Centre de Referència en Economia Analítica

Barcelona Economics Working Paper Series

Working Paper nº 45

Markups, Gaps, and the Welfare Costs of Business Fluctuations

Jordi Galí, Mark Gertler and J. David López-Salido

September 2003

Markups, Gaps, and the Welfare Costs of Business Fluctuations*

Jordi Galí †   Mark Gertler‡   J.David López-Salido§

Barcelona Economics WP n^{o} 45

September 2003
(first draft: May 2001)

Abstract

In this paper we present a simple, theory-based measure of the variations in aggregate economic efficiency associated with business fluctuations. We decompose this indicator, which we refer to as "the gap", into two constituent parts: a price markup and a wage markup, and show that the latter accounts for the bulk of the fluctuations in our gap measure. We also demonstrate the connection between our gap measure and the gap between ouput and its natural level, a more traditional indicator of aggregate efficiency. Finally, we derive a measure of the welfare costs of business cycles that is directly related to our gap variable. When applied to postwar U.S. data, for some plausible parametrizations, our measure indicates non-negligible welfare losses of gap fluctuations.

JEL Classification: E32

Key Words: Business Cycles, Countercyclical Markups, Welfare Costs

*We thank Jeff Amato, Andy Levin, Susanto Basu, Jonathan Parker and Bob Hall for helpful comments, as well as participants in seminars at LSE, Bocconi, IIES, UPF, Bank of Spain, Bank of England, NBER Summer Institute, CERGE, EUI, CEU, UAB, and the ECB Workshop on DSGE Models. Galí is thankful to the Bank of Spain, the MCYT, and DURSI for financial support. Gertler acknowledges the support of the NSF and the C.V. Starr Center.

†CREI and Universitat Pompeu Fabra.

‡New York University

§Bank of Spain

1 Introduction

To the extent that there exist price and wage rigidities, or possibly other types of market frictions, the business cycle is likely to involve inefficient fluctuations in the allocation of resources. Specifically, the economy may oscillate between expansionary periods where the volume of economic activity is close to the social optimum and recessions that feature a significant drop in production relative to the first best. In this paper we explore this hypothesis by developing a simple measure of aggregate inefficiency and examining its cyclical properties. The measure we propose - which we call "the inefficiency gap" or "the gap", for short - is based on the size of the wedge between the marginal product of labor and the marginal rate of substitution between consumption and leisure. Deviations of this gap from zero reflect an inefficient allocation of employment. By constructing a time series measure of the inefficiency gap, we are able to obtain some insight into both the nature and welfare costs of business cycles.

From a somewhat different perspective, we show that the inefficiency gap corresponds to the inverse of the markup of price over social marginal cost. Procyclical movements in the inefficiency gap accordingly mirror countercyclical movements in this markup. Our approach, however, differs from much of the recent literature on business cycles and markups by using the household's marginal rate of substitution between consumption and leisure to measure the price of labor, as opposed to wages.[1] As a matter of theory, of course, the household's consumption/leisure trade-off is the appropriate measure of the true social cost of labor. Wage data are not appropriate if either wages are not allocational or if labor market frictions are present that drive a wedge between market wages and the labor supply curve. As we demonstrate, our markup construct is highly countercyclical. In addition, it also leads directly to a measure of aggregate efficiency costs at each point in time.

[1] See Rotemberg and Woodford (1999) for a survey of the literature on business cycles and countercyclical markups.

Our approach builds on a stimulating paper by Hall (1997) that analyzes the cyclical behavior of the neoclassical labor market equilibrium. Specifically, Hall first demonstrates that the business cycle is associated with highly procyclical movements in the difference between the observable component of the household's marginal rate of substitution and the marginal product of labor. He then presents some evidence to suggest that this difference - which we refer to as the Hall residual - is of central importance to employment fluctuations. Also relevant is Mulligan (2002) who examines essentially the same measure of the labor market residual, though focusing on its low frequency movements. Specifically, he constructs an annual series of this variable, using data spanning more than a century. He finds that marginal tax rates correlate well at low frequencies with this labor market wedge.

As with Hall, we focus on the behavior of the labor market wedge at the business cycle frequency. We differ in several important ways, however. First, his framework treats this wedge simply as an exogenous driving force, interpretable for example as reflecting shifts in preferences.[2] We instead stress countercyclical markup variations as the key factor accounting for the cyclical fluctuations in this variable and present evidence in support of this general hypothesis. Second, given our "markup interpretation," we are able to use the Hall residual as the basis for a measure of the efficiency costs of business cycles.

From our gap variable it is possible to derive a measure of the lost surplus in the labor market at each point in time. Fluctuations generate efficiency costs on average because, as we show, the surplus lost from a decline in employment below its natural level exceeds the gain from a symmetric rise above its natural level. In this regard, our approach differs significantly from Lucas (1987, 2003) who examines the welfare costs

[2] To organize his approach, Hall (1997) modeled the labor market residual as an unobserved preference shock, though he did not take this hypothesis literally, but rather as a starting point for subsequent analysis. There has been a tendency in subsequent literature, however, (e.g. Holland and Scott (1998), Francis and Ramey (2001), Uhlig (2002)) to interpret this residual as an exogenous preference shock. Earlier literature as well offered a similar interpretation (e.g. Baxter and King, 1991). Our analysis will suggest that this residual cannot simply reflect exogenous preference shifts.

of consumption variability associated with the cycle. While the Lucas measure does not really take account of the sources of fluctuations, our measure instead isolates the costs associated with the inefficient component of fluctuations. In this regard, our metric may give a better sense of the potential gains from improved stabilization policy.

An equally important distinction is that our approach permits not only a measure of the costs of fluctuations on average, but also an assessment of the costs of particular episodes. We find, for example, that while efficiency costs of fluctuations are not large on average, they may be quite significant during major recessions, even after netting out the gains from the preceding boom. This consideration is highly relevant because it may be that the principle benefit from good stabilization policies may be avoiding severe recessions. To the extent that centrals banks have had either good skill or good luck in keeping to a minimum the number of severe downturns, it may be that on average the costs of fluctuations are not large. This kind of unconditional calculation, however, masks the kind of losses that can emerge if luck and/or skill suddenly turn bad. For this reason, an examination of episodes where matters clearly did seem to go awry can shed light on the importance of good policy management.

In section 2 we develop a framework for measuring the inefficiency gap in terms of observables, conditional on reasonably conventional assumptions about preferences and technology. We also show that it is possible to decompose the gap into price and wage markup components In section 3 we present empirical measures of this variable for the postwar U.S. economy. The inefficiency gap exhibits large procyclical swings. In addition, under the assumption that wages are allocational, most of its variation is associated with countercyclical movements in the wage markup.[3] The price markup shows, at best, a weak contemporaneous correlation. In section 4 we consider the possibility that purely exogenous factors (e.g. unobserved preference shifts) underlie

[3] In this respect our results are consistent with recent evidence in Sbordone (1999, 2000), Galí and Gertler (1999), Galí, Gertler and Lopez-Salido (2001) and Christiano, Eichenbaum and Evans (1997, 2001) that in somewhat different contexts similarly points to an important role for wage rigiditiy.

the variation in our gap measures. Specifically, we present evidence that suggests that the Hall residual is endogenous and thus cannot simply reflect exogenous variation in preferences. The evidence is instead consistent with our maintained hypothesis that endogenous variation in markups is largely responsible for the movement in the inefficiency gap. Section 5 characterizes both theoretically and empirically the link between the labor market distortion and the output gap. In Section 6 we then use this link to examine both the unconditional efficiency costs of recessions and the conditional costs associated with the major boom/bust episodes. Concluding remarks are in section 7.

2 The Gap and its Components: Theory

Let the *inefficiency gap* (henceforth, *the gap*) be defined as follows:

$$gap_t = mrs_t - mpn_t \tag{1}$$

where mpn_t and mrs_t denote, respectively, the (log) marginal product of labor and the (log) marginal rate of substitution between consumption and leisure.

As illustrated by Figure 1, our gap variable can be represented graphically as the vertical distance between the *perfectly competitive* labor supply and labor demand curves, evaluated at the current level of employment (or hours). In much of what follows we assume that our gap variable follows a stationary process with a (possibly nonzero) constant mean, denoted by gap (without any time subscript). The latter represents the steady state deviation between mrs_t and mpn_t. Notice that these assumptions are consistent with both mrs_t and mpn_t being nonstationary, as it is likely to be the case in practice as well as in the equilibrium representation of a large class of dynamic business cycle models.

We next relate the gap to the markups in the goods and labor markets. Under the assumption of wage-taking firms, and in the absence of labor adjustment costs, the nominal marginal cost is given by $w_t - mpn_t$, where w_t is (log) compensation per additional unit of labor input (including non-wage costs). Accordingly, we define the aggregate price markup as follows:

$$\mu_t^p = p_t - (w_t - mpn_t) \quad (2)$$

$$= mpn_t - (w_t - p_t) \quad (3)$$

The aggregate wage markup is given by:

$$\mu_t^w = (w_t - p_t) - mrs_t \quad (4)$$

i.e., it corresponds to the difference between the wage and the marginal disutility of work, both expressed in terms of consumption. Notice that the wage markup should be understood in a broad sense, including the wedge created by efficiency wages, payroll taxes paid by the firm and labor income taxes paid by the worker, search frictions, and so on.

