Resurrecting the Role of the Product Market Wedge in Recessions

By Mark Bils, Peter J. Klenow, and Benjamin A. Malin*

Employment and hours are more cyclical than dictated by productivity and consumption. This intratemporal labor wedge can arise from product or labor market distortions. Based on employee wages, the literature has attributed the intratemporal wedge almost entirely to labor market distortions. Because wages may be smoothed versions of labor's true cyclical price, we instead examine the self-employed and intermediate inputs, respectively. For recent decades in the United States, we find price markup movements are at least as cyclical as wage markup movements. Thus, countercyclical price markups deserve a central place in business-cycle research, alongside sticky wages and matching frictions. (JEL E24, E32, E63, J31, J41)

Employment and hours are more cyclical than can be explained by real labor productivity under conventional preferences for consumption and leisure. See Hall (1997); Mulligan (2002); and Chari, Kehoe, and McGrattan (2007), among others. This “labor wedge” could reflect either a gap between the marginal product and price of labor (i.e., a price markup) or a gap between the price of labor and the opportunity cost of supplying labor (i.e., a wage markup). The decomposition matters for both stabilization policy and understanding the nature of business cycles. Recently, a growing consensus argues that the wage markup plays a dominant role: see, for example, Galí, Gertler, and López-Salido (2007); Shimer (2009); Hall (2009); and Karabarbounis (2014a).

The consensus that the labor wedge reflects labor market distortions is based on measuring the price of labor using average hourly earnings. The gap between average hourly earnings and labor productivity is acyclical, suggesting that price markup movements are not cyclical. But it is not clear whether the marginal cost of labor to firms is well measured by average hourly earnings. Employee wages may not reflect the true marginal cost of labor to the firm. Wages may be smoothed versions of the shadow cost due to implicit contracting (e.g., for salaried workers). One obtains a
very different picture of the cyclical price of labor using the wages of new hires, as measured by Pissarides (2009) and Haefke, Sonntag, and van Rens (2013), or from the user cost of labor, as measured by Kudlyak (2014).

In this paper, we seek evidence on cyclical distortions in the product market that does not rely on wage data for workers. First, we estimate the labor wedge for the self-employed. If we observe significant cyclicality in the labor wedge for the self-employed, we presume it cannot be ascribed to wage rigidities or other labor market frictions. Second, we estimate the price markup from intermediate inputs (energy, materials, and services). Intermediate prices should provide a truer measure of that input’s cyclical price than do average hourly earnings for labor.

Our evidence is for the United States from 1987 onward. A benefit of using two distinct approaches is that they rely on different data sources. For the self-employed, we look at household data from the Current Population Survey (CPS) and the Consumer Expenditure Survey (CE), both conducted by the Bureau of Labor Statistics (BLS). For intermediates, we use the BLS Multifactor Productivity Database covering 60 industries. Our consistent finding is that, contrary to the emerging consensus, product market distortions are at least as important as labor market distortions in recent recessions.

The countercyclical price markup we estimate is compatible with prices being sticky (too high) in recessions, serving as a constraint on firm sales. Given its strong persistence, it is also consistent with firms purposefully choosing a higher price markup in recessions. As a recent example, Gilchrist et al. (2017) find that financially constrained firms chose higher markups at the expense of future market share during the Great Recession. Any model where expanding production has a component of investment (e.g., learning-by-doing) should have similar implications. The countercyclical price markup could also reflect greater uncertainty (or aversion to uncertainty) in recessions as in Arellano, Bai, and Kehoe (2012).

Our study is related to several earlier efforts. Bils (1987), Rotemberg and Woodford (1999), Bils, Klenow, and Malin (2013), and Nekarda and Ramey (2013) rely on measured earnings to estimate the marginal price of labor and thereby the cyclicality of the markup. Other studies infer the cyclicality of price markups from the cyclicality of investment (Galeotti and Schiantarelli 1998), final goods inventories (Bils and Kahn 2000; Kryvtsov and Midrigan 2013), and advertising (Hall 2014). While our results for intermediates have antecedents in the literature (e.g., Basu 1995), our analysis is the first to examine plant-level compositional biases in manufacturing, to look beyond manufacturing, and to address the cyclicality of the price markup versus the wage markup.

The paper proceeds as follows. Section I revisits the conventional labor wedge decomposition using wages, doing so for both the extensive and intensive margins. Section II looks at the self-employed. Section III investigates intermediate input use. Section IV relates our work to other efforts at measuring the cyclicality of markups. Section V concludes.

I. The Aggregate Labor Wedge

We begin by constructing the standard representative-agent labor wedge (RAW), defined as the ratio of the marginal product of labor (mpn) to the tax-adjusted
marginal rate of substitution of consumption for leisure (mrs). Constructing the wedge requires assumptions on preferences and technology; our baseline case follows Hall (1997) and Galí, Gertler, and López-Salido (2007). Production features a constant elasticity of output with respect to hours. Preferences are separable in consumption and hours and, over time, with a constant intertemporal elasticity for consumption and a constant Frisch elasticity of labor supply. These assumptions yield a log-linear labor wedge

\[
\text{RAW}_t \equiv \ln(m_{tn}) + \ln(1 - \tau_t) - \ln(mrs_t) \\
= \ln\left(\frac{y_t}{n_t}\right) + \ln(1 - \tau_t) - \left[\frac{1}{\sigma}\ln(c_t) + \frac{1}{\eta}\ln(n_t)\right],
\]

where \(y_t/n_t\) is output per hour, \(c_t\) is nondurables and services consumption per adult equivalent, \(n_t\) is hours per capita, and \(\tau_t \equiv \frac{\tau^c_t + \tau^n_t}{1 + \tau^n_t}\) is a combination of average marginal tax rates on consumption and labor. See the online Appendix for a precise description of all variables used.

For our baseline case, we use an intertemporal elasticity of substitution (IES) of \(\sigma = 0.5\), following Hall (2009), and a Frisch elasticity of labor supply of \(\eta = 1.0\), based on Chang et al. (2014). The latter argue, based on a heterogeneous-agent model with both intensive and extensive labor margins, that a representative-agent Frisch elasticity of 1 (or slightly higher) is reasonable.

To gauge the cyclicality of the RAW, we project it on real GDP and hours worked. (All variables are in logs and HP-filtered.) We use quarterly data from 1987 through 2012. Table 1 reports the cyclical elasticity of the wedge and its components: labor productivity, hours worked, consumption, and taxes. The wedge is strongly countercyclical (elasticity with respect to GDP: \(-2.69\)), reflecting mildly countercyclical productivity (\(-0.10\)), procyclical consumption (0.61), and highly procyclical hours (1.40). In recessions, the RAW increases as the mrs plummets but the mpn changes little. Using equation (1) and the results in Table 1, it is straightforward for the reader to recalculate the wedge’s cyclicality for alternative calibrations of \(\sigma\) and \(\eta\).

\[1\] We later entertain constant elasticity of substitution (CES) production in capital, labor, and intermediate inputs.

\[2\] Nonseparable utility in consumption and leisure would not alter our results significantly. Shimer (2009) and Karabarbounis (2014a) found this as well. We find the same if we calibrate the nonseparability to how consumption responds to retirement (Aguiar and Hurst 2013) or unemployment (Saporta-Eksten 2014).
Because the cyclical elasticity will be a key statistic summarizing wedge behavior throughout the paper, we point out a few other aspects of Table 1. First, the contribution of marginal tax rates to the cyclicity of the RAW is small. Because our tax measures have little impact on our results, we drop them in the remainder of the paper. Second, we have detrended our series with an HP-filter in order to focus on cyclical, rather than low-frequency, fluctuations in the wedge. Hamilton (forthcoming) has recently pointed out drawbacks of using the HP-filter and suggested an alternative method. In the online Appendix, we show that our main findings are robust to using Hamilton’s method.

Before proceeding, it is useful to take a step back and consider exactly what the “labor wedge” is. It measures the gap between the social cost and social benefit of a marginal unit of market labor under a given set of assumptions. It implies a deviation between hours worked in the data and what a social planner would choose. In practice, the measured gap will also reflect any mismeasurement in the data and/or misspecification of the social planner’s problem. Thus, it is worthwhile to consider possible mismeasurements and misspecifications and how they might affect our results.

The Table 1 calculations ignore fluctuations in the quality of the workforce and a role for overhead labor. We calculate that declining average workforce quality in expansions leads us to understate the cyclical elasticity of labor’s marginal product by perhaps 0.1 to 0.2 percent, and in turn overstate the cyclicity of the wedge. Ignoring overhead labor, conversely, overstates the procyclicality of labor’s marginal product (Rotemberg and Woodford 1999). For an overhead labor component of the magnitude suggested by Nekarda and Ramey (2013)—10 to 20 percent of employment—the effects of composition and overhead labor on the cyclicity of labor’s marginal versus average product (and on the estimated labor wedge) should approximately offset.

