also made for co-variation of utilization rates.

3. We constructed a full real marginal cost measure comprising counter-cyclical costs excluding utilization, plus pro-cyclical utilization rates. The *net* cyclicality of the full measure is then an empirical matter, dependent on the prevalence of demand and supply influences and the data weighting. Moreover, since utilization rates mimic an output gap, our "full" measure contributes to the emerging belief that Phillips curve approaches that merge new and old elements are helpful in accounting for inflation (i.e., Blanchard and Galí (2007, 2010)). The good tracking performance of the full marginal cost measure was clear, and, in terms of cost-push and demand-pull elements, provided insights into inflation developments.

4. Mis-specification of the driving variable is costly; failure to account for non-unitary factor substitution, non-neutral technical change, and factor utilization rates in driving variable biases upward the contribution of extrinsic inflation persistence, and exaggerates fixed-price contract lengths. Our results thus lend weight to a more balanced perspective on historical inflation dynamics, and are better reconciled with micro measures of price stickiness.

All Phillips curves (new and old) are driven by some real-activity measure. Richer, more plausible specifications for that driving variable contribute to better estimation across the board (and, one presumes, better policy making). The benchmarking of our results with others in the more recent literature (including forecasting performance) appears therefore a natural way to proceed. Incorporation of our driving variable into inflation equations embedded into general-equilibrium policy models is also an attractive research extension.

References

Acemoglu, D. (2002a). Directed technical change. *Review of Economic Studies*, 69:781–809.

Acemoglu, D. (2002b). Technical Change, Inequality and the Labor Market. *Journal of Economic Literature*, 40(1):7–72.

Anatolyev, S. (2005). GMM, GEL, serial correlation and asymptotic bias. *Econometrica*, 73:983–1002.

Barkume, A. J. (2007). Some New Evidence on Overtime Use, Total Job Compensation, and Wage Rates. Bureau of Labor Statistics Working Paper 402.

Barro, R. J. and Sala-i-Martin, X. (2004). *Economic Growth.* 2nd Edition, MIT Press, Cambridge, MA.

Basu, S. and Fernald, J. G. (1997). Returns to Scale in U.S. Manufacturing: Estimates and Implications. *Journal of Political Economy*, 105(2):249–283.

Basu, S., Fernald, J. G., and Kimball, M. S. (2006). Are Technology Improvements Contractionary? *American Economic Review*, 96(5):1418–1448.

Batini, N. (2009). Euro area inflation persistence. *Empirical Economics*, 31(4):977–1002.

Benati, L. (2007). Drift and breaks in labor productivity. *Journal of Economic Dynamics and Control*, 31(8):2847–2877.

Bils, M. (1987). The Cyclical Behavior of Marginal Cost and Price. *American Economic Review*, 77:838–855.

Bils, M. and Klenow, P. J. (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5):947–985.

Blanchard, O. J. (1997). The Medium Run. *Brookings Papers on Economic Activity*, 2:89–158.

Blanchard, O. J. and Galí, J. (2007). Real Wage Rigidities and the New Keynesian Model. *Journal of Money, Credit and Banking*, 39(1):35–66.

Blanchard, O. J. and Galí, J. (2010). Labor Market Frictions and Monetary Policy: A New Keynesian Model with Unemployment. *American Economic Journal: Macroeconomics*, 2(2):1–30.

Burnside, C. and Eichenbaum, M. (1996). Factor-hoarding and the propagation of business-cycle shocks. *American Economic Review*, 86(5):1154–74.

Chang, Y. and Park, J. Y. (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis*, 24:379–400.

Chirinko, R. S. (2008). σ: The Long and Short of It. *Journal of Macroeconomics*, 30(2):671–686.

Chirinko, R. S. and Fazzari, S. (1994). Economic fluctuations, market power, and returns to scale: Evidence from firm-level data. *Journal of Applied Econometrics*, 9(1):47–69.

Cooper, R. and Willis, J. L. (2009). The Cost of Labor Adjustment: Inferences from the Gap. *Review of Economic Dynamics*, 12(4):632–647.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–76.

Fuhrer, J. C. (1997). The (un)importance of forward-looking behavior in price specifications. *Journal of Money, Credit and Banking*, 29:338–350.

Fuhrer, J. C. (2010). Inflation persistence. In Friedman, B. M. and Woodford, M., editors, *Handbook of Monetary Economics*, volume 3 of *Handbook of Monetary Economics*, chapter 9, pages 423–486. Elsevier.