There are a variety of frictions (perhaps most prominently, wage and price rigidities) which may induce fluctuations in the markups: it is in this respect that these frictions are associated with inefficient cyclical fluctuations, or more precisely, with variations in the aggregate level of (in)efficiency. In particular, given that the marginal rate of substitution is likely to be procyclical, rigidities in the real wage–resulting either from nominal or real rigidities–will give rise to countercyclical movements in the wage markup.[4] Similar rigidities may give rise, in turn, to a countercyclical price markup in response to demand shocks since, holding productivity constant, the marginal product of labor is countercyclical.[5] Alternatively, procyclical movements in

[4]Models with countercyclical wage markups due to nominal rigidities include Blanchard and Kiyotaki (1987) and Erceg, Henderson and Levin (2000). Alexopolous (2000) develops a model with a real rigidity due to effiiciency wages that can generate a countercyclical wage markup. Alternatively, Hall (1997) stresses the possible role of countercyclical search frictions to account for the behavior of the labor market residual.

[5]With productivity shocks, the markup could be procyclical (since the marginal product of labor moves procyclically in that instance).

competitiveness could induce a countercyclical price markup, as in Rotemberg and Woodford (1996), for example.

To formalize the link between markup behavior and the gap, we first express equation (1) as

$$gap_t = -\{[mpn_t - (w_t - p_t)] + [(w_t - p_t) - mrs_t]\} \tag{5}$$

Combining equations (3), (4), and (5) then yields a fundamental relation linking the gap to the wage and price markups:

$$gap_t = -(\mu_t^p + \mu_t^w) \tag{6}$$

In the steady state, further:

$$gap = -(\mu^p + \mu^w) < 0 \tag{7}$$

where variables without time subscripts denote steady state values.

It is natural to assume that $\mu_t^p \geq 0$ and $\mu_t^w \geq 0$ for all t, implying $gap_t \leq 0$ for all t. In this case the level of economic activity is inefficiently low (i.e., the gap is always negative), so that (small) increases in our gap measure will be associated with a smaller distortion (i.e., an allocation closer to the perfectly competitive one). Notice also that countercyclical movements in these markups imply that the gap is high in booms and low in recessions.

To the extent that we can measure the two markups (or, at least their variation), we can characterize the behavior of the gap, as well as its composition. Constructing our gap variable requires some assumptions about technology and preferences. Below we consider a baseline case with reasonably conventional assumptions. Decomposing the resulting gap variable between wage and price markups requires an additional assumption, namely, that the observed wages used in the construction of the markup reflect the shadow cost of hiring an additional unit of labor. Since that assumption is likely to be more controversial, it is important to keep in mind that it is not necessary in order to measure the gap as a whole, but it is only used in computing its decomposition between the two markups.

Under the assumption of a technology with constant elasticity of output with respect to hours (say, α), we have (up to an additive constant):

$$mpn_t = y_t - n_t \tag{8}$$

where y_t is output per capita and n_t is hours per capita.[6]

We assume that the (log) marginal rate of substitution for a representative consumer can be written (up to an additive constant) as:

$$mrs_t = \sigma \ c_t + \varphi \ n_t - \overline{\xi}_t \tag{9}$$

where c_t is consumption per capita and $\overline{\xi}_t$ is a low frequency preference shifter. Parameter σ is related to coefficient of relative risk aversion and φ measures the curvature of the disutility of labor.[7] Following Hall (1997), we allow for the possibility of low frequency shifts in preferences over consumption versus leisure, as represented by movements in $\overline{\xi}_t$. These preference shifts may be interpreted broadly to include institutional or demographic changes that affect the labor market, but which are unlikely to be of relevance at business cycle frequencies. We differ from Hall, though, by restricting these shifts to the low frequency. In section 4 we provide evidence to justify this assumption.

Under the above assumptions our gap variable is thus given by:

$$gap_t = (\sigma \ c_t + \varphi \ n_t - \overline{\xi}_t) - (y_t - n_t) \tag{10}$$

[6]Under certain assumptions that specification is compatible with variable labor utilization, particularly if labor effort moves roughly proportionately with hours per worker, and the latter is highly positively correlated with aggregate hours (per capita), as the evidence suggests. See, e.g., Basu and Kimball (1997) for a detailed discussion.

[7]The parameter measures the curvature of the utility function under the standard assumption that labor supply adjusts along the intensive margin (i.e., over hours). As we show in Appendix A, however, under certain assumptions our framework also allows for labor supply adjustment to occur instead over the extensive margin (i.e., over participation.) Finally, this log-linear representation of the mrs has been reconciled with balanced growth in a model with household production (see Baxter and Jermann (1999), or in a generalized indivisible labor model (see King and Rebelo (1999).)

Furthermore, we can combine the above assumptions with the definition of the price markup to obtain:

$$\mu_t^p = (y_t - n_t) - (w_t - p_t) \quad (11)$$

$$\equiv - ulc_t \quad (12)$$

Hence the price markup can be measured (up to an additive constant) as *minus* the (log) *real* unit labor costs, denoted by ulc_t.. Similarly, the wage markup is given by:

$$\mu_t^w = (w_t - p_t) - (\sigma\ c_t + \varphi\ n_t) + \overline{\xi}_t \quad (13)$$

3 The Gap and Its Components: Evidence

We now use the theoretical relations in the previous section to construct measures of the gap and its two main components: the price and wage markups. Our evidence is based on quarterly postwar U.S. data over the sample period 1960:I - 2002:IV, and are drawn from the USECON database.[8]

Identification of gap and wage markup variations requires that we make an assumption on the coefficient of relative risk aversion σ and on φ, a parameter which corresponds to the inverse of the (Frisch) wage elasticity of labor supply. A vast amount of evidence from micro-data suggests wage elasticities mostly concentrated in the range of $0.05 - 0.5$.[9] On the other hand, the business cycle literature tends to

[8] The data used to construct the gap variable and its components were drawn from the USECON database commercialized by Estima in Rats format. The time series used (with corresponding mnemonics shown in brackets) include compensation per hour (LXNFC), hours all persons (LXNFH), real and nominal output (LXNFO and LXNFI), all of which refer to the nonfarm business sector. We also make use of the NIPA series for non-durable and services consumption (CNH+GSH). In addition we also use population over sixteen (POP16) to express variables in per capita terms, real GDP (GDPQ), implicit GDP deflator (GDPD), the Fed-funds rate (FFED), the spread between the 10-year government bond yield (FCM10) and the 3-month Treasury Bill rate (FTB3), and a commodity price index (PSCOM) for our VAR exercise in Figure 4.

[9] In his survey of the literature, Card (1994) concludes that the intertemporal elasticity of labor supply is "surely no higher than 0.5 and probably no higher than 0.2". However, whether it is appropriate to use the existing micro evidence to calibrate the intertermporal elasticity of labor supply is a matter of considerable controversy, particularly to the extent that employment adjusts

use values of unity and higher, using balanced growth considerations as a justification, as opposed to direct evidence (see, e.g., Cooley and Prescott, 1995). We use as a baseline value $\varphi = 1$, which we view as a reasonable compromise between the values suggested in the micro and macro literature[10]. In addition, because it will turn out that the costs of fluctuations vary inversely with the Frisch labor supply elasticity, we are biasing our analysis against finding large welfare costs by choosing an elasticity that is above most of the direct estimates in the literature.

There is a similar controversy over the choice of the coefficient of relative risk aversion, which corresponds to the inverse of the intertemporal elasticity of substitution. Direct estimates of the latter tend to fall in the range $0.1 - 0.3$. This evidence suggests a value of σ that varies from 10 to 3.3.[11] Below we assume a value of 5 in our baseline calibration, but then also experiment with alternative values.

In addition, we need to make an assumption to identify the low frequency shifter $\overline{\xi}_t$. Let $\widetilde{\mu}_t^w \equiv (w_t - p_t) - (\sigma c_t + \varphi\ n_t)$ be the observable component of the wage markup. It follows that

$$\widetilde{\mu}_t^w = \mu_t^w - \overline{\xi}_t \tag{14}$$

From this perspective, the wage markup μ_t^w is the "cyclical" component of $\widetilde{\mu}_t^w$ and $\overline{\xi}_t$ is (minus) the "trend" component. Following Hall (1997) we approximate the low frequency movements of $\widetilde{\mu}_t^w$ by fitting a fifth-order polynomial of time to $\widetilde{\mu}_t^w$.[12]

along the extensive margin as well as the intensive margin (see, e.g., the discussion in Mulligan (1998)).

[10] It is true that our framework allows for labor market frictions, while most estimates of the intertemporal elasticity of labor supply elasticity are conditional on the assumption of frictionless labor markets. However, to the extent that a countercyclical wage markup contributes to the positive relation between wages and hours, the estimates of the intertemporal elasticity are likely to be biased upwards. In this respect, our parametrization may be conservative. See Blundell, Ham and Meghir (1998) for an attempt to factor in labor market frictions for case of female labor supply.

[11] Using micro-data, Barsky *et al.* (1997) estimate an intertemporal elasticity of substitution of 0.18, implying a coefficient of relative risk aversion slightly above 5. Using macro-data, Hall (1988) concludes that the intertemporal elasticity of substitution ($1/\sigma$) is likely below 0.2.