A representative agent requires not only workers of equal quality but also perfect consumption sharing. To the extent this is not true, the cyclicity of the wedge may be overstated. In the online Appendix we calculate that the consumption of those who are continuously employed does not move as much as per capita consumption (cyclical elasticity with respect to GDP of 0.52 for the employed versus 0.61 per capita), which implies a slightly less countercyclical wedge (elasticity of −2.51 instead of the −2.69 in Table 1).

3 Mulligan (2012) contends that changes in effective marginal tax rates influenced labor market behavior during the Great Recession. His focus is on how lower-income workers have been affected by the expansion of means-tested assistance programs.

4 The calculations also ignore home production, which Karabarbounis (2014b) suggests can explain part of the cyclicity in the labor wedge. Countercyclical hours worked at home imply smaller cyclical movements in effective leisure, whereas countercyclical consumption of home-produced goods implies smaller movements in effective consumption. Note, however, that our intertemporal and Frisch elasticities are calibrated to a literature that estimates models without home production. Matching the same empirical moments in a model with home production would therefore require lower intertemporal and Frisch elasticities. These lower elasticities would (at least partly) offset the impact of the smaller movements in effective leisure and consumption on the labor wedge.
A. Extensive and Intensive Margin Wedges

We next depart from the representative agent framework by constructing separate wedges on the extensive margin (EMW) and the intensive margin (IMW). These wedges distinguish between the two components of hours worked (employment and hours per worker) but still measure gaps between the social benefit and social cost of additional labor, either in the form of an additional employee or an additional hour per worker. They will also be useful for decomposing the labor wedge into product and labor market components for three reasons. First, we can calibrate the Frisch elasticity of labor supply to micro estimates at the hours margin. Second, we can compare the intensive margin here with the intensive margin for the self-employed (in Section II). Third, product market distortions should have an impact on the wedge on both margins. So, if the labor wedge is important at only one margin, that would be suggestive that the price markup has little cyclical importance.

To analyze the extensive margin, we make some additional assumptions. We consider a representative household that consists of many members. Consumption is perfectly shared across members, and labor supply decisions are made on both the extensive and intensive margins. Preferences are

$$E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \frac{c_t^{1-1/\sigma}}{1-1/\sigma} - \nu \left( \frac{h_t^{1+1/\eta}}{1+1/\eta} + \psi \right) e_t \right\},$$

where $c_t$ denotes consumption per adult equivalent, $e_t$ employment, and $h_t$ hours worked per employee. The variable $\psi$ is a fixed cost of employment, which guarantees an interior solution for hours versus employment. The marginal disutility of employment is $\nu \left( \frac{h_t^{1+1/\eta}}{1+1/\eta} + \psi \right) \equiv \nu \Omega_t h_t$, while the marginal disutility of an extra hour per worker is $\nu h_t^{1/\eta} e_t$.

For firms we assume (i) a constant output elasticity with respect to labor, and (ii) employment and hours per worker are perfect substitutes (i.e., production depends on total hours, $n_t = e_t h_t$). The marginal product of labor is thus proportional to output per hour $y_t/n_t$, while the marginal product of employment is $mpn_t^{ext} \propto \frac{y_t}{n_t} h_t$ and the marginal product of hours per worker is $mpn_t^{int} \propto \frac{y_t}{n_t} e_t$.

Adding employment involves frictions. Firms post vacancies at the beginning of the period, while matches form and produce during the period. The matching technology is $m_t = \nu_t^\phi f(u_t)$, where $m_t$ are matches, $\nu_t$ vacancies, and $u_t$ unemployment. Let the variable $\kappa$ denote the opportunity cost of creating a vacancy, expressed in labor input as a fraction of the steady-state workweek $h$. Let $\delta$ be the per-period

---

5 The EMW is also interesting in its own right and, to our knowledge, it has not been constructed empirically (though it appears in many theoretical models).  
6 Some labor market distortions may hit both margins proportionally. Thus, finding equally important wedges at both margins does not rule for the price markup. But if we find the wedge is only important at one margin, this is indirect evidence against product market distortions.  
7 Blanchard and Gali (2010) and Gali (2011) use this timing, although it is more conventional for matches to start producing in the following period. The former timing gives cleaner results, but it could be altered without changing our analysis substantially.
separation rate, and \( \gamma \) the fraction of the initial period of employment devoted to training.\(^8\)

In this environment, the intensive margin wedge is given by

\[
(2) \quad IMW_t \equiv \ln(mpn^{int}_t) - \ln(mrs^{int}_t) = \ln\left(\frac{y_t}{n_t}\right) - \left[\frac{1}{\sigma} \ln(c_t) + \frac{1}{\eta} \ln(h_t)\right].
\]

The IMW in (2) differs from the standard RAW in equation (1) in two ways: hours per worker \( h_t \) replaces hours per capita \( n_t \), and we calibrate \( \eta = 0.5 \). A lower \( \eta \) is appropriate given that it now reflects the Frisch elasticity only at the intensive (hours) margin (Chetty et al. 2013).

On the extensive margin, consider creating one more vacancy in period \( t \) and reducing vacancies in \( t + 1 \) by just enough to keep employment unaffected from \( t + 1 \) forward. Spending \( \kappa h_t \frac{y_t}{n_t} \) to create an additional vacancy generates \( \phi m_t/v_t \) additional matches, of which \( (1 - \delta) \) survive to \( t + 1 \). This perturbation requires lower spending on vacancies at \( t + 1 \) by \( (1 - \delta) \kappa h_t \frac{y_{t+1}}{n_{t+1}} \frac{m_t/v_t}{v_{t+1}} \). A social planner would set

\[
(3) \quad \frac{\phi m_t}{v_t} \left[u'(c_t)\left(1 - \gamma \frac{h_t}{h_t}\right) h_t \frac{y_t}{n_t} - \Omega_t h_t\right] - u'(c_t)\kappa h_t \frac{y_t}{n_t} + \beta(1 - \delta) E_t \left\{ \frac{\phi m_t}{v_t} \right\} \frac{\frac{y_{t+1}}{n_{t+1}} \frac{m_t/v_t}{v_{t+1}}}{\gamma} = 0.
\]

The marginal benefit of an extra vacancy (utility from consuming increased output today) equals its marginal cost (disutility of employment plus the cost of creating an added vacancy today less the resource savings from creating fewer future vacancies). Rearranging (3) to get a (log) ratio of the marginal benefit to the marginal cost of additional employment, we have

\[
(4) \quad EMW_t = \ln\left(\frac{y_t}{n_t}\right) - \left[\frac{1}{\sigma} \ln(c_t) + \ln(\Omega_t)\right] - S_t,
\]

\[
S_t \approx \kappa V_m \frac{\gamma}{\Omega t} \left[1 - \beta(1 - \delta) E_t \left\{ u'(c_{t+1}) \frac{\gamma_{t+1} \gamma_{t+1}}{\gamma_t \gamma_t}\right\} + \gamma \left[1 - \beta(1 - \delta) E_t \left\{ u'(c_t) \frac{\gamma_{t+1} \gamma_{t+1}}{\gamma_t \gamma_t}\right\} \right]\right]^{-1}.
\]

where \( \Omega_t \) is the marginal disutility of employment (per hour worked).\(^9\)

The EMW, like the IMW, reflects movements in labor productivity, \( \ln\left(\frac{y_t}{n_t}\right) \), and the marginal utility of consumption, \( \frac{1}{\sigma} \ln(c_t) \). But the EMW also differs from the IMW. Whereas the IMW reflects the marginal disutility of an extra hour, which is highly procyclical for reasonable Frisch elasticities, the extensive margin reflects

\(^8\)The assumptions that \( \kappa, \delta, \) and \( \gamma \) are constant over time (and thus exogenous) are not necessary for the derivation of the EMW but do simplify the data needed for its construction.

\(^9\)The approximation stems from our use of several first-order Taylor series approximations. The online Appendix provides more details on the EMW (and IMW) construction.
the average disutility of adding a worker. We find this average disutility to be nearly acyclical. In addition, the term $S_t$, which is specific to the EMW, reflects the efficacy of spending on vacancies. In recessions, $S_t$ declines as vacancies are more likely to yield a match. This lends a countercyclical component to the EMW. The cyclicity of the EMW vis-à-vis the IMW essentially reduces to whether cyclicity in the hiring term $S_t$ exceeds that in the marginal disutility of working a longer workweek.