Gabriel, V. J. and Martins, L. F. (2009). New Keynesian Phillips Curves and potential identification failures: A Generalized Empirical Likelihood analysis. *Journal of Macroeconomics*, 31(4):561–571.

Gagnon, E. and Khan, H. (2005). New Phillips curve under alternative production technologies for Canada, the United States, and the Euro area. *European Economic Review*, 49(6):1571–1602.

Galí, J. and Gertler, M. (1999). Inflation Dynamics: A Structural Econometric Analysis. *Journal of Monetary Economics*, 44:195–222.

Galí, J., Gertler, M., and López-Salido, J. D. (2007). Markups, gaps, and the welfare costs of business fluctuations. *Review of Economics and Statistics*, 89(1):44–59.

Gollin, D. (2002). Getting Income Shares Right. *Journal of Political Economy*, 110(2):458–474.

Granger, C. W. J. (1983). Co-Integrated Variables and Error-Correcting Models. UCSD Discussion Paper 83-13.

Hamermesh, D. S. (2006). Overtime Laws and the Margins of Work Timing. In Askenazy, P., Carton, D., de Coninck, F., and Gollac, M., editors, *Organisation et Intensit du Travail*. Paris: Octares.

Hansen, B. E. (2001). The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity. *Journal of Economic Perspectives*, 15(4):117–128.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, 14:262–280.

Hart, R. A. (2004). *The Economics of Overtime Work*. Cambridge University Press.

Jones, C. I. (2003). Growth, Capital Shares, and a New Perspective on Production Functions. Mimeo, University of California Berkeley.

Kinoshita, T. (1987). Working Hours and Hedonic Wages in the Market Equilibrium. *Journal of Political Economy*, 95(6):1262–77.

Klump, R., McAdam, P., and Willman, A. (2007). Factor Substitution and Factor Augmenting Technical Progress in the US. *Review of Economics and Statistics*, 89(1):183–92.

León-Ledesma, M. A., McAdam, P., and Willman, A. (2010a). Identifying the Elasticity of Substitution with Biased Technical Change. *American Economic Review*, 100(4):1330–1357.

León-Ledesma, M. A., McAdam, P., and Willman, A. (2010b). In Dubio pro CES: Supply Estimation with Mis-Specified Technical Change. Working Paper No. 1175, European Central Bank.

Madeira, J. (2013). Overtime Labor, Employment Frictions and the New Keynesian Phillips Curves. *Review of Economics and Statistics*, (forthcoming).

Mazumder, S. (2010). The New Keynesian Phillips curve and the cyclicality of marginal cost. *Journal of Macroeconomics*, 32(3):747–765.

McAdam, P. and Willman, A. (2011). Technology, Utilization and Inflation: What drives the New Keynesian Phillips Curve? Working paper series, European Central Bank.

McAdam, P. and Willman, A. (2013). Medium Run Redux. *Macroeconomic Dynamics*, (forthcoming).

McConnell, M. and Perez-Quiros, G. (2000). Output Fluctuations in the United States: What has Changed since the Early 80s? *American Economic Review*, 90(5):1464–1476.

Nekarda, C. J. and Ramey, V. A. (2010). The Cyclical Behavior of the Price-Cost Markup. mimeo.

Newey, W. K. and Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719.

Oliner, S. D. and Sichel, D. E. (2000). The Resurgence of Growth in the Late 1990s: Is Information Technology the Story? *Journal of Economic Perspectives*, 14(4):3–22.

Park, J. Y. (2003). Bootstrap unit root tests. *Econometrica*, 71:1845–1895.

Roberts, J. M. (1995). New Keynesian Economics and the Phillips Curve. *Journal of Money, Credit and Banking*, 27(4):975–984.

Rogerson, R. (1988). Indivisible Labor, Lotteries and Equilibrium. *Journal of Monetary Economics*, 21:3–16.

Rotemberg, J. J. and Woodford, M. (1999). The Cyclical Behavior of Prices and Costs. In Taylor, J. and Woodford, M., editors, *Handbook of Macroeconomics, Vol. 1B*. Elsevier.

Rudd, J. and Whelan, K. (2007). Modeling Inflation Dynamics: A Critical Survey of Recent Research. *Journal of Money Credit and Banking*, 39:155–170.