[12] Because we use the gap measure in subsequent time series analysis, we opt for a high order polynominal instead of a band pass filter to detrend the data. The fifth order polynominal detrended, however, produces cycles that closely resemble those that arise from a band pass filter that removes

Finally, before proceeding, we note that the relationships derived in the previous section hold only up to an additive constant. Accordingly, our framework only allows us to identify the *variations* over time in the markup and its components, but not their levels. Our baseline results thus employ measures of the price and wage markups and the gap constructed using, respectively, equations (12), (13), and (6), expressed in terms of deviations from their respective sample means.

Figure 2 presents the times series measure of our gap variable under our baseline assumptions of $\sigma = 5$ and $\varphi = 1$. Notice that this variable comoves strongly with the business cycle, displaying large declines during NBER-dated recessions (represented by the shaded areas in the graph). It is also interesting to observe that the increase in the gap during the rapid growth period post-1995 is much less significant than in earlier expansions. A natural interpretation is that the rapid output growth over this period must have been supported by underlying real factors (e.g. technology improvements), as opposed to excess demand.

We next decompose the movements of the gap into its wage and price markup components. The wage markup measures were constructed using (13).[13] The price markup corresponds to minus the log of real unit labor costs, as implied by (12). Figure 3 shows the behavior of the gap against the wage markup (both relative to their means). To facilitate visual inspection, we plot the inverse of the wage markup (i.e., minus the log wage markup). By definition, the difference between the gap and the inverse wage markup is the inverse price markup. What is striking about the pictures is the strong co-movement between the gap and the (inverse) wage markup. Put differently, our evidence suggests that the inefficiency gap seems to be driven largely by countercyclical movements in the wage markup.[14]

frequencies above 2 quarters and below either 32 or 64 quarters.

[13]The results are robust to simple adjustments for compositional bias of the real wage, based on Barsky, Solon and Parker (1994).

[14]As a somewhat cleaner way to illustrate the strong countercylical relation between the gap and the wage markup, we show later that this pattern also holds conditional on a shock to monetary policy.

To be clear, our conclusion that countercyclical wage markup variation drives the variation in the gap rests on the assumption that wages are allocational and can thus be used to construct a relevant cost measure.[15] While this assumption is standard in the literature on business cycles and markups (e.g., Rotemberg and Woodford, 1999), it is not without controversy. Notice, however, that even if observed wages are not allocational, our gap variable is still appropriately measured, since its construction does not require the use of wage data. Thus our welfare analysis, which depends on the overall gap and not its decomposition, is not affected by this issue.

Table 1 reports some basic statistics that support the visual evidence in Figure 3. In particular, the Table reports a set of second moments for the gap and its two components: the wage and price markup, and also for detrended (log) GDP, a common indicator of the business cycle. Note first that the percent standard deviation of the gap is large (relative to detrended output) and that departures of the gap from steady state are highly persistent. In addition, the wage markup is nearly as volatile as the overall gap, and is strongly negatively correlated with the latter, as well as with detrended GDP. This confirms the visual evidence that movements in the gap are strongly associated with countercyclical movements in the wage markup. On the other hand, the price markup is less volatile than the wage markup and does not exhibit a strong contemporaneous correlation with the gap.[16]

In results reported in an earlier version of the present paper (Gali, Gertler and Lopez-Salido (2002)) we demonstrate that our gap measure is robust to alternative assumptions about production that yield alternative measures of the marginal product of labor. Our baseline case assumes constant elasticity of output with respect

[15] Some indirect evidence that wages are allocational is found in Sbordone (1999) and Galí and Gertler (1999) who show that firms appear to adjust prices in response to measures of marginal cost based on wage data. In turn, as shown in Galí (2001), they do not respond to marginal cost measures that employ the household's marginal rate of substitution in place of the wage, as would be appropriate if wages were not allocational.

[16] However, the relatively weak co-movement of the price markup with detrended output is useful for understanding the dynamics of inflation and the recent evidence on the New Keynesian Phillips curve. See Sbordone (1999) and Gali and Gertler (1999).

to hours. We consider three alternative assumptions suggested by Rotemberg and Woodford (1999): (i) Cobb-Douglas modified to allow for overhead labor; (ii) CES; and (iii) Cobb-Douglas with labor adjustment costs. In each case we follow the parametrization recommended in Rotemberg and Woodford (1999). In all cases, our gap and markup measures are quite robust to these alternative assumptions. We also show that our results are robust to allowing for a measure of the marginal rate of substitution based on time dependent preferences in leisure, following Eichenbaum, Hansen and Singleton (1988).[17]

To summarize: the results thus far suggest that the business cycle is associated with large co-incident movements in the efficiency gap. Thus, under our framework, the evidence suggests that countercyclical markup behavior is an important feature of the business cycle. A decomposition of the gap, further, suggests that the countercyclical movement in the wage markup is by far the most important source of overall variations in the gap. Thus, to the extent wages are allocational, some form of wage rigidity, either real or nominal, may be central to business fluctuations.

4 Labor Supply Shifts, Hall's Residual and the Gap

We have proceeded under the interpretation that our measured gap between the marginal rate of substitution and the marginal product of labor reflects countercyclical markup behavior. In his baseline identification scheme, however, Hall modeled this gap as an unobserved preference shock, though he was clear to state that he did not take this hypothesis literally. Subsequent literature, however, (e.g., Holland and Scott (1998), Francis and Ramey (2001) and Uhlig (2002)) has indeed interpreted this residual as reflecting either exogenous labor supply shifts or some other unspecified exogenous driving force. In this section we show that the high frequency movements

[17] In Gali, Gertler and Lopez-Salido (2002) we show that it also remains true that the movements in the gap for alternative values of labor supply elasticity and the risk aversion are associated largely with countercyclical movements in the wage markup.

in the gap cannot be simply due to exogenous preference shifts, etc. Rather, the evidence is instead compatible with our countercyclical markup interpretation.

Let us follow Hall (1997) by assuming that the marginal rate of substitution is now augmented with a preference shock ξ_t that contains a cyclical component, $\widetilde{\xi_t}$, as well as a trend component, $\overline{\xi_t}$:

$$mrs_t = c_t + \varphi\, n_t - \xi_t \tag{15}$$

with

$$\xi_t = \overline{\xi_t} + \widetilde{\xi_t}$$

where we maintain our baseline assumption that the coefficient of relative risk aversion, σ, is unity. Hall then defines the residual x_t as the difference between the "observable" component of the marginal rate of substitution, $c_t + \varphi\, n_t$, and the marginal product of labor, $y_t - n_t$:

$$x_t \equiv (c_t + \varphi\, n_t) - (y_t - n_t) \tag{16}$$

The issue then is how exactly to interpret the movement in Hall's residual. Using the augmented specification of the marginal rate of substitution allowing for preference shocks (15), together with (8) and the definition of the inefficiency gap (1), it is possible to express Hall's residual as follows:

$$\begin{aligned} x_t &\equiv (mrs_t - mpn_t) + \xi_t && (17)\\ &= -(\mu_t^p + \mu_t^w) + \xi_t && (18) \end{aligned}$$

Hall's assumption of perfect competition in both goods and labor markets implies $\mu_t^p = \mu_t^w = 0$. This allows him to interpret variable x_t as a preference shock, since under this assumption $x_t = \xi_t$.[18] Notice that under these circumstances the efficiency gap is zero, as there are no imperfections in either goods or labor markets. On the other hand, if preferences are not subject to shocks ($\xi_t = 0$, all t), and we allow for departures from perfect competition, x_t will purely reflect movements in markups, i.e., $x_t = -(\mu_t^p + \mu_t^w)$. In the latter instance, Hall's residual corresponds exactly to our inefficiency gap, i.e., $x_t = gap_t$, for all t.

Note that if the Hall residual indeed reflects exogenous preference shocks, it should be invariant to any other type of disturbance. In other words, the null hypothesis of preference shocks implies that the Hall residual should be exogenous. We next present two tests that reject the null of exogeneity, thus rejecting the preference shock hypothesis.

First, we test the hypothesis of no-Granger causality from a number of variables to our gap measure. The variables used are: detrended GDP, the nominal interest rate, and the yield spread. Both the nominal interest rate and the yield spread may be thought of as a rough measure of the stance of monetary policy, while detrended GDP is just a simple cyclical indicator. Table 2 displays the *p-values* for several Granger-causality tests. These statistics correspond to bivariate tests using alternative lag lengths. They indicate that the null of no Granger-causality is rejected for all specifications, at conventional significance levels. That finding is robust to reasonable alternative calibrations of σ and φ. Overall, the evidence of Granger causality is inconsistent with the hypothesis that the Hall residual mainly reflects variations in preferences.

As a second test, we estimate the dynamic response of our gap variable to an identified exogenous monetary policy shock. The identification scheme is similar to the one proposed by Christiano et al. (1999), and others. It is based on a VAR that includes measures of output, the price level, commodity prices, and the Federal Funds rate, to which we add our gap measure (or, equivalently, Hall's residual) and the price

[18] See also Baxter and King (1991). Holland and Scott (1998) construct similar measures for the U.K.

markup. From the gap and the price markup response we can back out the behavior of the wage markup, using equation (6). We identify the monetary policy shock as the orthogonalized innovation to the Federal Funds rate, under the assumption that this shock does not have a contemporaneous effect on the other variables in the system. Figure 4 shows the estimated responses to a contractionary monetary policy shock. The responses of the nominal rate, output, consumption and prices are similar to those found in Christiano *et al.* (1999), Bernanke and Mihov (1998), and other papers in the literature. Most interestingly for our purposes, the inefficiency gap declines significantly in response to the unanticipated monetary tightening. Its overall pattern of response closely mimics the response of output. This endogenous reaction, of course, is inconsistent with the preference shock hypothesis, but fully consistent with our hypothesis that countercyclical markups may underlie the cyclical variation in the Hall residual. In this respect, note that the tight money shock induces a rise in the wage markup that closely mirrors the decline in the gap, both in the shape and the magnitude of the response. This countercyclical movement in the wage markup is consistent with evidence on unconditional variation presented in Table 1. The price markup also rises, though with a significant lag. Apparently, the sluggish response of wages, which gives rise to strong countercyclical movement in the wage markup, delays the rise in the price markup.[19] In any event, the decline in the inefficiency gap is clearly associated with a countercyclical rise in markups.