It is well established that cyclical movements in total hours are primarily driven by employment fluctuations. That holds true for our 1987 to 2012 period, where cyclical employment fluctuations (with respect to real GDP or total hours) are about four times larger than in the workweek. One might jump to the conclusion that cyclicity in the extensive (employment) margin wedge would similarly dominate cyclicity in the intensive (workweek) margin wedge. That jump would be unwarranted, however. Important components of the wedge, labor’s marginal product and consumption’s marginal utility, display the same cyclicity with respect to both margins. Any wedge differences, as just discussed, reduce to cyclicity in the hiring term $S_t$ versus that in the marginal disutility of working a longer workweek. Although employment fluctuations are larger than those in the workweek, the elasticity of the marginal disutility in response to the workweek may exceed that of the hiring cost $S_t$ to employment.

We treat the elasticity of labor supply at the intensive margin as reflecting curvature in preferences, while that at the extensive margin reflects the impact of adjustment costs. Of course, a more general calculation might allow for adjustments costs at the intensive margin as well, or allow for upward-sloping labor supply at the extensive margin, beyond that generated by adjustment costs. For instance, consider heterogeneity in worker productivity. This introduces comparative advantage in labor supply. Ignoring search costs, the social planner would select the most productive workers for employment first. If expanding employment requires adding a worker with less market comparative advantage in a boom than in a bust, then this should act to stabilize employment. So by ignoring comparative advantage, we may understate the cyclicality of the wedge at the extensive margin.\footnote{Heterogeneity in preferences through $\Omega_t$ acts similarly, as comparative advantage is then dictated by a worker’s productivity relative to $\Omega_t$.}

To construct the EMW, we use the same variables required by the IMW, plus vacancies ($v_t$), matches ($m_t$), and additional parameters. A quarterly separation rate of $\delta = 0.105$ matches the average rate of quits, layoffs, and discharges in the BLS Job Openings and Labor Turnover Survey. The discount factor $\beta = 0.996$ implies a steady-state annual real interest rate of 1.6 percent, the average of the three-month T-bill rate less core PCE inflation over our sample. Hiring costs per match, $\kappa v / m$, are set to 0.4 quarters of output, and training costs to $\gamma = 0.16$, consistent with estimates by Barron, Berger, and Black (1999). Finally, the elasticity of matches to vacancies is set to $\phi = 0.5$. These parameter values imply a steady-state ratio of $mrs$ to $mpn$ on the extensive margin of about 0.90.

We find that the EMW and IMW elasticities are quite similar. Figure 1 shows the unfiltered extensive and intensive margin wedges from 1987 to 2012. Table 2 reports their cyclical elasticities with respect to real GDP and hours worked. The
elasticities of EMW and IMW with respect to GDP are \(-1.99\ (0.26)\) and \(-1.91\ (0.13)\), respectively.

B. Decomposing the Wedge

We now empirically decompose the labor wedge into product market and labor market components (or, equivalently, price and wage markups). This requires a measure of the marginal cost of labor to firms. As stressed by Galí, Gertler, and López-Salido (2007), the assumption that any particular wage measure reflects labor’s true marginal cost is controversial. We will show that alternative wage measures lead to vastly different conclusions about the relative importance of the price markup versus wage markup. This motivates our subsequent analysis, which decomposes the labor wedge without using wage data.

11 EMW and IMW elasticities are smaller than for the RAW; an aggregate Frisch elasticity of 2.3 would make the RAW behave similarly to the EMW and IMW.
The IMW decomposition is standard and given by

\[ IMW_t = \left( \ln \left( \frac{y_t}{n_t} \right) - \ln \left( \frac{w_t}{p_t} \right) \right) + \left( \ln \left( \frac{w_t}{p_t} \right) - \frac{1}{\sigma} \ln(c_t) - \frac{1}{\eta} \ln(h_t) \right) = \mu^{p,int}_t + \mu^{w,int}_t, \]

where \( \frac{w_t}{p_t} \) is the (real) marginal cost of labor to firms. The price markup based on the intensive margin (\( \mu^{p,int}_t \)) is the gap between the firm’s marginal product and marginal cost of labor. The wage markup based on the intensive margin (\( \mu^{w,int}_t \)) is the gap between the firm’s marginal cost and the household’s cost of providing an additional hour.

The EMW decomposition is

\[ EMW_t = \left( \ln \left( \frac{y_t}{n_t} \right) - \bar{S}_t - \ln \left( \frac{w_t}{p_t} \right) \right) + \left( \ln \left( \frac{w_t}{p_t} \right) + \bar{S}_t - S_t - \frac{1}{\sigma} \ln(c_t) - \ln(\Omega_t) \right) = \mu^{p,ext}_t + \mu^{w,ext}_t, \]

where \( \bar{S}_t \) takes the same form as \( S_t \) in equation (4) but with \( \phi = 1 \). The presence of \( \bar{S}_t \) rather than \( S_t \) reflects the fact that firms do not internalize the congestion effects of their decision to post another vacancy; each firm views the probability of filling a vacancy as \( m/v \), whereas the social planner knows one more vacancy generates \( \phi m/v \) additional matches. For intuition, momentarily consider \( \bar{S}_t = S_t \) in equation (6). Doing so, it is apparent that the wage markup on the extensive margin (\( \mu^{w,ext}_t \)) mirrors that on the intensive margin (\( \mu^{w,int}_t \)), but with the household’s \( mrs \) calculated at the employment margin. For the price markup on the extensive margin (\( \mu^{p,ext}_t \)), an additional employee produces \( \ln \left( \frac{y_t}{n_t} \right) - S_t \) because firms pay the vacancy cost (\( S_t \)).

Table 3 decomposes the EMW and IMW into the price markup and wage markup using average hourly earnings (AHE) as the measure of the firm’s marginal cost of labor (\( w_t \)). This wage measure is conventional and would reflect the true marginal cost if all workers were employed in spot markets. Here the price markup \( \ln \left( \frac{y_t}{n_t} \right) - \ln \left( \frac{w_t}{p_t} \right) \) is simply the (log) inverse labor share. Using AHE, the price markup accounts for only 2 to 6 percent of the cyclicity of the labor wedge on the intensive margin, and between 16 and 23 percent on the extensive margin. These results are in line with Karabarbounis’s (2014a) conclusion that the price markup is relatively unimportant.

<table>
<thead>
<tr>
<th></th>
<th>Elasticity w.r.t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDP</td>
</tr>
<tr>
<td>Extensive margin wedge</td>
<td>-1.99 (0.26)</td>
</tr>
<tr>
<td>Price markup (AHE)</td>
<td>-0.32 (0.14)</td>
</tr>
<tr>
<td>Intensive margin wedge</td>
<td>-1.91 (0.13)</td>
</tr>
<tr>
<td>Price markup (AHE)</td>
<td>-0.04 (0.08)</td>
</tr>
</tbody>
</table>

Notes: Each entry is from a separate regression. Sample covers 1987:1–2012:IV. All variables are in logs and HP-filtered. Expectation terms in the EMW are constructed using a VAR. The price markup inferred from the extensive margin (\( \mu^{p,ext} \)) follows equation (6), and the price markup from the intensive margin (\( \mu^{p,int} \)) follows (5).
Alternative frameworks for understanding the labor market, however, emphasize the durable nature of the firm-worker relationship and imply that the contemporaneous wage plays no allocative role. For example, in matching models with search frictions, what matters is the expected surplus generated over the life of the match and not the wage payment at any one time. Implicit contracting models similarly imply that the flow wage payment is not allocative. Barro (1977) and Rosen (1985) forcefully drive home that an acyclical wage, or even a countercyclical wage, can coincide with a highly procyclical true price of labor (see also Basu and House 2016). Recent examples of empirical support for implicit contracting include Ham and Reilly (2013) and Bellou and Kaymak (2012).

In Table 4 we examine how cyclicality in two alternative measures of the marginal cost of labor affect the wedge decomposition. We use cyclicality for the wages of new hires (NH) and the user cost of labor (UC) as reported by Kudlyak (2014) (please see online Appendix). Both are arguably more relevant for job formation. We therefore focus on the extensive margin. We find that, depending on the wage measure used, the price markup can account for almost none or essentially all of the EMW, from 16 percent using average hourly earnings to 109 percent using user cost. (Using the user cost of labor, the price markup also accounts for all cyclicality in the intensive margin wedge from Table 3.)