Shapiro, M. D. (1986). The Dynamic Demand for Capital and Labor. *Quarterly Journal of Economics*, 101:513–542.

Solow, R. M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics*, 39(3):312–320.

Stock, J. H. (1987). Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica*, 55(5):1035–56.

Trejo, S. J. (1991). The Effects of Overtime Pay Regulation on Worker Compensation. *American Economic Review*, 81(2):719–740.

Tsoukis, C., Kapetanios, G., and Pearlman, J. (2011). The Elusive Persistence: Wage and Price Rigidities, the Phillips Curve, and Inflation Dynamics. *Journal of Economic Surveys*, 25(4):737–768.

Varian, H. (1974). A Bayesian approach to real estate assessment. In Feinberg, S. E. and Zellner, A., editors, *Studies in Bayesian Economics in Honour of L. J. Savage*, pages 195–208. Amsterdam: North Holland.

APPENDICES

A Section 4 Proofs

A.1 Proof of Equations (11) and (12)

Let us first solve the firm's profit maximizing problem in the absence of any frictions in price setting. This allows us to define real marginal costs also capturing the costs resulting from time-varying factor utilization rates. In addition, the first-order conditions of profit maximization gives us the equilibrium system used in estimating the parameters of the production-technology system needed for constructing real marginal costs. Assume firm $f$faces demand function $Y_{ft} = \left(\frac{P_{ft}}{P_t}\right)^{-\varepsilon} Y_t$. Its profit function is

$$\Pi_t = P_t \left\{ \begin{array}{c} Y_{ft}^{1-\frac{1}{\varepsilon}} Y_t^{\frac{1}{\varepsilon}} - \frac{W(\overline{W}_t, h_{ft})}{P_t} N_{ft} - \frac{\overline{W}_t}{P_t} \Omega_N (N_{ft}, N_{ft-1}) - I_{ft} \\ -\Phi(\kappa_{ft}) K_{ft} - \Omega_K (K_{ft}, K_{ft-1}, K_{ft-2}) - (1+i_{t-1}) \frac{P_{t-1}}{P_t} b_{ft-1} + b_{ft} \end{array} \right\} \tag{A.1}$$

where $\Omega_j$refers to an adjustment cost function associated to factor $j = N,\ K,\ \delta \in (0,1)$ is the depreciation rate, i denotes the nominal interest rate, and b_{ft} denotes a one-period real corporate bond reflecting the possibility of external finance for the firm. The firm maximizes the discounted sum of profits, subject to its production constraints,

$$\max \sum_{s=t}^{\infty} \prod_{j=0}^{s-t} R_j \left\{ \Pi_s + P_s \Lambda_{fs}^Y \left[(1+\mu) F\left(\Gamma_s^K \widetilde{\kappa}_{fs} K_{fs}, \Gamma_s^N h_{fs} N_{fs} \right) - \chi_{fs} - Y_{fs} \right] \right\} \tag{A.2}$$

where $\widetilde{\kappa}_{ft} = \frac{\kappa_{ft}}{\kappa_{ss}}$ (κ_{ss}being the equilibrium utilization rate). The first-order conditions are:

$$Y_f : \Lambda_{ft}^Y = \frac{P_{ft}}{(1+\mu) P_t} \tag{A.3}$$

$$\kappa_f : \Lambda_{ft}^Y = \frac{\Phi'(\kappa_{ft})}{(1+\mu) F_{K_f}} \widetilde{\kappa}_{ft} \tag{A.4}$$

$$h_f : \Lambda_{ft}^Y = \frac{W_{h_f}}{P_t (1+\mu) F_{N_f}} h_{ft} \tag{A.5}$$

$$N_f : \frac{\partial \Omega_N \left(N_{ft}, N_{ft-1} \right)}{\partial N_{ft}} + \mathbb{E}_t \left\{ R_{t+1} \frac{\overline{W}_{t+1}}{\overline{W}_t} \frac{\partial \Omega_N \left(N_{ft+1}, N_{ft} \right)}{\partial N_{ft}} \right\} = \frac{P_t}{\overline{W}_t} \Lambda_{ft}^Y (1+\mu) F_{N_f} - \frac{W\left(\overline{W}_t, h_{ft} \right)}{\overline{W}_t} \tag{A.6}$$