To be clear, because preference shocks are not observable, it is not possible to directly determine the overall importance of these disturbances. While our evidence rejects the hypothesis that exogenous preference variation drives all the movement in our gap measure, it cannot rule out the possibility that some of this movement

[19] As Gali and Gertler (1999) and Sbordone (1999) observe, the sluggish behavior of the price markup helps explain the inertial behavior of inflation, manifested in this case by the delayed and weak response of inflation to the monetary shock. Staggered pricing models relate inflation to an expected discounted stream of real marginal costs, which corresponds to the inverse of the price markup. The sluggish response to the price markup translates into sluggish behavior of real marginal cost.

is due to preference shocks. Yet, to the extent that preference shocks are mainly a low frequency phenomenon then they are likely to be captured by the trend component associated with our low frequency filter (together with other institutional and demographic factors which may lead to low frequency variations in markups). In this instance our filtered gap series, which isolates the high frequency movement in this variables, is likely to be largely uncontaminated by exogenous preference variations.

5 Relation to the Output Gap

As a necessary prelude to the welfare analysis that follows, in this section we illustrate the connection between our gap measure and the output gap, a more traditional indicator of cyclical utilization. The output gap is commonly meant to refer to the deviation of output from its *natural* level, defined as the equilibrium value in the absence of nominal rigidities and their associated cyclical variations in price and wage markups. Formally,

$$\widetilde{y}_t \equiv y_t - \overline{y}_t \tag{19}$$

where $\widetilde{y}_t$, and $\overline{y}_t$ denote the output gap and the natural level of output, respectively. While it is not possible to derive an exact relation between the output gap and the inefficiency gap without specifying a complete model, we demonstrate how it is possible to derive a relatively tight band simply conditional on a minimal set of plausible assumptions.

First, we need a restriction on technology. For simplicity, we assume that the reduced form aggregate production function can be written as:

$$y_t = a \ n_t + z_t \tag{20}$$

where z_t is exogenous or, at least, invariant to the degree of nominal rigidities.[20] Think of z_t as including both technology and capital, where we treat capital as ex-

[20] Simulations using a model with capital accumulation suggests that such an assumption is a good approximation.

ogenous on the grounds that the percent fluctuations in capital at the business cycle frequency a relatively small. Note that equation (20) allows the possibility of variable capital utilization, following Burnside and Eichenbaum (1996) and King and Rebelo (1999). Here we simply observe that variable capital utilization will raise the *effective* output elasticity of employment, a. Given equations (10) and (20), we can derive the following expression for the gap:

$$gap_t = \left(\frac{1-a+\varphi}{a}\right)\ y_t + \sigma\ c_t - \left(\frac{1+\varphi}{a}\right)\ z_t - \overline{\xi}_t \tag{21}$$

Second, we assume that the only source of gap variation lies in the presence of nominal rigidities in labor and/or goods markets. In other words, frictionless or desired markups are assumed to be constant. This assumption permits us to interpret the natural level of output as the level of output consistent with a *constant* gap (which corresponds to its steady state value, gap). Accordingly, if we let $\overline{c}_t$ be the level of consumption in the absence of nominal rigidities, then it follows that $\overline{y}_t$ and $\overline{c}_t$ must satisfy

$$gap = \left(\frac{1-a+\varphi}{a}\right)\ \overline{y}_t + \sigma\ \overline{c}_t - \left(\frac{1+\varphi}{a}\right)\ z_t - \overline{\xi}_t \tag{22}$$

To obtain a relation between the output gap our demeaned gap measure $\widehat{gap}_t \equiv gap_t - gap$ (e.g., as portrayed in figure 2), first combine equations (21) and (22):

$$\widehat{gap}_t = \left(\frac{1-a+\varphi}{a}\right)\ \widetilde{y}_t + \sigma\ \widetilde{c}_t \tag{23}$$

where $\widetilde{c}_t = c_t - \overline{c}_t$. Without loss of generality, we can express the consumption gap, $\widetilde{c}_t$, as a time varying proportion of the output gap, as follows

$$\widetilde{c}_t = \eta_t\ \widetilde{y}_t \tag{24}$$

Conditional on η_t, we can express $(y_t - \overline{y}_t)$ as a function of our gap measure. To see this, notice that by substituting equation (24) into (23), we obtain,

$$\widetilde{y}_t = \left(\frac{a}{1 + \varphi + a(\sigma\eta_t - 1)} \right) \widehat{gap}_t \tag{25}$$

To illustrate, we compute the output gap for our baseline preference specification, with $\sigma = 5$ and $\varphi = 1$. We follow King and Rebelo (1999), who argue that the evidence is consistent with a value of a of roughly unity. Unfortunately, we do not have any direct estimates of η_t. To calibrate its value we can make use of the evidence portrayed in our Figure 4, in which, by construction the natural levels of both consumption and output remain constant. As can be seen in this Figure, this elasticity remains quite stable around 0.5, given the parallel responses of consumption and output.[21] In addition, to that evidence Deaton and Campbell (1989) report the standard deviation of consumption is roughly 0.8 the standard deviation of output, with a correlation of roughly 0.9. To the extent that much of the high frequency variation is output reflects variation in the output gap, a value of η_t around 0.7 seems reasonable. We accordingly use this value as our benchmark number but also experiment with values in the range between 0.5 and 1.0.

Figure 5 plots our output gap with a commonly used measure of the output gap, constructed by the Congressional Budget Office (CBO). Broadly speaking, our measure of the output gap has similar properties to the CBO gap. The correlation between the two series 0.6 for the whole sample and 0.76 for the period since 1970. Though we do not report the results here, our output gap measure is largely robust to letting η_t vary between 0.5 and 1.0.

6 Welfare and the Gap

We now use the results of the previous section to derive a simple way to measure the welfare costs of fluctuations in the degree of inefficiency of aggregate resource allocations, as captured by our gap variable. We then apply this methodology to

[21] Simulations of a sticky price with capital accumulation confirm the proportionality of the consumption/output responses to different kind of shocks.

postwar U.S. data. In addition to obtaining a measure of the average cost of gap fluctuations, we also compute the welfare losses during particular episodes, including the major postwar recessions.

As we noted in the introduction, our approach differs from Lucas (1987) and others by focusing on the costs stemming from fluctuations in the degree of inefficiency of the aggregate resource allocation, as reflected by the movements in our gap variable.[22] As in Ball and Romer (1987), the cycle generates losses on average within our framework because the welfare effects of employment fluctuations about the steady state are asymmetric. As Figure 1 illustrates, given that the steady state level of employment is inefficient (due to positive steady state price and wage markups), the efficiency costs of an employment contraction below the steady state will exceed the benefits of a symmetric increase. In particular, note that the vertical distance between the labor demand and supply curves rises as employment falls below the steady state and falls when employment moves above. The quantitative effect of this nonlinearity on the welfare cost of fluctuations ultimately depends on the slopes of the labor demand and supply curves, and on the steady state distance relative to the first-best, perfectly competitive steady state. Below we show that under some plausible parameter values the resulting average welfare cost can be substantial.

6.1 A Welfare Measure

We now proceed to derive our welfare measure. We continue to assume a production technology with constant elasticity of output with respect to hours, as given by equation (20). We also assume, following King and Rebelo, that the output elasticity with respect to hours is unity ($a = 1$). Accordingly, equation (20) can be written (in levels) as

[22] For other approaches to measuring the unconditional costs of fluctuations see, e.g., Barlevy (2000) and Beaudry and Pages (2001). For a very early attempt to measure the welfare cost of inefficiently high unemployment, see Gordon (1973).

$$Y_t = Z_t \; N_t \tag{26}$$

Next, we obtain a measure of the utility gain or loss, Δ_t, from reallocating employment at time t from its existing level N_t to the level that would arise in the frictionless equilibrium, $\overline{N}_t$. Let $W(N_t)$ be the period t utility value of output, net of the utility cost of working, conditional on employment level N_t. Accordingly,

$$\Delta_t = W(N_t) - W(\overline{N}_t) \tag{27}$$

Assuming that, at the margin, the household is indifferent between consuming and saving any additional unit of output, the shadow value of the latter will equal the marginal utility of consumption. Accordingly, if $U(C_t, N_t)$ is the period utility function of the representative consumer, then the gross utility gain from raising employment is given by the marginal utility of consumption times the marginal product of labor, i.e. $U_{c,t} \; Z_t$. The cost is $U_{n,t} < 0$, the marginal disutility of hours. On net the welfare effect is thus given by $W'(N_t) = U_{c,t} \; Z_t + U_{n,t}$.