### II. The Self-Employed Wedge

Given the ambiguity of the decomposition results just presented, it is useful to explore decompositions of the labor wedge that do not depend on wage data. In the online Appendix we present a static decision problem of individuals choosing whether to work for others, work for themselves as sole proprietors, or work for themselves with employees. Talent draws (or occupation-specific human capital) determine which of these choices maximize utility. Whether a self-employed person has employees or not, they choose how many hours to work, leading to a first-order condition on the intensive margin. In this first-order condition, their marginal self-employed earnings are proportional to their average self-employed earnings per

---

**Table 4—Wedge Decomposition: Alternative Wage Measures**

<table>
<thead>
<tr>
<th></th>
<th>Elasticity w.r.t. GDP</th>
<th>Elasticity w.r.t. Total hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extensive margin wedge</td>
<td>−1.99 (0.26)</td>
<td>−1.55 (0.14)</td>
</tr>
<tr>
<td>Price markup (AHE)</td>
<td>−0.32 (0.14)</td>
<td>−0.35 (0.09)</td>
</tr>
<tr>
<td>Price markup (NH)</td>
<td>−0.98 (0.16)</td>
<td>−0.81 (0.09)</td>
</tr>
<tr>
<td>Price markup (UC)</td>
<td>−2.17 (0.21)</td>
<td>−1.65 (0.09)</td>
</tr>
</tbody>
</table>

Notes: Each entry is from a separate regression. Sample covers 1987:I–2012:IV. All variables are in logs and HP-filtered. Expectation terms in the EMW are constructed using a vector autoregression (VAR). The price markup follows equation (6). See the online Appendix for details on construction of NH and UC.

---

12 See Pissarides (2009); Haefke, Sonntag, and van Rens (2013); and Kudlyak (2014). Kudlyak finds that the new hire wage falls 1.2 percent relative to all workers’ wages for each percentage point rise in the unemployment rate, whereas the relative fall in the user cost of labor is 3.4 percent.
hour. This leads to a familiar labor wedge for the self-employed with the usual decomposition:

\[
\mu_{SE} = \ln(mpn)_{SE} - \ln\left(\frac{w}{P}\right)_{SE} + \mu^p_{SE} - \ln(mrs)_{SE} - \ln\left(\frac{w}{P}\right)_{SE}.
\]

We assume \(\mu^w_{SE} = 0\). That is, for the self-employed, we assume that their marginal earnings, i.e., the marginal price of their own labor \(\left(\frac{w}{P}\right)_{SE}\), is equal to their \(mrs\) because no wage rigidities or other labor market distortions impinge on their decision to supply labor to their own business. This implies that any labor wedge for the self-employed must be due to the price markup:

\[
\mu_{SE} = \ln(mpn)_{SE} - \ln(mrs)_{SE} + \mu^p_{SE}.
\]

Note that, by assuming \(\mu^w_{SE} = 0\), we can use \(mrs_{SE}\) as our measure of the unobserved shadow wage for the self-employed \(\left(\frac{w}{P}\right)_{SE}\).

We maintain our Section I assumptions on preferences. For technology, unlike in Section I, where we implicitly assumed employee and self-employed hours were perfect substitutes, here we assume that self-employed hours enter production alone with a constant elasticity. The resulting \(\mu_{SE}\) takes the same form as equation (2), though we must now measure hours worked, productivity, and consumption for the self-employed. To do so, we turn to the CPS and CES. We first document the cyclical behavior of self-employed hours worked and then check whether movements in productivity and consumption can explain these fluctuations.

As a starting point, we note that self-employment has been just as cyclical as total employment. The share of self-employed in nonagricultural industries declined slightly during each of the past two NBER-defined recessions: from 10.1 to 10.0 percent during 2001 and from 10.5 to 10.4 percent from 2007 to 2009.\(^{13}\) The self-employment share exhibits lower frequency fluctuations, but if we HP-filter, the resulting series is completely acyclical with respect to GDP or aggregate hours. Becoming self-employed requires starting a business, so fluctuations in self-employment could be affected by financing costs and constraints. In particular, the decline in self-employment during the Great Recession may partly reflect financing constraints. Going forward, we thus focus on the intensive (hours) margin for the self-employed.

\(^{13}\)Hipple (2010) reports annual rates of self-employment in the United States for 1994 to 2009 based on the monthly CPS, and these series are extended through 2012 (based on Hipple input) by Heim (2014). The Hipple series includes both incorporated and unincorporated self-employed. Incorporated self-employed constitute about one-third of total self-employed. The numbers reported here are for non-agriculture, which represents 94 percent of the self-employed. For agriculture, self-employment (again from Hipple) is acyclical. We focus on non-agriculture workers because top- and bottom-coding of income in the CPS is extreme for farmers. For farmers, it is also implausible to treat realized income as known at the time that labor input is chosen, an assumption implicit in calculations of the labor wedge.
We base our analysis on the Annual Social and Economic supplements to the CPS, typically referred to as the March CPS. In the March supplement, household members report their hours and income for the previous calendar year. They also report the income and class of workers at their primary job, the job held longest during the prior year. The class-of-worker variable allows us to distinguish the self-employed separately for agriculture and non-agriculture. We begin our sample in 1987, the first year that data on primary-job income are available. Advantages of the March supplement are: (i) it is large; (ii) its top-coding of income is less extreme than in the monthly surveys; and (iii) some households can be matched across two consecutive March surveys, allowing us to examine year-over-year changes for a given set of workers. Our unmatched sample contains 197,723 self-employed individuals for 1987 to 2012 (as well as 1,901,936 wage earners).14

Figure 2 reports usual weekly hours and total annual hours worked separately for the self-employed (nonagricultural) and for those earning wages and salaries for 1987 to 2012. The intensive margin is clearly more cyclical for the self-employed. If we regress hours per week on real GDP (both series are in logs and HP-filtered), the elasticity for the self-employed (0.37, standard error 0.13) is nearly twice that for wage earners (0.20, s.e. 0.02). During the Great Recession (2007–2009), the workweek for the self-employed declined by 4.9 percent (two full hours) compared with only 1.7 percent for wage earners. Similarly, annual hours declined by 6.9 percent for the self-employed, compared with 3.2 percent for wage earners.

Figure 2 might be influenced by composition bias. For example, if workers becoming self-employed in expansions work more hours than the typical self-employed,

---

14 We require that workers be between ages 20 and 70 and work at least 10 hours per week and at least 10 weeks during the year. Some income and hours responses are top-coded. In each year we trim the top and bottom 9.6 percent of observations by income on primary job; 9.6 percent is just large enough to remove the top-coding of business income for the self-employed in all 26 years. We trim the bottom for symmetry; this also serves to remove all negative income entries. Usual hours are top-coded at 99 per week. We trim the top 1.2 percent of workers by weekly hours. This is the minimal trimming that removes top-coded hours for all years. We also report results below making the minimal trims necessary in each year of matched data. This reduces the average sample share trimmed by top coding of earnings from 9.6 to 4.9 percent, and for the workweek from 1.2 to 0.5 percent. (We still trim the bottom of earnings distribution symmetrically.) Our results are essentially unaffected.
then hours in Figure 2 will have a procyclical bias. For this reason, we match self-employed individuals across consecutive March supplements, constructing growth rates for their hours and income. Using these growth rates, we express hours and income relative to 1987. We are not able to match across 1994 and 1995 calendar years because of a CPS sample redesign. For 1994–1995 we impute to each series its mean growth rate. We then create a level series indexed to 1987. In all subsequent statistics, we exclude years 1994 and 1995. Our matched sample includes 39,306 self-employed individuals, prior to trimming to deal with top-coding.

When we compare these hours indices for the self-employed and wage earners, we see that the workweek is more cyclical for the self-employed. The elasticity of the workweek with respect to real GDP (both variables are HP-filtered) is 0.28 (s.e. 0.07) for the self-employed versus 0.17 (s.e. 0.03) for wage earners. Annual hours are slightly more cyclical for wage earners, with an elasticity with respect to real GDP of 0.57 (s.e. 0.08) compared with 0.54 (s.e. 0.13) for the self-employed.

Similar remarks apply if we measure the cycle by aggregate hours.

In Table 5 we report the cyclicality of the intensive margin labor wedge. The first column is estimated for all workers, not just the self-employed. It repeats the analysis from equation (2), but uses workweek fluctuations constructed from the matched-CPS surveys. It is also annual rather than quarterly and excludes 1994 and 1995 because we cannot match those years in the CPS. It dispenses with the tax wedge, which we found to have little impact. As in Section I, we find a strongly countercyclical wedge. The elasticities of the wedge with respect to real GDP and aggregate hours, −1.87 and −1.20, are modestly smaller than reported in Table 2 (estimates there being −1.91 and −1.38), with the difference reflecting a slightly less procyclical workweek.

Columns 2–4 of Table 5 show how the wedge’s cyclicality changes as we sequentially replace the estimates of cyclicality in hours, productivity, and consumption for
all workers with estimates for the self-employed. Column 2 constructs the wedge using fluctuations in the workweek just for the self-employed, maintaining the same aggregate series for productivity and consumption. Not surprisingly, given the greater cyclicality of the workweek for the self-employed described previously, this results in a modestly more cyclical labor wedge. The elasticity with respect to real GDP goes from $-1.87$ (s.e. 0.09) to $-2.06$ (s.e. 0.15).