$$b_f : \mathbb{E}_t R_{t+1} = \frac{1}{1+i_t} \tag{A.7}$$

$$K_f : \frac{\partial \Omega_K (K_{ft}, K_{ft-1}, ..)}{\partial K_{ft}} + \mathbb{E}_t \left\{ R_{t+1} \frac{P_{t+1}}{P_t} \frac{\partial \Omega_K (K_{ft+1}, ..)}{\partial K_{ft}} \right\} + \mathbb{E}_t \left\{ R_{t+1} \frac{P_{t+1}}{P_t} R_{t+2} \frac{P_{t+2}}{P_{t+1}} \frac{\Omega_K (K_{ft+2}, ..)}{\partial K_{ft}} \right\}$$

$$= \frac{P_{ft}}{P_t} F_{K_f} - \left(1 - \mathbb{E}_t \left\{ R_{t+1} \frac{P_{t+1}}{P_t} (1-\delta) \right\} + \Phi(\kappa_{ft}) \right) \tag{A.8}$$

$$\Lambda_f^Y : Y_{ft} = (1+\mu) F\left(\Gamma_t^K \widetilde{\kappa}_{ft} K_{ft}, \Gamma_t^N h_{ft} N_{ft} \right) - \chi_{ft} \tag{A.9}$$

$1+\mu = \frac{\varepsilon}{\varepsilon - 1}$ represents the equilibrium mark-up of prices over costs, $F_{K_f} = \frac{\partial F}{\partial \left(\Gamma_s^K \widetilde{\kappa}_{fs} K_{fs} \right)} \Gamma_t^K \widetilde{\kappa}_{ft}$ and $F_{N_f} = \frac{\partial F}{\partial \left(\Gamma_s^N h_{fs} N_{fs} \right)} \Gamma_t^N h_{ft}$.From (6) in the main text we note the derivative $W_{h_f} = \overline{W}_t \left(1 + a \left(h_{ft} - 1 \right) \right)$. Conditions (A.3-A.5) define the shadow price (or marginal cost) of output. Conditions (A.4) and (A.5) further highlight that an optimizing firm would equalize the marginal cost of raising output across all factor margins. Conditions (A.6) and (A.8) define dynamic demands for the number of employees and capital, (A.7) defines the discount factor and (A.9) retrieves the production function. Given (A.7), the inverse of gross real interest rate is,

$$(1+r_t)^{-1} = \mathbb{E}_t\left\{R_{t+1}\frac{P_{t+1}}{P_t}\right\} = \frac{1+\mathbb{E}_t\pi_{t+1}}{1+i_t} \tag{A.10}$$

where π denotes inflation. Conditions (A.7) and (A.8) solve for the firm's real user cost of capital, r_{ft}^K,

$$r_{ft}^K = \underbrace{\frac{r_t+\delta}{1+r_t} + \Phi\left(\kappa_{ss}\right)}_{r_t^K} + \left[\Phi\left(\kappa_{ft}\right) - \Phi\left(\kappa_{ss}\right)\right] \tag{A.11}$$

where r_t^K is the equilibrium component common to all firms. Equations (A.3)-(A.5) imply,

$$\frac{\Phi'\left(\kappa_{ft}\right)}{F_{K_f}}\widetilde{\kappa}_{ft} = \frac{w_t\left[1+a\left(h_{ft}-1\right)\right]}{F_{N_f}}h_{ft} \tag{A.12}$$

where $w_t = W_t/P_t$ is the real wage. As regards marginal productivities F_{K_f} and F_{N_f} consider their behavior in equilibrium, i.e., $h_{ft} = \widetilde{\kappa}_{ft} = 1$. Now (A.6), (A.8) and (A.11) implies that $\frac{F_{N_f|h_f=1}}{F_{K_f|\widetilde{\kappa}_f=1}} = \frac{\overline{w_t}}{r_t^K}$hold in the full-capacity equilibrium. The CES production function can then be shown to imply the following marginal product conditions,

$$F_{N_f} = F_{N|h=1}\cdot h_{ft}^{\frac{\sigma-1}{\sigma}} = \overline{w_t}h_{ft}^{\frac{\sigma-1}{\sigma}} \tag{A.13}$$

$$F_{K_f} = F_{K|\widetilde{\kappa}=1}\cdot \widetilde{\kappa}_{ft}^{\frac{\sigma-1}{\sigma}} = r_t^K\widetilde{\kappa}_{ft}^{\frac{\sigma-1}{\sigma}} \tag{A.14}$$