Ultimately, however, we are interested in an expression for the total utility gap, Δ_t. If the percent difference between N_t and $\overline{N}_t$ is not large, then it is reasonable to approximate $W(N_t)$ with a second order Taylor expansion about $W(\overline{N}_t)$. Accordingly, assuming that utility is separable in consumption and leisure (i.e., $U_{cn} = 0$) we have:

$$\begin{aligned} \Delta_t \;\; \simeq \;\; & [\overline{U}_{c,t} Z_t + \overline{U}_{n,t}] \; (N_t - \overline{N}_t) \\ & + \frac{1}{2} \left(\overline{U}_{cc,t} \; \frac{\partial \overline{C}_t}{\partial \overline{Y}_t} \; Z_t + \overline{U}_{nn,t} \right) \; (N_t - \overline{N}_t)^2 \end{aligned} \tag{28}$$

where, in order to lighten the notation we have defined $U_t \equiv U(C_t, N_t)$, $\overline{U}_t \equiv U(\overline{C}_t, \overline{N}_t)$, and $U \equiv U(C, N)$.

Assume that preferences are given by $U(C,N) = \frac{C^{1-\sigma}}{1-\sigma} - \frac{N^{1+\varphi}}{1+\varphi}$. Then, as we show in the Appendix B, it is possible to express Δ_t as the following quadratic function of the percent deviation of employment from its value in the frictionless equilibrium, $\widetilde{n}_t \equiv \log\left(\frac{N_t}{\overline{N}_t}\right)$:

$$\Delta_t \simeq \frac{\overline{U}_{c,t}\overline{Y}_t}{(1+\mu)} \left\{\mu\, \widetilde{n}_t \ - \frac{1}{2}\left[\varphi + \sigma(1+\mu)\eta_t - \mu\right]\ \widetilde{n}_t^2\right\} \tag{29}$$

where, as before, η_t is the elasticity of the consumption gap with respect to the output gap, and where μ is the steady state net markup.

Since we do not have a direct measure of the frictionless level of employment, $\widetilde{n}_t$ is not observable. We can, however, derive a relation between $\widetilde{n}_t$ and our gap variable, $\widehat{gap}_t$, which is measurable. Given the production function (26), it follows that $\widetilde{n}_t = \widetilde{y}_t$. We can then obtain a relation for $\widetilde{n}$ in terms of $\widehat{gap}_t$, by exploiting the analysis of the previous section, in particular equation (25).

Let $\widetilde{\sigma} = \eta_t\ \sigma$, given this restriction and our earlier assumption that $a = 1$, from expression (25) it follows that

$$\widetilde{n}_t = \left(\frac{1}{\widetilde{\sigma} + \varphi}\right)\ \widehat{gap}_t \tag{30}$$

Combining equations (29) and (30) then yields an expression for Δ_t in terms of $\widehat{gap}_t$:

$$\Delta_t \simeq \overline{U}_{c,t}\overline{Y}_t\ \omega(\widehat{gap}_t) \tag{31}$$

where

$$\omega(\widehat{gap}_t) = \frac{1}{(1+\mu)(\widetilde{\sigma}+\varphi)}\ \left[\mu\ \widehat{gap}_t - \frac{1}{2}\left(1 + \frac{\mu(\widetilde{\sigma}-1)}{\widetilde{\sigma}+\varphi}\right)\ \widehat{gap}_t^2\right] \tag{32}$$

Observe that $\omega(\widehat{gap}_t)$ is the efficiency loss or gain from gap deviations from its steady value, expressed as a percent of the frictionless level of output $\overline{Y}_t$ (since $\omega(\widehat{gap}_t) = \Delta_t/(\overline{U}_{c,t}\overline{Y}_t)$). The first term in brackets, the linear term, reflects the symmetric costs and benefits from the gap moving below and above the steady state, due to the positive steady state markup μ. The second term, the quadratic term, captures the asymmetric effect of gap fluctuations on welfare. In particular, a reduction in the gap below the steady state results in an efficiency loss that exceeds the gain stemming from a commensurate increase in the gap above the steady state.

Finally, in keeping with the literature, we express the welfare losses from gap fluctuations measure terms of an equivalent loss in consumption. In particular, we divide Δ_t by $\overline{U}_{c,t}\overline{C}_t$ to make the metric the percent of consumption in the frictionless equilibrium:. Let $\Phi = Y/C$ be the steady state output/consumption ratio. Then the efficiency loss (or gain) as a percent of $\overline{C}_t$ is given by

$$\begin{aligned} \frac{\Delta_t}{\overline{U}_{c,t}\overline{C}_t} &\approx \frac{\overline{Y}_t}{\overline{C}_t}\, \omega(\widehat{gap}_t) \\ &\approx \Phi\, \omega(\widehat{gap}_t) \end{aligned} \tag{33}$$

We can use equation (33) to calculate a time series of the efficiency gain or loss in each quarter t. To obtain a measure of the average welfare cost over time, we take the unconditional expectation of equation (33) to obtain:

$$E\left\{\frac{\Delta_t}{\overline{U}_{c,t}\overline{C}_t}\right\} = -\frac{\Phi}{(1+\mu)(\widetilde{\sigma}+\varphi)}\left[\frac{1}{2}\left((1+\mu)+\frac{\mu(\widetilde{\sigma}-1)}{\widetilde{\sigma}+\varphi}\right) var(gap_t)\right] \tag{34}$$

where $var(gap_t)$ is the variance of the inefficiency gap. Notice that, as a result of the concavity of ω, the expected welfare effects of fluctuations in the gap variable are negative, i.e. these fluctuations imply *losses* in expected welfare. This loss, further, is of "second order," as it is linearly related to the variance of the inefficiency gap. It is, however, potentially large, depending in particular on the magnitude $var(gap_t)$. As section 3 suggests, $var(gap_t)$, is potentially large if labor supply is relatively inelastic or risk aversion is relatively high.

To be clear, our approach provides a lower bound on the measure of the total welfare costs of fluctuations. The reason is simple: it does not include the welfare costs from *efficient* fluctuations in consumption and employment. Suppose, for example, that the data were generated by a real business cycle model with frictionless, perfectly competitive markets. We should then expect to see no variation in our gap measure, as the resource allocation would always be efficient. Our metric would then indicate no welfare costs of fluctuations, while some losses would still be implied by the variability of consumption and leisure (under standard convexity assumptions on preferences). It is also important to stress that, to the extent that the steady state value of the gap corresponds also to the average value around which the economy fluctuates, as assumed above, *average* welfare losses will only be of second order. On the other hand, our efficiency cost measure suggests possible first order effects at any moment in time: As reflected in equation (32) and illustrated further below, deviations in the gap variable from that steady state may have non-negligible first-order welfare effects, with gap declines generate greater losses than symmetric gap increases due to the concavity of the welfare loss function.

6.2 Some Numbers

Equation (33) provides a real time measure of the efficiency costs of deviations of our gap variable from steady state. Accordingly, we construct a quarterly time series of $\Delta_t/\overline{U}_{c,t}\overline{C}_t$, taking as input our baseline measure of the gap, with parameter values $\sigma = 5$ and $\varphi = 1$. In addition, to construct the welfare measure, we continue to assume that the elasticity of the consumption gap with respect to η, equals 0.7. Figure 6 plots the resulting time series over the sample 1960:IV-2002:IV. The value at each period t can be interpreted as the efficiency gain or loss in percentage units of consumption associated with the deviation of the inefficiency gap from its steady state. The asymmetric nature of the gains and losses is clear. As the figure shows, significant efficiency losses arise in recessions that do not appear to be offset by commensurate gains during booms. Note also that the efficiency losses are largest during the major recessions, ranging between 3.0 and 3.5 percent of consumption per period around the time of the respective troughs. During the major recessions, further, these losses persist for for several years, suggesting nontrivial costs around these periods. We return subsequently to this issue.

We next present a measure of the average welfare cost of the cycle, based on equation (34). As we noted earlier, the measure is simply proportionate to the square of the gap. We construct estimates for alternative values of the parameters φ and σ. For the inverse of the Frisch labor supply, φ, we consider in addition to our benchmark value of unity, two values that are more in line with the micro evidence: 2 (implying a Frisch elasticity of 0.5) and 5 (implying a Frisch elasticity of 0.2). For the coefficient of relative risk aversion, in addition to our benchmark value of 5, we consider 1.0 and 10, which spans the set of values typically studied in the literature. For the parameter μ, the sum of the steady state wage and price markups, we assume a value of 0.40. A value of 0.20 plausible for the steady state price markup (see Rotemberg and Woodford (1999)). Since the steady state wage markup depends on tax distortions as well as workers' market power, 0.20 seems a reasonable lower bound given the evidence on average labor tax rates.

For a parametrization that corresponds to our baseline case of section 3 ($\varphi = 1, \sigma = 5$), we estimate the average efficiency costs of postwar U.S. business fluctuations to be quite modest, roughly 0.28 percent of consumption. The estimates of welfare costs are nonetheless sensitive to the labor supply elasticity and the intertemporal elasticity of substitution. For high values of the two elasticities (corresponding in our case to $\sigma = 1$ and $\varphi = 1$), the welfare costs are small, on the order of Lucas' benchmark estimate of 0.07. In this case, roughly speaking, the labor supply curve is relatively flat, implying small cyclical fluctuations in the inefficiency gap. On the other hand, in the case of low elasticities, the costs can become fairly significant. With $\varphi = 5$ (corresponding to a labor supply elasticity of 0.2), the costs vary from of 0.3 percent of consumption for the case of $\sigma = 1$, to roughly 0.75 for the case of $\sigma = 10$. In this latter case, the labor supply curve is very steep, implying very large fluctuations in the inefficiency gap. It is worth noting that, while higher than the values normally used in business cycle calibration exercises, $\varphi = 5$ and $\sigma = 10$ fall within the range of estimates in the literature, as we discussed in section 3.