We next replace aggregate labor productivity with a measure of productivity specific to the self-employed. Our measure is self-employed annual business income divided by their annual hours worked.\(^1\) Self-employed business income per hour is proportional to the marginal product of self-employed labor assuming: (i) a constant elasticity of output with respect to self-employed labor, and (ii) self-employed income is proportional to business output. This second assumption is likely to be violated: self-employed income probably includes some returns to equity in the business, not just self-employed labor. This should lead us to overstate the procyclicality of the marginal product of labor for the self-employed, because their residual equity claim on business revenue is likely to be more procyclical than business output. This means we probably understate the countercyclicality of the self-employed labor wedge.

An additional factor that works to understate the countercyclicality of the labor wedge is that our self-employed productivity measure ignores any overhead component of self-employed labor. This could be especially important given the small scale of operations for much self-employed production. A final concern is that reported income could misstate actual income. The self-employed tend to understate income. Hurst, Li, and Pugsley (2014) find the ratio of consumption to income is higher for the self-employed, consistent with the self-employed understating income. The concern for us would be if the self-employed underreport at a lower rate in recessions.

Going from column 2 to 3 of Table 5, we replace aggregate labor productivity with self-employed business income per hour. Aggregate labor productivity has been modestly countercyclical since 1987, with an elasticity with respect to real GDP of $-0.21$ (s.e. 0.05). Self-employed business income per hour has been less cyclical (elasticity $-0.13$, s.e. 0.16). Thus, the estimated labor wedge becomes slightly less countercyclical, with an elasticity of $-1.97$ (s.e. 0.21) with respect to real GDP. In summary, the wedge calculated with measured productivity and hours for the self-employed is just as cyclical as that for all workers.\(^2\) Figure 3 plots the time series of these two wedges.

We have assumed the cyclicality of consumption for the self-employed is the same as for consumption per capita. For robustness, we estimate self-employed consumption relative to aggregate consumption based on quarterly growth rates in household spending on nondurables and services in the Consumer Expenditure Surveys (CE). We add these estimates of relative consumption to aggregate consumption to obtain an estimate of consumption for the self-employed.

---

\(^1\) We deflate business income by the nondurables and services PCE deflator. We calculate the percentage change in self-employed income across matched years by the midpoint formula. This avoids extreme values for individuals with very low income in one of the years.

\(^2\) We reestimate for column 3 making the minimal trims necessary to remove top-coding in any given matched year of CPS data, rather than trimming the same shares in all years. This yields a trimmed share that averages about 10 percent. Our results are little affected. The elasticity of the self-employed wedge with respect to real GDP goes from coefficient $-1.97$ to $-1.96$ (s.e. 0.25). The elasticity with respect to total hours goes from $-1.29$ to $-1.20$ (s.e. 0.14).
The elasticity of aggregate consumption with respect to real GDP is 0.64 (s.e. 0.02). Self-employed consumption is even more procyclical, with an elasticity of 1.27. But the standard error is too large, at 0.51, to reliably infer that the self-employed have more procyclical consumption. The big standard error reflects the small number of self-employed observations in the CE. If we do use this measure of consumption, however, we get an even more cyclical wedge for the self-employed. This is illustrated in the last column of Table 5. The self-employed labor wedge now exhibits an elasticity of $-3.23$ (s.e. 0.96) with respect to real GDP. In this section’s subsequent exercises, we revert to measuring self-employed consumption by aggregate consumption, rather than adopting such a noisy measure.\(^{20}\)

Table 6 presents three robustness exercises.\(^{21}\) First, self-employed who are incorporated might take income in the form of corporate profits rather than business income. It is not obvious how the incorporated self-employed treat these profits in answering the CPS question about their business income. For this reason, as an alternative measure of labor productivity, we consider business income per hour excluding the incorporated self-employed. This series is more procyclical than business income per hour for all self-employed. Its elasticity with respect to real GDP is 0.28 (s.e. 0.27), whereas the measure for all self-employed is slightly countercyclical. The first column of Table 6 repeats column 3 from Table 5, while column 2

---

\(^{20}\)We also examined the cyclicality of consumer expenditure for the self-employed in the Panel Study of Income Dynamics (PSID). The PSID has reasonably broad expenditure measures starting in 1999, but only biannually. In the PSID, self-employed consumption relative to consumption of all households dropped by 2.8 percent from 2007 to 2009, corresponding to the most recent recession. (Unfortunately, biannual observations miss the timing of the eight-month recession during 2001.) But, overall, the PSID numbers suggest similar cyclicality of consumption for the self-employed as for all households.

\(^{21}\)In addition to these exercises, we checked how our results change if the self-employed supply labor more elastically or are more willing to substitute consumption across time than are other workers. With a Frisch elasticity of 1.0 rather than 0.5, the elasticity of the self-employed wedge with respect to real GDP is $-1.69$. With an intertemporal elasticity of substitution of 1.0 rather than 0.5, the self-employed wedge cyclical elasticity is $-1.33$. In both cases, the elasticity of the self-employed wedge is a significant portion of the overall wedge ($-1.87$).
of Table 6 measures productivity by business income per hour for those not incorporated. The wedge becomes less countercyclical, with an elasticity with respect to real GDP of \(-1.57\) (s.e. 0.22). Nevertheless, the self-employed labor wedge remains nearly as cyclical as that for all workers (Table 5, column 1).

A second robustness exercise considers that the self-employed are distributed differently across industries than are wage earners. For instance, self-employment is about twice as frequent in construction, a highly cyclical industry, than overall. Self-employment is less common in durable manufacturing, which is also highly cyclical. We constructed a self-employed labor wedge reweighting observations by industry so that the weighted shares of the self-employed by industry mimic those for all workers. We do this for 12 major industries. For example, if self-employment is twice as frequent in construction, then the self-employed in construction are down-weighted by a factor of one-half. The results are given in column 3 of Table 6. The cyclicality of the self-employed wedge is modestly reduced. The elasticity is now \(-1.64\) (s.e. 0.29) with respect to real GDP. Again, the wedge remains nearly as cyclical as for all workers.\(^{22}\)

Another possible concern is that some self-employed might contract with just a single buyer. If so, perhaps they are quasi-employees, with their price setting resembling wage setting. For many self-employed we believe this depiction is implausible, for instance owners of repair shops or professionals such as plumbers and doctors. As a barometer of whether a self-employed person is likely to be independent rather than a quasi-employee, we use whether they have employees of their own. Our presumption is that a self-employed person with their own workforce is less likely to be a quasi-employee. Beginning with the 2014 March CPS supplement, the self-employed are asked if they have paid employees.\(^{23}\) Based on these 2014 and 2015 CPS supplements, we contrast the fraction of self-employed with employees across the 12 major industries, further broken down by three occupational classifications. These are (i) managers and proprietors, (ii) professionals, craftsmen,

\(^{22}\) As discussed in Section I, if the calibration is misspecified by ignoring countercyclical home production, this can impart some cyclicity to the labor wedge (Karabarbounis 2014b). Might this misspecification fall disproportionately on the self-employed? This would require that the self-employed have a comparative advantage in home production. Since the self-employed exhibit higher average market earnings, they would need to be even more able at home production.

\(^{23}\) We merge the 2014 and 2015 March supplements with the basic March files, so the sample here reflects self-employed whose survey months 4 and 8 coincide with the March supplement. The sample size is 2,121.
and operatives, and (iii) clerical, sales, service workers, and laborers. Overall, 27 percent of the self-employed report having employees. It is highest in retailing at 47 percent. (Across industry/occupations, it varies from 7 percent to 70 percent.)

For the last column of Table 6, we reweight the self-employed for our 1987 to 2012 sample, based on their industry/occupation cell, by the share of self-employed with employees in those cells in the 2014–2015 March supplements. That is, we weight observations only to the extent they reflect industries/occupations where the self-employed employ workers. The cyclicality of the self-employed wedge is reduced, but quite modestly. The elasticity is −1.64 (s.e. 0.40) with respect to real GDP, nearly as cyclical as for all workers from Table 5.

We conclude that the self-employed exhibit a highly countercyclical labor wedge. Depending on the specification, it is either as cyclical as the wedge for all workers or nearly so. Because this wedge is presumably not driven by wage or other labor market distortions, it is evidence of highly countercyclical price markups. By extension, we find it suggestive of countercyclical price markups for the overall economy, assuming the cyclicality of price markups is similar for owner/operators as for other firms.