That is to say, the firm's total marginal product equals its marginal product when the factor is fully utilized times a term in the utilization rate itself. Inserting (A.13) and (A.14) into (A.12) yields,

$$\widetilde{\kappa}_{ft}^{\frac{1}{\sigma}}\Phi'\left(\kappa_{ft}\right) = r_t^K\left[1+a\left(h_{ft}-1\right)\right]h_{ft}^{\frac{1}{\sigma}} \tag{A.15}$$

Equation (A.15) defines the relationship between capital and labor utilization rates. A closed-form is obtained after applying a first-order Taylor approximation to $\log\left[\Phi'\left(\kappa_{ft}\right)\right] \approx \log\left[\Phi'\left(\kappa_{ss}\right)\right] + \frac{\kappa_{ss}}{\Phi'\left(\kappa_{ss}\right)}\log\widetilde{\kappa}_{ft}$ and to $\log\left[1+a\left(h_{ft}-1\right)\right] \approx a\log h_{ft}$. After observing that the full-capacity equilibrium form of (A.12) gives $\Phi'\left(\kappa_{ss}\right) = r_t^K$,equation (A.15) becomes,

$$\log\widetilde{\kappa}_{ft} = \rho_t^{\widetilde{\kappa},h}\log h_{ft} \tag{A.16}$$

where

$$\rho_t^{\widetilde{\kappa},h} = \left(\frac{1}{\sigma}+a\right)/\left(\frac{1}{\sigma}+\frac{\kappa_{ss}}{\Phi'\left(\kappa_{ss}\right)}\right) \tag{A.17}$$

A.2 Proof of Footnote 20

Take the steady state of (A.15), which, together with (A.11), implies,

$$\Phi'\left(\kappa_{ss}\right) - \Phi\left(\kappa_{ss}\right) - \frac{r_t+\delta}{1+r_t} = 0 \tag{A.18}$$

then differentiate with respect to r_t:

$$\frac{\partial \kappa_{ss}}{\partial r} = \frac{1}{(\Phi'' - \Phi')} \frac{(1-\delta)}{(1+r)^2} > 0 \Leftrightarrow \Phi'' > \Phi' \tag{A.19}$$

Thus, an increase in the real interest rate raises equilibrium utilization closer to its technical upper bound if capital utilization costs are sufficiently convex. This partly reduces further the need to invest in the more expensive capital stock.

A.3 Proof of Equations (14) and (15)

Equations (A.6), (A.8) and (A.11) imply that in the full-capacity equilibrium $\frac{F_{N_f|h_f=1}}{F_{K_f|\widetilde{\kappa}_f=1}} = \frac{F_{N|h_f=1}}{F_{K|\widetilde{\kappa}_f=1}} = \frac{\overline{w_t}}{r_t^K}$. Further with the properties of the homogenous production function this implies that capital intensities, on one hand, and marginal productivities of labor and capital, on the other hand, are equal across firms. Now, under the assumption that the fixed costs of firms are proportional to full capacity aggregate output $Y_t^*$and that the sum $\sum \chi_{ft} = \mu Y_t^*$, the aggregation of the firm level full-capacity output results in the following homogenous full-capacity aggregate production function,

$$Y_t^* = \sum Y_{ft}^* = (1+\mu) \sum F\left(\Gamma_t^K K_{ft}, \Gamma_s^N N_{ft}\right) - \mu Y_t^* = F\left(\Gamma_t^K K_t, \Gamma_t^N N_t\right) \tag{A.20}$$

where $K_t = \sum K_{ft}$and $N_t = \sum N_{ft}$. After defining $\widetilde{\kappa}_t = \sum \frac{K_{ft}}{K_t} \widetilde{\kappa}_{ft}$ (recall $\widetilde{\kappa}_{ft} = \kappa_{ft} / \kappa_{ss}$) and $h_t = \sum \frac{N_{ft}}{N_t} h_{ft}$ the aggregate counterpart of the firm level production function can be written as an expansion of (A.20),

$$Y_t = (1+\mu) F\left(\Gamma_t^K \widetilde{\kappa}_t K_t, \Gamma_s^N h_t N_{fs}\right) - \mu Y_t^* \tag{A.21}$$