Any measure of the average cost of business cycles, however, obscures the fact that individual recessionary episodes may be rather costly. What moderates the impact of these episodes on the overall welfare measure is the fact they have been relatively infrequent, particularly over the last several decades. One reason for this may be that stabilization policy has been reasonably effective. Another possibility is that the economy has been subject to smaller shocks. In either event, it is of interest to examine efficiency losses during the major boom-bust episodes. Doing so provides a sense of the gains from avoiding future recessions (either by good policy or by good luck.)

We accordingly consider the two major episodes in our sample where the economy experienced full cycle (i.e. a boom followed by a deep recession) in terms of our gap variable. They correspond to the periods 1972:II-1978:I and 1978:II-1984:IV In each instance we measure the boom as the period where the gap variable climbs above zero up to the point where it returns to zero. The recession is the period that follows, where the gap turns negative up to the time it returns to steady state. We measure the gains and losses as a percent of one year's consumption. For each episode, Table 4 reports the efficiency gains from the boom and the corresponding costs from the recession, followed by the net loss (the sum of the first two columns.)

We consider the same range of parameter values as in Table 3. In every case, the costs of the recession outweigh the benefits of the boom, as we might expect, given the asymmetric nature of the cyclical efficiency gains and losses. For our baseline case ($\sigma = 5$ and $\varphi = 1$), the net cost of the 80-82 recession was quite significant, nearly seven percent of one year's consumption, while the gross cost was just over eight percent. The 74-75 recession was also costly, though not of the same magnitude, with the net costs being 2.6 percent of one year's consumption and the gross costs

4.0 percent. The 74-75 recession was not as protracted and also did not feature as sharp a decline in aggregate consumption.

The results, of course are sensitive to the degree of risk aversion, as well as the elasticity of labor supply. Risk aversion matters both because it affects the slope of the labor supply curve and also because it affects how households value the efficiency losses in recessions relative to the gains in booms. With $\sigma = 1$, the net costs of the 80-82 recession decline to to 1.5 percent, and the gross costs fall to 3.3 percent. On the other hand, with $\sigma = 10$, the net costs rise to nearly eleven and a half percent and the gross costs jump to over twelve percent. Given that a risk aversion of ten is not unreasonable, the results suggests potentially huge efficiency of the 80-82 downturn. Because the consumption decline was relatively modest during the 74-75, the associated efficiency costs are not as sensitive to the degree of risk aversion.

Labor supply elasticity is also relevant since the efficiency losses in downturns depend inversely on the slope of the labor supply curve. For example, given our baseline risk aversion of five, the net costs of the 74-75 recession increase from 2.6 percent of steady state consumption to 6.0 percent, as φ varies from unity to five. The gross costs, in turn, vary from 4.0 percent to over 7.0 percent. The costs of the $80 - 82$ recession increase only modestly as the labor supply elasticity declines, even though these costs are quite large in our high elasticity case, as we noted earlier. Overall, our results suggest only modest *average* efficiency losses from fluctuations. However, major boom-bust episodes appear entail rather significant losses.

7 Concluding Comments

At the risk of considerable oversimplification, it is possible to classify modern business cycle models into two types. The first class attempts to explain quantity fluctuations by appealing to high degrees of intertemporal substitution in an environment of frictionless markets. The second instead appeals to countercyclical markups owing to particular market frictions. In this regard, there has been a considerable debate as to whether the markup is indeed countercyclical (see Rotemberg and Woodford (1999) for a summary). Much of this debate has been centered around markup measures that use wage data to calculate the cost of labor. We show, however, that the markup is highly countercyclical, using the household's consumption/leisure trade-off as the shadow cost of labor, as theory would suggest. Under this identification scheme, the markup corresponds exactly to the labor market residual studied by Hall (1997) and others. Whether the countercyclical markup variation is driven primarily by product market or labor market behavior is, however, an open question. To the extent that wages are allocative, we find that labor market frictions are the key factor. As we discussed, however, the exact form that these frictions may take (e.g., nominal wage rigidity, efficiency wages, search frictions, etc.) is also an open question.

A second message of this paper is that to the extent that our markup interpretation of the efficiency gap is correct, business cycles may involve significant efficiency costs. To be sure, our results suggest that these efficiency losses are modest when averaged over time. This result occurs, however, because -whether by good luck or good policy- significant recessions have not often occurred in the post war. We find, however that when they do occur, the efficiency costs may indeed be quite large, even after netting out the gains from the preceding boom. These results obtain for reasonably standard assumptions on preferences (e.g., a coefficient of reltive risk aversion of five and a unit-elastic Frisch labor supply). Thus, while the gains from eliminating all fluctuations may not be large–as suggested by the existing literature–there nonetheless do appear to be significant efficiency benefits from avoiding severe recessions.

Finally, we observe that our calculation ignores at least several important considerations that might be leading us to understate the efficiency costs of recessions. First, within our framework, a reduction in hours leads to increased enjoyment of leisure, which partially offsets the impact of the output decline. In reality, workers who are laid off during recessions do not simply get to enjoy the time off, but rather have to look for a new job. In addition, there is often a loss of human capital that was specific to the previous employer. Second, our calculation ignores the costs of inflation associated with the economy moving above steady state output (see, e.g. Woodford (1999)). For this reason, our metric may overstate the gains from booms. To the extent that the costs of high inflation roughly offset the efficiency gains from the boom, our measure of the gross efficiency loss of the recession may provide a more accurate indicator of the costs of these episodes. Taking into account these considerations is on the agenda for future research.

Table 1. Basic Statistics: 1960-2002

Baseline Calibration ($\sigma = 5$, $\varphi = 1$)

Variable	s.d.(%)	ρ	Cross Correlation			
			GDP	Gap	Price Mkup	Wage Mkup
GDP	2.9	0.91	1			
Gap	9.2	0.95	0.83	1		
Price Markup	4.6	0.96	0.09	0.01	1	
Wage Markup	10.3	0.96	-0.79	-0.90	-0.45	1

Table 2. Granger Causality Tests (1960-2002)

Baseline Calibration ($\sigma = 5$, $\varphi = 1$)

Bivariate VAR

Variable	4-lags	5-lags	6-lags
GDP	0.000	0.000	0.000
Nominal Interest Rate	0.018	0.018	0.036
Yield Spread	0.026	0.001	0.002

Note: The values reported are p-values for the null hypothesis of no Granger causality from each variable listed to Hall x (F-test). Filtered data using fith order polynomial in time.

Table 3. Welfare Costs of Fluctuations (1960-2002)

(percent of average consumption)

	$\varphi = 1$	$\varphi = 2$	$\varphi = 5$
$\sigma = 1$	0.071	0.125	0.296
$\sigma = 5$	0.284	0.332	0.485
$\sigma = 10$	0.582	0.624	0.764

Note: Calibration $a = 1$, $\mu = 0.4$. The average output/consumption ratio is 1.70. The data was filtered for the period 50:1-02:4 using fifth order polynomial in time (e.g. Hall (1987).) Welfare computations cover the sample period 60:1-02:4.

Table 4 : Costs(−) and Benefits(+) of Boom/Recession Episodes

(percent of one year's consumption)

		72:2-78:1			78:2-84:4		
		Boom	*Reces.*	*Net*	*Boom*	*Reces.*	*Net*
$\sigma = 1$							
	$\varphi = 1$	0.99	-5.38	-4.39	1.78	-3.29	-1.51
	$\varphi = 2$	1.18	-6.12	-4.94	2.00	-3.85	-1.85
	$\varphi = 5$	1.24	-7.98	-6.74	1.93	-5.23	-3.30
$\sigma = 5$							
	$\varphi = 1$	1.40	-4.13	-2.73	1.24	-8.10	-6.86
	$\varphi = 2$	1.32	-4.88	-3.56	1.34	-8.31	-6.97
	$\varphi = 5$	1.19	-7.21	-6.02	1.43	-8.34	-6.91
$\sigma = 10$							
	$\varphi = 1$	1.25	-4.07	-2.82	0.95	-12.38	-11.43
	$\varphi = 2$	1.19	-4.83	-3.64	1.01	-12.45	-11.44
	$\varphi = 5$	1.04	-7.09	-6.05	1.09	-12.69	-11.60

Note: See Table 3.

Appendix A: The Household's MRS

Here we illustrate that the expression we use for the household's marginal rate of substitution between consumption and leisure, equation (9), may be motivated either by making the standard assumption that labor supply adjusts along the intensive margin or, under certain assumptions, that the adjustment is along the extensive margin. Our argument is based on Mulligan (1998).

Case I: Labor Supply Adjustment Along the Intensive Margin
Let C_t and N_t denote consumption and hours worked, respectively. Assume a representative agent with preferences given by

$$\frac{1}{1-\sigma}\ C_t^{1-\sigma} - \frac{1}{1+\varphi}\ N_t^{1+\varphi}$$

It follows that

$$MRS_t = C_t^{\sigma}\ N_t^{\varphi}$$

By taking the log of each side of this relation we obtain equation (9).