### III. Intermediate Inputs

The conventional way to estimate price markups ($\mu^P$) is based on the inverse labor share of income, such as Karabarbounis (2014a). But any input with a well-measured marginal product and marginal price can be used to infer marginal cost and thus price markups. Here we investigate intermediate inputs: materials, energy, and services.

Intermediate inputs are promising for several reasons. First, intermediates are used by all industries. Second, adjustment costs for intermediates are believed to be low relative to adjustment costs for capital or even labor (see Basu 1995; Levinsohn and Petrin 2003; and De Loecker and Warzynski 2012). Third, the assumption of no overhead component seems more defensible for intermediates than for labor.

One question is whether intermediate prices reflect the marginal cost of intermediate inputs. Long-term relationships between firms and suppliers could raise the same implicit contracting issues that arise with labor. Still, intermediates offer an independent piece of evidence vis-à-vis labor. And, as with labor, one would expect price smoothing relative to true input costs to impart a procyclical bias to the estimated $\mu^P$. See Carlton (1986) for evidence and arguments along this line.

#### A. Technology for Gross Output

We assume a CES production function for gross output in an industry:

$$y = \left[ \theta m^{1-\frac{1}{\varepsilon}} + (1 - \theta) \left[ z_v \left[ \alpha k^{1-\frac{1}{\omega}} + (1 - \alpha) (z_n n)^{1-\frac{1}{\omega}} \right]^{\frac{\mu}{\omega - 1}} \right]^{1-\frac{1}{\varepsilon}} \right]$$

---

24 As shown in the online Appendix, this finding is robust to detrending our data using Hamilton’s (forthcoming) method rather than the HP-filter.

25 In any case, an overhead component of intermediates would actually reinforce our findings, just like overhead labor would do. With overhead inputs, variable inputs are more procyclical than total inputs, so that marginal cost is more procyclical than suggested by total inputs, and price markups, in turn, are more countercyclical.
where \( y \) denotes gross output, \( m \) intermediate inputs, \( k \) capital, and \( n \) labor. Technology shocks can be specific to value added \((z_v)\) or labor \((z_n)\). The elasticity of substitution between intermediates and value added is \( \varepsilon \), and the elasticity of substitution between capital and labor within value added is \( \omega \).

With this technology, the marginal product of intermediates is

\[
\frac{\partial y}{\partial m} = \theta \left( \frac{y}{m} \right)^{\frac{1}{\varepsilon}}.
\]

Based on this marginal product, we can estimate price markups as

\[
\mu^p = \frac{p}{p_m} \frac{\partial y}{\partial m} = \frac{p\theta \left( \frac{y}{m} \right)^{\frac{1}{\varepsilon}}}{p_m}.
\]

In the special case of Cobb-Douglas aggregation of intermediates and value added \((\varepsilon = 1)\), the price markup is the inverse of intermediates’ share:

\[
\mu^p = \frac{\theta p y}{p_m m}.
\]

A higher price-cost markup boosts gross output relative to spending on intermediates. This is analogous to using inverse labor’s share to measure price markup movements. Note, one must assume Cobb-Douglas production in labor and intermediates to estimate markups from the inverse labor share in value added, as we did in Section I. If production is Cobb-Douglas in intermediates and value added, then a countercyclical markup would show up as a procyclical intermediate inputs share.

Although an overhead component to intermediates is unlikely, one may still be concerned about fixed costs of production. Suppose \( y_t = f_t - \Gamma \), where \( f_t \) is our CES production function and \( \Gamma \) is a fixed cost. In the Cobb-Douglas case, this would imply \( \mu^p_t = \theta \frac{p y_t}{p_m m} \left( 1 + \frac{\Gamma}{y_t} \right) \). Thus, if fixed costs are important, our estimates will underestimate the countercyclicality of price markups.

**B. Evidence on the Cyclicality of Intermediate Inputs**

We use the BLS Multifactor Productivity Database on industry gross output and KLEMS inputs (capital, labor, energy, materials, and services). The KLEMS dataset is annual from 1987 to 2012, covering 60 industries, including 42 outside of manufacturing. Other studies of the cyclicality of intermediates (Basu 1995; Vaona 2016; and Kim 2015) have focused on manufacturing, but manufacturing constitutes only about 12 percent of US GDP. Furthermore, the KLEMS data reflect significant revising of industry GDP accounts by the Bureau of Economic Analysis (BEA), starting about a decade ago, to much more reliably divide industry output between value added and intermediate purchases.\(^{26}\)

---

\(^{26}\)The KLEMS dataset incorporates data from a number of BLS and BEA programs. See *BLS Handbook of Methods*, chapter 11 (http://www.bls.gov/opub/hom/pdf/homch11.pdf) for a description of methodology. Thanks to improvements in census data collection on output and intermediates, in August 2010 KLEMS data became available beyond manufacturing. See Harper et al. (2010) for details. The comprehensive revision of BEA methods, improving measures of intermediates, is outlined in Moyer et al. (2004). These improvements are not captured
Figure 4 plots the weighted-average KLEMS industry intermediate share from 1987 to 2012. It fluctuates but exhibits no clear trend. If the production elasticity for intermediates is stable, this is consistent with no trend in the price-cost markup. This is somewhat surprising in light of De Loecker and Eeckhout (2017), who infer sharply rising markups in recent decades. Their sample is Compustat firms, whereas KLEMS is based on listed and unlisted firms alike. Moreover, our evidence is for intermediate share alone, whereas theirs is for a combination of labor and intermediate share (cost of goods sold).

Figure 5 plots the weighted-average KLEMS industry intermediate share against GDP, where both variables are in logs and HP-filtered. As shown, spending on intermediates relative to gross output is highly procyclical in the KLEMS data. In the Cobb-Douglas case of $\varepsilon = 1$, a procyclical intermediate share implies a countercyclical markup.

To explore the cyclicality of markups more systematically, we next run regressions of the intermediate-based price markup on the cycle. Based on equation (7),

$$\log \left( \frac{p_{it}}{p_{mit}} \left( \frac{y_{it}}{m_{it}} \right)^{\varepsilon} \right) = \alpha_i + \beta \log(cyc_t) + \varepsilon_{it},$$

where $cyc_t$ is either real GDP or hours worked, and all variables are HP-filtered. The industry fixed effects ($\alpha_i$) should take out changes in the aggregate share due to shifting industry composition over the cycle. We weight industries by the average share of their value added in all industry value added from 1987 to 2012. Standard errors are clustered by year.

In data based solely on the Annual Surveys of Manufacturing (ASM), employed in a number of papers. KLEMS is also preferred to ASM because its division of gross output between value added and intermediates adjusts for intra-industry purchases and better reflects accumulation of inventories.
Table 7 presents the results. Consider first the case of $\varepsilon = 1$, for which production is Cobb-Douglas in intermediates and value added. The price markup is estimated to be highly countercyclical. This is true for both measures of the cycle (based on GDP or total hours worked), for all industries together, and separately for manufacturing and nonmanufacturing industries. Though not reported in the table, it is also true if we weight by annual Tornqvist value-added shares rather than shares over the entire sample, and if we use growth rates rather than HP-filtered series.

For manufacturing, we can break intermediate inputs into materials, energy, and services. As Table 7 shows, the inverse shares for materials and energy are both countercyclical, and significantly so. In contrast, the inverse share of spending on services is procyclical, perhaps because services are contracted less in spot markets than materials or energy.\footnote{Outside manufacturing, we can only break intermediates into these components for the period 1997–2012. For all industries together, the inverse services share is acyclical, whereas the inverse shares of materials and energy are countercyclical.}

Table 7—Cyclicality of Price Markups Using Intermediates

<table>
<thead>
<tr>
<th>Elasticity w.r.t.</th>
<th>$\varepsilon = 1$</th>
<th>$\varepsilon = 0.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDP (0.24)</td>
<td>Hours (0.15)</td>
</tr>
<tr>
<td>All industries</td>
<td>$-0.94$ ($0.24$)</td>
<td>$-0.59$ ($0.15$)</td>
</tr>
<tr>
<td>Nonmanufacturing</td>
<td>$-0.76$ ($0.21$)</td>
<td>$-0.70$ ($0.19$)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>$-0.76$ ($0.21$)</td>
<td>$-0.70$ ($0.19$)</td>
</tr>
<tr>
<td>Materials</td>
<td>$-1.24$ ($0.63$)</td>
<td>$-1.15$ ($0.40$)</td>
</tr>
<tr>
<td>Services</td>
<td>$1.49$ ($0.68$)</td>
<td>$0.93$ ($0.43$)</td>
</tr>
<tr>
<td>Energy</td>
<td>$-1.97$ ($1.05$)</td>
<td>$-1.58$ ($0.69$)</td>
</tr>
</tbody>
</table>

Notes: Each entry is from a separate regression. Annual data are from 1987 to 2012 for 60 industries (1,560 industry-year observations): 18 manufacturing and 42 nonmanufacturing. All variables are in logs and HP-filtered. Regressions include industry fixed effects and use industry average value-added shares as weights. Standard errors are clustered by year.