Taking a first-order approximation of (A.21) around $\widetilde{\kappa}_t = h_t = 1$ yields the overall capacity utilization rate, u_t,

$$\underbrace{\log \frac{Y_t}{Y_t^*}}_{u_t \;-\; 1} \approx \underbrace{\frac{\Gamma_t^K K_t}{Y_t^*} \frac{\partial Y_t^*}{\partial\left(\Gamma_t^K K_t\right)}}_{(1+\mu)\alpha} (\widetilde{\kappa}_t - 1) + \underbrace{\frac{\Gamma_t^N N_t}{Y_t^*} \frac{\partial Y_t^*}{\partial\left(\Gamma_t^N N_t\right)}}_{(1+\mu)(1-\alpha)} (h_t - 1) \tag{A.22}$$

which is given by the (factor-income-share) weighted average of factor utilization rates. Under CD and CES with Harrod neutrality, approximation (A.22) is exact. The quality of the approximation is also relatively good under factor-augmenting technical progress unless factor income shares contain very strong trends. Substituting the aggregate counterpart of (A.16) into (A.22), we further derive a relationship between labor utilization and total capacity utilization,

$$\log h_t = \frac{1}{(1+\mu)\left[1 + \alpha\left(\rho_t^{\widetilde{\kappa},h} - 1\right)\right]} \log u_t = \frac{\Theta}{1+\mu} \log u_t \tag{A.23}$$

A.4 Frictionless Price Level and Full Marginal costs, Proofs of Section (4.1)

Equations (A.3), (A.5) and (A.13) imply that in the absence of friction the in price setting the firm's optimal log reset price p_{ft}^* is,

$$p_{ft}^* = \log \frac{\overline{W}_t}{F_{N|h=1}} + \underbrace{\log\left[1 + a\left(h_{ft} - 1\right)\right] + \frac{1}{\sigma}\log h_{ft}}_{\approx\left(a+\frac{1}{\sigma}\right)\log h_{ft}} + \mu \tag{A.24}$$

Define the aggregated optimal frictionless price level as weighted average of individual goods prices with output shares, $s_{ft} = \frac{Y_{ft}^*}{Y_t^*}$ as weights, ,

$$p_t^* = \sum s_{ft} p_{ft}^* = \log \frac{\overline{W}_t}{F_{N|h=1}} + \left(a + \frac{1}{\sigma}\right) \sum s_{ft} \log h_{ft} + \mu \tag{A.25}$$

As in full-capacity equilibrium capital intensities and marginal costs of labor and capital are equal across firms, then also $\frac{N_{ft}}{N_t} = \frac{Y_{ft}^*}{Y_t^*} = s_{ft}$ and(A.25) can be presented in terms of the aggregate level utilization rates,

$$p_t^* = \log \frac{\overline{W}_t}{F_{N|h=1}} + \underbrace{\left(a + \frac{1}{\sigma}\right)\log h_t}_{\frac{\Theta}{1+\mu}\left(a+\frac{1}{\sigma}\right)\log u_t} + \mu \tag{A.26}$$

Equation (A.26) can also be presented, instead of the straight time wage rate $\overline{W}_t$, also in terms of the observed wage rate per employed worker $W_t = \overline{W}_t\left[h_t + \frac{a}{2}\left(h_t - 1\right)^2\right]$. After using the Taylor approximation $\log\left[h_t + \frac{a}{2}\left(h_t - 1\right)^2\right] \approx \log h_t$ and denoting conventional nominal marginal costs as $cmc_t^n = \log \frac{W_t}{F_{N|h=1}}$, we have,

$$p_t^* = \underbrace{\underbrace{cmc_{ft}^n}_{\text{Conventional}} + \underbrace{\left(a + \frac{1}{\sigma} - 1\right)\log h_t}_{\text{Utilization}}}_{fmc_{ft}^n} + \mu \tag{A.27}$$

$$p_t^* = \underbrace{\underbrace{cmc_t^n}_{\text{Conventional}} + \underbrace{\frac{\Theta}{1+\mu}\left(a + \frac{1}{\sigma} - 1\right)\log u_t}_{\text{Utilization}}}_{fmc_{ft}^n} + \mu \tag{A.28}$$

Real marginal costs are then,

$$\begin{aligned} p_t^* - p_t &= \mu + cmc_t^r + \varphi^h \log h_t \\ &= \mu + cmc_t^r + \varphi^u \log u_t \end{aligned} \tag{A.29}$$

The composite parameters are given by $\varphi^h = a + \frac{1}{\sigma} - 1$ and $\varphi^u = \frac{\Theta}{1+\mu}\varphi^h = \frac{\varphi^h}{(1+\mu)\left[1+\alpha\left(\rho_t^{\widetilde{\kappa},h} - 1\right)\right]}$. Following this, the derivation of the NKPC (with and without intrinsic persistence) follows in the normal way.