Case II: Labor Supply Adjustment Along the Extensive Margin
Now assume that individuals either do not work or work a fixed amount of hours per week. Suppose there is a representative household with a continuum of members represented by the unit interval, and who differ according to their disutility of work. Specifically, let us assume that j^{φ} is the disutility of work for member j. Under perfect consumption insurance within the household, and interpreting N_t as the fraction of working household members in period t, total household utility will be given by

$$\frac{1}{1-\sigma}\ C_t^{1-\sigma} - \int_0^{N_t} j^{\varphi}\ dj$$

Note that

$$\int_0^{N_t} j^{\varphi} dj = \frac{1}{1+\varphi} N_t^{1+\varphi}$$

Accordingly, the utility function for the family in this case is isomorphic to case of adjustment along the intensive margin. It follows that the marginal rate of substitution has the same form as well.

Appendix B: The Welfare Metric

As shown in the text, the effect on welfare of a deviation from the frictionless level of employment can be approximated by

$$\Delta_t \simeq \left(\overline{U}_{c,t} Z_t + \overline{U}_{n,t}\right)\ (N_t - \overline{N}_t) + \frac{1}{2}\ \left(\overline{U}_{cc,t} \frac{\partial \overline{C}_t}{\partial \overline{Y}_t} Z_t + \overline{U}_{nn,t}\right)\ (N_t - \overline{N}_t)^2 + o(\|a\|^3)$$

where $o(\|a\|^n)$ represents terms that are of order higher than n^{th}, in the bound $\|a\|$ on the amplitude of the relevant shocks. A straightforward manipulation yields

$$\begin{aligned} \Delta_t \ \simeq\ & \overline{U}_{c,t} \overline{Y}_t \ \left(1 + \frac{\overline{U}_{n,t}}{\overline{U}_{c,t} Z_t}\right) \left(\frac{N_t - \overline{N}_t}{\overline{N}_t}\right) + \\ & \frac{1}{2} \left(\frac{\overline{U}_{cc,t} \overline{C}_t}{\overline{U}_{c,t}} \eta_t + \frac{\overline{U}_{n,t}}{\overline{U}_{c,t} Z_t} \frac{\overline{U}_{nn,t} \overline{N}_t}{\overline{U}_{n,t}}\right) \left(\frac{N_t - \overline{N}_t}{\overline{N}_t}\right)^2 \end{aligned}$$

where $\eta_t = \frac{\partial \overline{C}_t}{\partial \overline{Y}_t} \frac{\overline{Y}_t}{\overline{C}_t}$ is the elasticity of consumption with respect to output in the frictionless equilibrium. We now make use of the second order approximation of relative deviations in terms of log deviations:

$$\frac{X - \overline{X}}{\overline{X}} = \widetilde{x} + \frac{1}{2}\ \widetilde{x}^2 + o(\|a\|^3)$$

where $\widetilde{x} \equiv \log\left(\frac{X}{\overline{X}}\right)$. Hence, the previous expression for Δ_t can be written as

$$\Delta_t \simeq \overline{U}_{c,t} \overline{Y}_t \left\{ \left(1 - \frac{1}{1+\mu}\right) \left(\widetilde{n}_t + \frac{1}{2}\ \widetilde{n}_t^2\right) - \frac{1}{2}\ \left(\sigma \eta_t + \frac{\varphi}{1+\mu}\right)\ \widetilde{n}_t^2 \right\},$$

where we have used that $MPN_t = \frac{Y_t}{N_t} = Z_t$, $\frac{\overline{U}_{n,t}}{\overline{U}_{c,t}} / \overline{MPN}_t = GAP = \frac{1}{1+\mu}$, $\frac{\overline{U}_{cc,t} \overline{C}_t}{\overline{U}_{c,t}} = \sigma$, $\frac{\overline{U}_{nn,t} \overline{N}_t}{\overline{U}_{n,t}} = \varphi$. Finally, from the previous expression it is straightforward to obtain the expression for Δ_t that is in the main text:

$$\Delta_t \approx \frac{\overline{U}_{c,t} \overline{Y}_t}{(1+\mu)}\ \left\{ \mu\ \widetilde{n}_t\ \ - \frac{1}{2} \left[\varphi + \sigma \eta_t (1+\mu) - \mu\right]\ \widetilde{n}_t^2 \right\}$$

References

[1] Alexopolous, Michelle, (2001), "Unemployment and the Business Cycle," mimeo, University of Toronto.

[2] Ball, Lawrence and David Romer, (1989), "Are Prices Too Sticky?" *Quarterly Journal of Economics*, 1987, 104, 507-524.

[3] Barlevy, Gadi, 2000, "Evaluating Costs of Business Cycles of Business Cycles in Models of Endogenous Growth", mimeo.

[4] Barsky, Robert, Jonathon Parker, and Gary Solon,(1994), "Measuring the Cyclicality of Real Wages: How Important is Compostion Bias?," *Quarterly Journal of Economics 109, 1-25.*

[5] Barsky, Robert, F.Thomas Juster, Miles S. Kimball and Matthew D. Shapiro, (1997), "Preference Parameters and Behavioral Heterogenity: An Experimental Approach in the Health and Retirment Study," *Quarterly Journal of Economics*, 112, 537-579.

[6] Basu, S. and M. Kimball (1997), "Cyclical Productivity with Unobserved Input Variation", NBER WP 5915.

[7] Baxter, M. and Robert King (1991), "Productive Externalities and Business Cycles", Discussion Paper no 53. Federal Reserve Bank of Minneapolis, Institute for Empirical Macroeconomics, November 1991.

[8] Baxter, M. and U.J. Jermann (1999), "Household Production and the Excess of Sensitivity of Consumption to Current Income," *American Economic Review,* 89(4), 902-920.

[9] Beaudry, Paul and Carmen Pages (2001): "The Cost of Business Cycle and the Stabilization Value of Unemployment Insurance," *European Economic Review* 45, no. 8, 1545-1572.

[10] Bernanke, Ben and Ilyan Mihov (1998), "Measuring Monetary Policy," *Quarterly Journal of Economics, CXIII, 869-902.*

[11] Blanchard Olivier, J. and Nobuhiro Kiyotaki, (1987), "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review,* LXXVII, 647-66.

[12] Burnside, Craig, and Martin Eichenbaum, (1996), "Factor Hoarding and the Business Cycel," *American Economic Review, 40, 861-69.*

[13] Card, David, (1994), "Intertemporal Labor Supply: An Assessment," in Frontiers of Econometrics, Christopher Sims, editor.

[14] Christiano, Lawrence, Martin Eichenbaum and Charle Evans, (1997), "Sticky Prices and Limited Participation Models: A Comparison," *European Economic Review,* 41, 1201-1249.

[15] Christiano, Lawrence, Martin Eichenbaum and Charle Evans, (1999), "Monetary Policy Shocks: What Have we Learned and to What End?," *Handbook of Macroeconomics, Vol.1A,*.Amsterdam North-Holland.

[16] Christiano, Lawrence, Martin Eichenbaum and Charle Evans, (2001), "Nominal Rigidities and the Effects of a Shock to Monetary Policy." mimeo Northwestern University.

[17] Cooley, Thomas and Edward Prescott, (1995) "Economic Growth and Business Cycles," in *Frontiers of Business Cycle Research,* Thomas Cooley (editor), Princeton University Press.

[18] Eichenbaum, Martin, Lars Hansen and Kenneth Singleton, (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics , 103 (1), February 1988, 51-78.*

[19] Erceg C., Dale Henderson, and Andrew Levin, (2000), "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics, 46(2), 281-313.*

[20] Francis, Neville and Valerie Ramey (2001), "Is the Technology-Driven Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," mimeo U.C.S.D, December.

[21] Galí, J. and Mark Gertler (1999), "Inflation Dynamics: A Structural Econometric Analysis", *Journal of Monetary Economics,* 44, 195-222.

[22] Galí, J., Mark Gertler and J.David López-Salido (2001), "European Inflation Dynamics", *European Economic Review, 45 (7),* June 2001, 1121-1150.

[23] Galí, J., Mark Gertler and J.David Lopez-Salido (2002): "Markups, Gaps, and the Welfare Costs of Business Fluctuations" NBER WP # 8850.

[24] Galí, Jordi (2001): "The Case for Price Stability. A Comment" in *Why Price Stability?*, edited by A. Garcia-Herrero et al., European Central Bank.

[25] Gordon, Robert J. (1973), "The Welfare Cost of Higher Unemployment," *Brookings Papers on Economic Activity no. 1,* 133-195. .

[26] Hall, Robert, E., (1988), "Intertemporal Substitution in consumption," *Journal of Political Economy, 96 (2), 339-357.*

[27] Hall, Robert, E. (1997): "Macroeconomic Fluctuations and the Allocation of Time", *Journal of Labor Economics*, vol.15,no.1,pt.2. 1997, S223-S250.

[28] Holland, A. and A. Scott, (1998), "Determinants of U.K. Business Cycles," *Economic Journal,* 108, 1067-1092.

[29] King, Robert and Sergio Rebelo, (1999):"Resuscitating Real Business Cycles", *The Handbook of Macroeconomics, Vol 1B,* Amsterdam North-Holland.