Figure 5. Cyclicality of Intermediate Share
It is often argued that it is tough to substitute between intermediates and value added. Bruno (1984) and Rotemberg and Woodford (1996) estimate elasticities of 0.45 and 0.69, respectively, for US manufacturing. Oberfield and Raval (2014) obtain estimates ranging from 0.63 to 0.90 by looking across regions in US manufacturing, while Atalay (2017) estimates an elasticity of 0.5, but with large standard errors. We therefore report results using $\varepsilon = 3/4$ in Table 7. A smaller elasticity makes the $\mu^p$ based on intermediates more countercyclical. Because firms shift toward intermediates in booms, the marginal product of intermediates falls faster if substitutability is limited, making marginal cost more procyclical. Thus, the price-cost markup implied by intermediate inputs becomes more countercyclical.

In the KLEMS data, both the price and quantity of intermediates are procyclical relative to labor, if one uses average hourly earnings (AHE) as the price of labor. The elasticity of $p_{m/n}$ with respect to real GDP is 0.65 (s.e. 0.16), while the elasticity of $m/n$ with respect to real GDP is 1.19 (s.e. 0.37). Why do firms shift toward intermediates over labor in booms, if intermediates become relatively expensive? One answer is that AHE understates cyclicality in the price of labor, perhaps because of wage smoothing, with labor’s price even more procyclical than that of intermediates. That implies a much more procyclical marginal cost and, in turn, a much more countercyclical markup than judged by AHE.

What share of the labor wedge might be accounted for by the $\mu^p$ we obtain from intermediates? To answer this, we construct an industry-specific labor wedge that is consistent with our gross output production function. We replace aggregate labor productivity ($v/n$) with nominal gross output per hour in each KLEMS industry (relative to the consumption deflator). We also consider preferences that allow for an industry-specific marginal rate of substitution $^{28}$ The industry-$i$ (intensive margin) labor wedge is thus

$$\ln(\mu^i) = \ln\left(\frac{p_i \text{mpn}_i^\text{int}}{\text{pmr}_i^\text{int}}\right) = \ln\left(\frac{p_i v_i}{p_m v_m}\right) + \ln\left(\frac{y_i}{V_i}\right) - \frac{1}{\eta} \ln\left(\frac{h_i}{h}\right) + \ln\left(\frac{\text{mpn}_i^\text{int}}{\text{mr}_s^\text{int}}\right).$$

This industry labor wedge differs from the aggregate labor wedge in three possible ways. Value added per hour could be more or less cyclical in the KLEMS industries than GDP per hour (the first term on the right side). The cyclicality of gross output may differ from value added (the second term). Finally, hours worked per worker, and thus the $\text{mr}_s$, could be more (or less) cyclical for the KLEMS industries than for all industries (the third term).

Table 8 presents the cyclical elasticities. The KLEMS industry wedge has a smaller elasticity ($−0.89$ with respect to GDP) than the aggregate IMW ($−1.91$) from Section I. Why? Nominal value-added labor productivity is more procyclical in the KLEMS industries than GDP per hour (cyclical elasticity with respect to GDP of 0.33), and gross output is more procyclical than value added (cyclical elasticity of $y/v$ with respect to GDP of 0.49). Workweeks, which are industry-specific and thus

$^{28}$ Preferences are $E_0 \sum_{t=0}^{\infty} \beta^t \left\{ \frac{c^{1-\sigma}}{1-1/\sigma} \right\} - \sum_i \left[ r^{h_i} \left( \frac{1}{1+1/\eta} + \psi \right) e_t \right\}$ and thus $\ln(\frac{\text{mr}_s^\text{int}}{\text{mr}_s^\text{int}}) = \frac{1}{\eta} \ln\left(\frac{h_i}{h}\right) + \ln\left(\frac{c}{\psi}\right)$.

In the online Appendix we consider alternative preferences, alternative technology (e.g., $\varepsilon \neq 1$), and the extensive margin.
affect the all-industry, manufacturing, and nonmanufacturing wedges in different ways, account for the remainder.29

Comparing Tables 7 and 8, the intermediates-based \( \mu^p \) accounts for essentially all of the cyclical labor wedge.30 Figure 6 provides visual corroboration by plotting the weighted-average industry (inverse) intermediate share against the labor wedge, constructed for \( \varepsilon = 1 \).31

When we consider \( \varepsilon = 0.75 \), \( \mu^p \) becomes more countercyclical in Table 7. At the same time, a smaller \( \varepsilon \) makes the labor wedge less countercyclical because firms shift away from value added (labor) in booms. Thus, price-cost markups loom even larger relative to the labor wedge if intermediates and value added have more limited substitutability.

The results above are at the level of 60 industries, reflecting the industry makeup of the KLEMS dataset. We see the KLEMS data as the preferred source for examining the cyclicality of intermediates because they extend beyond manufacturing and incorporate the highest quality industry information, especially on intermediates. One question that arises is whether the strong cyclicality in the intermediate share occurs at the plant level, or instead reflects a compositional shift of output toward plants with high intermediate shares during booms. For manufacturing, plant-level statistics from the US Longitudinal Research Database (LRD), which is based on the Census of Manufacturing and Annual Survey of Manufacturing, are reported in Garcia-Macia, Hsieh, and Klenow (2016). Though not directly comparable to KLEMS, one can check whether high intermediate share plants are more cyclical in the LRD moments.

Across plants \( j \) within each industry \( i \), one can construct the standard intermediate share as

\[
s_{it} = \frac{\sum_{j \in i} p_{yjt} m_{jt}}{\sum_{j \in i} p_{yjt} y_{jt}} = \frac{\sum_{j \in i} p_{yjt} y_{jt} \cdot p_{yjt} m_{jt}}{\sum_{j \in i} p_{yjt} y_{jt} \cdot p_{yjt} y_{jt}}.
\]

29 Because some of our industries only have workweek data starting in 1990, we use the aggregate average workweek from 1987 through 2012, which had an elasticity with respect to GDP of 0.32 (s.e. 0.03), adjusted by the relative elasticity of industry-specific workweeks to the aggregate from 1990 through 2012. These elasticities are 0.32 (s.e. 0.03) for the aggregate, 0.22 (s.e. 0.04) for all KLEMS industries, 0.41 (0.08) for manufacturing KLEMS industries, and 0.16 (0.04) for nonmanufacturing KLEMS industries. Online Appendix Table A1 reports the cyclicality of the labor wedge using a common workweek for all industries (i.e., omitting the industry-specific adjustments).

30 As shown in the online Appendix, this finding is robust to detrending our data using Hamilton’s (forthcoming) method rather than the HP-filter.

31 The labor wedge in Figure 6 is constructed using the aggregate average workweek. Using industry-specific workweeks produces similar plots beginning in 1990 rather than 1987.

### Table 8—Cyclicality of the Intensive Margin Labor Wedge

<table>
<thead>
<tr>
<th></th>
<th>Elasticity w.r.t.</th>
<th>Total hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GDP</td>
<td></td>
</tr>
<tr>
<td>All industries</td>
<td>−0.89 (0.26)</td>
<td>−0.59 (0.13)</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>−0.72 (0.39)</td>
<td>−0.39 (0.20)</td>
</tr>
<tr>
<td>Nonmanufacturing</td>
<td>−0.93 (0.24)</td>
<td>−0.65 (0.12)</td>
</tr>
</tbody>
</table>

Notes: Each entry is from a separate regression. Annual data are from 1987 to 2012 for 60 industries (1,560 industry-year observations): 18 manufacturing and 42 nonmanufacturing. All variables are in logs and HP-filtered. Regressions include industry fixed effects and use industry average value-added shares as weights. Standard errors are clustered by year.
For contrast, one can construct a fixed-composition intermediate share as

$$\bar{s}_{it} = \sum_{j \in i} \frac{p_{yjt}}{p_{yit}} \cdot s_{jC},$$

where $s_{jC}$ denotes the average intermediate share for plant $j$ across the census years in which it operates. Note that $s_{jC}$ only uses the most reliable LRD components, namely intermediates in census years and gross output in all years. 32 If gross output shares are no more cyclical for high intermediate share plants, then $\bar{s}_{it}$ should not project on the business cycle.