B Data Construction

We use US 1953q1 to 2011q4 data from NIPA-BEA and BLS.

B.1 Raw Data Series

National Income and Product Accounts (NIPA-BEA)

1. Table 1.3.5 [Gross Value Added by Sector, Billions of Dollars] Line 1: GDP, Line 8: General Government, Line 11: Gross Housing Value Added
2. Table 1.3.6 [Real Gross Value Added by Sector, Chained Dollars] Line 1: GDP, Line 8: General Government, Line 11: Gross Housing Value Added
3. Table 1.12 [Net Income by Type of Income, Billions of Dollars] Line 2: PrivateCompensation of Employees (Line 2 minus Line 4 minus Line 8), Lines 19, 20: Indirect taxes less subsidies.
4. Table 5.2.6U[34] [Real Gross and Net Domestic Investment by Major Type, Chained Dollars] Line 12: Fixed Net Non-Residential Investment.
5. Table 6.5, Fulltime Equivalent Employees by Industry (Annual Data, thousands) Private Fulltime Equivalent (Line 1 minus Line 75)
6. Table 6.7, Self-Employed Persons by Industry (Annual Data, thousands) Line 1

Fixed Asset Tables (FAT-BEA)

1. Table 4.2. Chain-Type Quantity Indexes for Net Stock of Private Nonresidential Fixed Assets by Industry Group and Legal Form of Organization [Index numbers, 2005=100], Annual Series, Line 1: Private Non-Residential Fixed Assets

Bureau of Labor Statistics (BLS)

1. Employee Number Private: Current Employment Statistics survey (National), Series id, ES0500000001.

B.2 Fuller Description

Total Income , Real Production and the Price of Production

Our production concept covers private sector excluding Gross Housing Value Added. This adjustment is done because the latter component (mainly rent and imputed rent income of housing) is by definition closely related housing stock and housing price developments. Also to eliminate the effects of indirect taxation on prices, total nominal income is calculated excluding indirect taxes less subsidies (NIPA Table 1.12, Lines 11 and 12). Hence, total income is defined as, Gross Domestic Product (GDP) minus Gross

[34] "U" indicates supplementary tables.

Value Added of General Government (G) minus Gross Housing Value Added (H) minus Indirect Taxes less Subsidies (TS),

$$P * Y = GDP - G - H - TS$$

These gross value added product figures are taken from NIPA Table 1.3.5. NIPA Table 1.3.6., in turn, gives the corresponding Real Gross Value Added figures (in Chained Dollars). Indirect taxation is accounted by scaling the real gross domestic product down in proportion to the tax content of indirect taxes less subsidies in the base year. The price of private non-housing production is calculated as the ratio of nominal and real production series.

Capital stock

In the US there are no published series for the real capital stock on quarterly level. However, the U.S. Bureau of Economic Analysis (BEA) publishes quarterly figures of Real Net Non-Residential Investment (NIPA Table 5.2.6 U). Accumulating this series over the sample period gives, bar the initial value, the development of the capital stock. The initial level (1951q4 level) of our capital stock series is evaluated so that the growth of the sum of the initial capital stock and the cumulated net investment from 1951q4 to 2005q4 equals the growth of the annual index of Net Non-Residential Fixed assets (BEA, Fixed Assets Accounts Tables/Standard Fixed asset Tables, Table 4.2) over the period of 1951 - 2005.

Labor Series

Annual level of our employment series is the sum of the Full-time Equivalent Employees (NIPA Table 6.5, private sector) and Self-Employed (NIPA Table 6.7, private sector). As these series are available only as annual levels the quarterly variation has moved to them from the Bureau of Labor Statistics series: Employee Number Private: Current Employment Statistics survey (National), Series id, ES0500000001.

Labor income

Labor income is calculates as a sum of Compensation of Employees (NIPA Table 1.12 , Private Sector) and the labor income of Self-Employed Persons. As the NIPA figures of proprietors' income contain besides labor income also a capital income component of self-employed persons, the imputed labor income of self-employed persons is evaluated by assuming that the shadow wage rate of self-employed persons equals the average compensation rate per employee. Our historical data set was originally collected from the NIPA tables with 2000 as the base year over the period 1952:1 -2007:2. We extended this data set to 2011:4 by the percentage changes of the corresponding NIPA series of the year 2005 as the base.

C A Stationary Equilibrium System

Rewrite the aggregate real marginal cost equation (A.29) in the form,

$$\log P_t - \log(1+\mu) - \log\left(\frac{W_t}{F_{N|h_t=1}}\right) = \varepsilon_t^P \tag{C.1}$$

FOCs (A.8) and (A.9), with the discussed production-function and aggregation properties, imply the relations,

$$F_{K|\widetilde{\kappa}=1} - (1+\mu)\, r_t^K = \varepsilon_t^{F_K} \tag{C.2}$$

$$\log Y_t - \log F\left(\Gamma_t^K K_t, \Gamma_t^N N_t\right) = \varepsilon_t^Y \tag{C.3}$$

where,

$$\varepsilon_t^P = -\underbrace{\log\left(\frac{P_t^*}{P_t}\right)}_{\log[(1+\mu)CMC_t^r]} + \varphi^u \log\left(\frac{Y_t}{Y_t^*}\right)$$

$$\varepsilon_t^{F_K} = F_{K|\widetilde{\kappa}=1}\left(1 - \frac{P_{ft}}{P_t}\widetilde{\kappa}_{ft}^{\frac{\sigma-1}{\sigma}}\right) + (1+\mu)\left[\left(r_{ft}^K - r_t^K\right) + \left\{\begin{array}{c} \frac{\partial\Omega_K\left(K_{ft},,..\right)}{\partial K_{ft}} + \frac{1}{1+r_t}\mathbb{E}_t\frac{\partial\Omega_K\left(K_{ft+1},..\right)}{\partial K_{ft}} \\ + \left(\frac{1}{1+r_t}\right)\mathbb{E}_t\left[\left(\frac{1}{1+r_{t+1}}\right)\frac{\Omega_K\left(K_{ft+2},..\right)}{\partial K_{ft}}\right] \end{array}\right\}\right]$$

$$\varepsilon_t^Y = \log\left(\frac{Y_t}{Y_t^*}\right)(= u_t - 1)$$

These residuals are all by definition stationary. After the left hand sides of (C.1)-(C.3) are expressed in terms of observable aggregate level I(1) variables. The right hand sides are not directly observable but have exact economic interpretations and are, by definition, stationary. Hence, using the terminology of the cointegration literature we have the three equation system of long-run equilibrium relationships: i.e., a relationship between observable variables which has, on average, been maintained for a long period. The estimation of this equilibrium system, i.e. by treating the stationary right hand sides as estimation residuals, allows us to extract consistent – indeed super consistent, Stock (1987) – estimates of the parameters of interest.

C.1 Residual Interpretation

It turns out that the residuals in the system above have an important – and, in the literature, overlooked – property. The residual from (C.3) gives the capacity utilization rate. The residual of (C.1) is the difference of the markup over full real marginal cost plus $\widehat{\varphi^u}$ times the capacity utilization rate (i.e. the residual of (C.3)). Real marginal costs – except for the exact parameter value φ^u multiplying the capacity utilization rate – are fully determined by these two residuals, and can be consistently substituted into the dynamic NKPC specification presented below. In the NKPC all variables are $I(0)$ series. Hence, in terms of the cointegration literature the estimation the NKPC equation represents the second step of the Engle-Granger two-step approach to estimate a dynamic equation of cointegrated variables (Granger (1983), Engle and Granger (1987)) In turn, estimation of the NKPC allows us to estimate φ^u (alternatively, φ^h), as well as β, θ and ω. Further, it is interesting to see that, if utilization margins matter for the correct measurement of real marginal costs (i.e., implying $\varphi^h, \varphi^u \neq 0(\varphi^h, \varphi^u > 0 \text{ if } \sigma < 1)$) then the estimation residuals of (C.1) and (C.3) must be correlated (positively if $\sigma < 1$). In the special case of frictionless price setting – i.e. $P_t = P_t^f$ and $MC_t^r = (1+\mu)^{-1}$ – this correlation would be perfect. Friction in price setting, via a time-varying dynamic markup (i.e., P_t^f/P_t being non constant over time), decreases the correlation between these two residuals.