The LRD moments for 1987–2010 include $\bar{s}_{it}$ for 17 industries that closely correspond to KLEMS industries. A regression of $\bar{s}_{it}$ on the cyclical component of log GDP, controlling for industry effects, shows that the fixed composition intermediate shares are indeed procyclical. Comparing these estimates to Table 7, about one-third of the procyclical intermediate share at the industry level in manufacturing reflects plant composition, and the remaining two-thirds can be attributed to within-plant movements. If we remove this compositional component, we still find that the lion’s share of cyclicality in the labor wedge for manufacturing can be ascribed to price markups.

IV. Discussion and Relationship to Literature

How does our work relate to other attempts at measuring the cyclicality of price markups? The challenge is capturing cyclicality in the marginal cost of production. Researchers must make assumptions about firm production functions in order to infer marginal cost from quantities and prices of inputs and output. Marginal cost

32 Other than the census years 1992, 1997, 2002, and 2007, LRD coverage of intermediate inputs does not include purchased services. Identification of intermediate inputs in non-census years is considered sparse and of lower quality (Moyer et al. 2004).
should be equated across input margins, so one can consider the cost of increasing output via any input. Many studies have focused on labor: for example, Bils (1987); Rotemberg and Woodford (1999); and Nekarda and Ramey (2013).

Labor’s share of income, the average price of labor divided by its average product, often serves as the baseline measure of marginal cost, with corrections made to address concerns that average prices and products may not equal marginal ones. We argued in Section I that using wage data to infer the marginal price of labor may be especially fraught with difficulty, and subsequently eschewed wage data altogether by examining self-employed labor as well as intermediate inputs.

Rotemberg and Woodford (1999) pointed out that one could use intermediate inputs to infer markups. Basu (1995) found that quantities of intermediate inputs rose relative to real output in expansions, but did not explore relative price movements. The work of Murphy, Shleifer, and Vishny (1989) on relative price movements for broad categories suggested the relative price of intermediates was likely to be procyclical. By looking at the intermediate share of income, we combine quantities and prices to obtain a measure of marginal cost in Section III. We also consider a more recent sample and more disaggregated industries.

Like us, Vaona (2016) uses total intermediate inputs to estimate markups. His sample is 1959–1996 US manufacturing industries, and he estimates the marginal product of intermediates nonparametrically. He emphasizes the response of industry markups to the industry cycle. Our focus is on how average industry markups respond to the aggregate cycle. We provide evidence beyond manufacturing, adjust for the composition of industries (and even plants within manufacturing), and compare it to the overall labor wedge quantitatively.

Kim (2015) estimates markup fluctuations from 1958 to 2009 for manufacturing industries from spending on energy intermediates. Energy is a fairly small component of our KLEMS intermediates: for 2010, it constituted only 4.4 percent of total intermediate spending in manufacturing, 6.2 percent in other industries. For his sample, Kim estimates that markups increase with the industry cycle. He includes year fixed effects, so his results are silent on the movement of markups over the aggregate cycle. If we similarly consider only energy spending in manufacturing, for our 1987–2012 sample we find energy’s share of gross output is procyclical and estimate a countercyclical price markup across a wide range of elasticities of substitution.

Other researchers have also used approaches that do not require wage data. Galeotti and Schiantarelli (1998) measure marginal cost using capital inputs. Like our approach, this requires assumptions on the production function to infer a marginal product (of capital, in this case). In addition, they must take a stand on the stochastic discount factor, because the marginal price (i.e., rental rate) of capital is not directly observable but depends on both today’s acquisition price of capital and tomorrow’s expected discounted price. They find evidence of countercyclical markups.

Hall (2014) considers a model in which advertising shifts the static demand curve faced by a firm. If the markup is high, then the firm will want to advertise more. The

---

33 They incorporate adjustment costs, requiring a further functional form assumption.
model implies that the ratio of advertising expenditures to revenue is proportional to the markup. Since the advertising expenditure share is acyclical in the data, Hall concludes that markups are also acyclical.

Alternative models of advertising will have different implications for markups. In the online Appendix we show that, if advertising affects consumer reservation prices rather than shifting quantity demanded, then changes to the price elasticity of demand have no effect on the advertising expenditure share, but do cause markup changes. A second alteration that breaks the tight contemporaneous link between advertising expenditures and markups is when advertising affects future demand (e.g., Bagwell 2007). Some evidence that advertising is a durable investment is provided by Campello, Graham, and Harvey (2010). They report that planned marketing expenditures fell much more for financially constrained firms than for unconstrained firms during the Great Recession.34

Another approach, pursued by Bils and Kahn (2000) and Kryvtsov and Midrigan (2013), exploits the close theoretical relationship between markups and finished goods inventories. Consider a firm’s decision to add a unit of finished inventories. If the additional unit is sold in the current period, the net benefit is the markup of price over marginal cost. If the additional unit is not sold, the net benefit is the expected discounted ratio of future to current real marginal cost (the firm will not have to produce as much in the future). If the markup is low, holding all else equal, the benefit of holding inventories is smaller and the firm will reduce its inventories relative to sales.

Note, however, that empirically all else might not be equal. In particular, both papers must proxy for how firms’ discount rates vary over the cycle. It also necessary to capture any cyclicality in the elasticity of sales with respect to inventories. Importantly, both studies assume \( \frac{\partial \text{sales}}{\partial \text{inventories}} \) is a time-invariant function of the sales-to-inventory ratio.35 Given their assumptions and a highly procyclical sales-to-inventory ratio in the data, both studies conclude that markups must be countercyclical.

In the online Appendix we consider work-in-process (WIP) inventories and again infer countercyclical markups. The intuition for the relationship between inventories and markups is similar to that described above. If markups are high (i.e., real marginal costs low) relative to the future, a firm should shift production from tomorrow to today and increase its stock of WIP inventories. The WIP framework is somewhat simpler than that for finished goods: only the relative (intertemporal) markup appears rather than both the relative markup and the level of the markup. Also, one does not need to take a stand on how inventories affect sales, but instead on how WIP inventories enter the production function.

In summary, these other nonwage approaches to measuring price markups, with the possible exception of Hall (2014), yield results broadly consistent with our own: namely, countercyclical markups. However, they all involve dynamics, requiring one
to measure any adjustment costs and the stochastic discount factor. Our self-employed and intermediates approaches, on the other hand, require only static measurements.

V. Conclusion

Hours worked fall more in recessions than can be explained by optimal changes in labor supply in response to real labor productivity. This labor wedge could reflect distortions in the labor market (e.g., sticky wages) and/or distortions in the product market (e.g., sticky prices).

Research has increasingly focused on problems in labor markets, in particular for firms hiring workers. Using average hourly earnings, the labor wedge seems to arise between the cost of labor to firms and the value of jobs to workers. But this inference could be mistaken if the true cost of labor to firms is more cyclical than average hourly earnings. If labor’s price is measured instead by the wages of new hires or the user cost of labor, then the resulting labor wedge occurs mostly between the cost of labor and real labor productivity.

To bring new evidence to bear on this debate, we estimated the product market component of the labor wedge without relying on workers’ wages. First, we looked at the self-employed. The labor wedge appears nearly as cyclical for the self-employed as for wage earners, even though sticky wages and matching frictions should not be barriers to the self-employed working more hours. The hours of the self-employed appear to fall in recessions because of difficulty, or reluctance, in selling their output (for example, because of sticky prices). Second, we presented evidence on intermediate inputs. In recent recessions, prices rise relative to the level of marginal cost we infer from intermediate prices and quantities. Again, this suggests that firms face difficulty in converting production into revenue during recessions. We stress that these two approaches rely on completely different data sources (households versus industries).

Our point estimates imply that the labor wedge’s cyclical variation reflects product market distortions as opposed to labor market distortions. We cannot reject that labor market distortions matter, though they appear less important than has been inferred using data on average hourly earnings. Our evidence is consistent with a price of labor that is at least as cyclical as the new hire wage.

Our evidence does not determine the exact nature of these product market distortions, which is critical for informing stabilization policy. One explanation would be price stickiness that constrains sales. Another would be countercyclical desired markups (see Rotemberg and Woodford 1999 for a review). If producing has an investment component (e.g., the customer base model in Gilchrist et al. 2017), then tightening financial constraints could make firms raise prices relative to marginal cost in recessions as they cut such investments. Our evidence is also consistent with models where expanding production puts firms in a riskier position, and risk (or risk avoidance) heightens during recessions (e.g., Arellano, Bai, and Kehoe 2012).

Our findings are also relevant for the puzzle of unemployment’s high cyclical relative to labor productivity, the Shimer (2005) puzzle. A highly countercyclical price markup translates into strongly procyclical labor demand, beyond what might be attributed to labor productivity. It provides a rationale for firms to create less
employment in recessions without a decline in productivity, and even absent important wage stickiness.

REFERENCES