[30] Lucas, Robert E., (1987), *Models of Business Cycles, Oxford University Press.*

[31] Lucas, Robert E., (2003),:"Macroeconomic Priorities," *American Economic Review,* 93,1, 1-14.

[32] Mulligan, Casey B.(1998), "Substitution Over Time: Another Look at Life-Cycle Labor Supply," *NBER MacroAnnual*, vol.13, Ben Bernanke and Julio Rotemberg, editors.

[33] Mulligan, Casey B. (2002): "A Century of Labor-Leisure Distortions," NBER WP#8774.

[34] Pencavel, John, (1998), "The Market Work Behavior and Wages of Women, 1975-94," *Journal of Human Resources,* 33, 771-804.

[35] Rotemberg, Julio, and Michael Woodford (1999), "Oligolistic Pricing and the Effects of Aggregate Demand on Economic Activity," *Journal of Political Economy 100: 1153-1207.*

[36] Rotemberg, Julio, and Michael Woodford (1999), "The Cyclical Behavior of Prices and Costs," *The Handbook of Macroeconomics, Vol 1B,* Amsterdam North-Holland.

[37] Sbordone, Argia, M. (1999), "Prices and Unit Labor Costs: A New Test of Price Stickiness". mimeo, Rutgers University.

[38] Sbordone, Argia, M., (2000), "An Optimizing Model of Wage and Price Dynamics," mimeo Rutgers University.

[39] Staiger, Douglas, James Stock and Mark Watson (2001), "Prices, Wages and the U.S. NAIRU in the 1990s," mimeo.

[40] Uhlig, Harald, (2002), "What Moves Real GNP?," mimeo, paper presented at CEPR-ESSIM Conference, 22-26 May Tarragona, Spain.

[41] Woodford, M. (1999): "Inflation Stabilization and Welfare" NBER Working Paper # 5684.

Figure 1. The Gap: A Diagrammatic Exposition

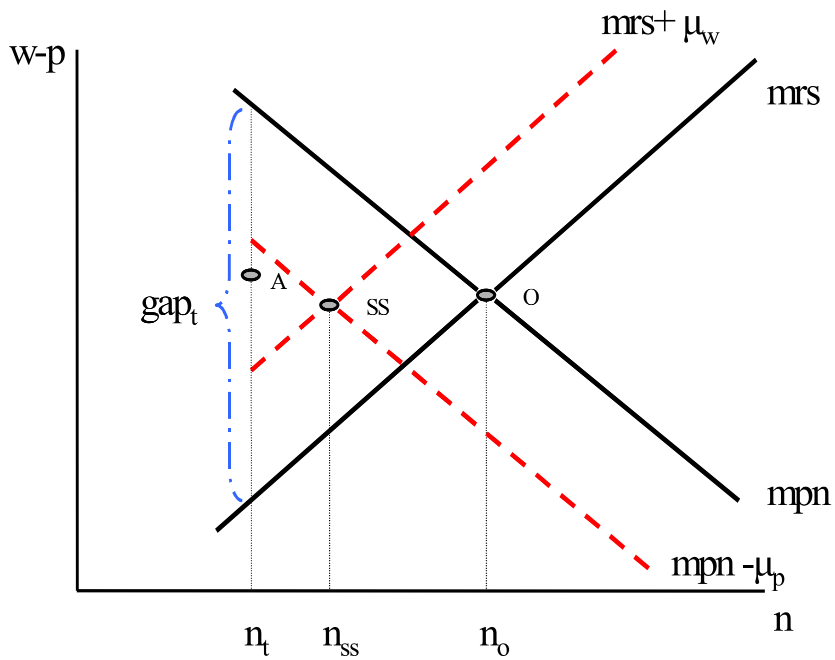

Figure 2. The Gap

Baseline Calibration (σ=5 , φ=1)

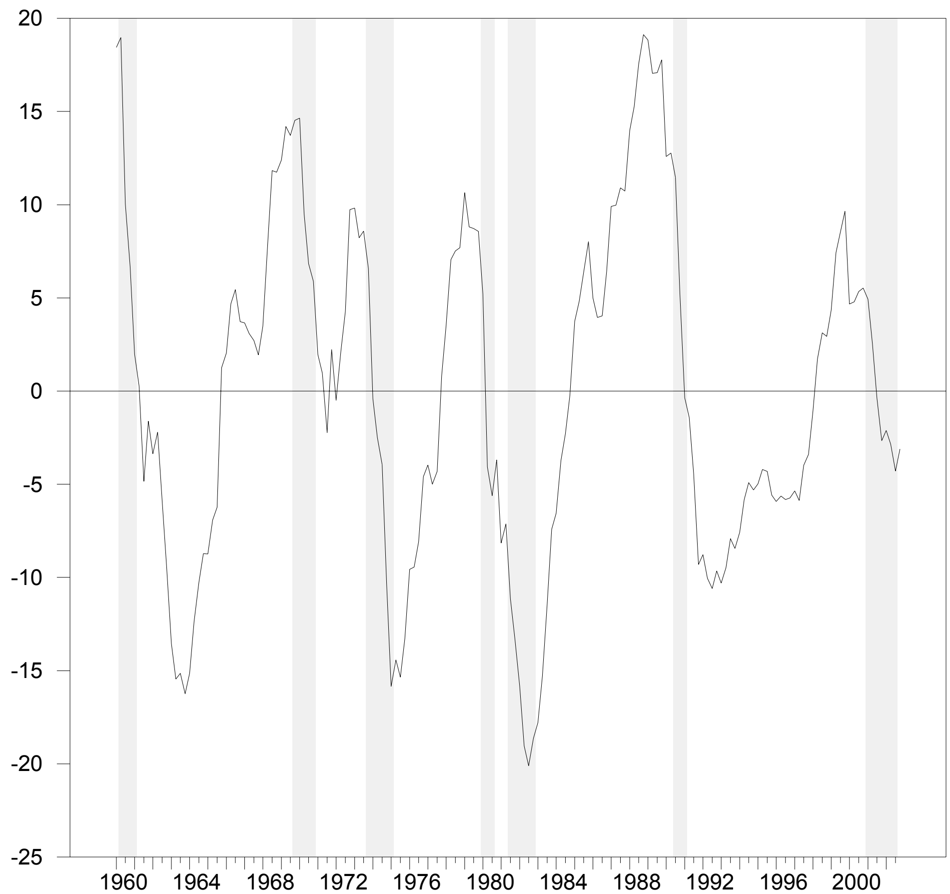

Figure 3. The Gap and the Wage Markup

Baseline Calibration

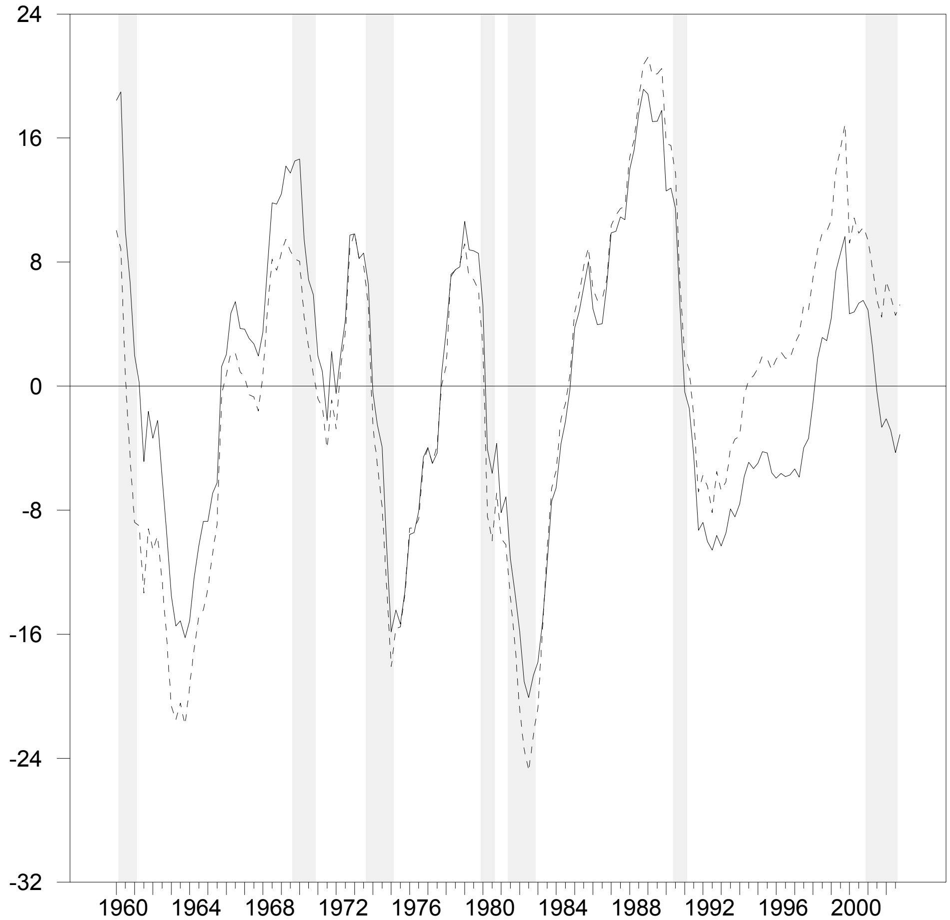

Figure 4. Dynamic Effects of Monetary Policy Shocks

Baseline Calibration

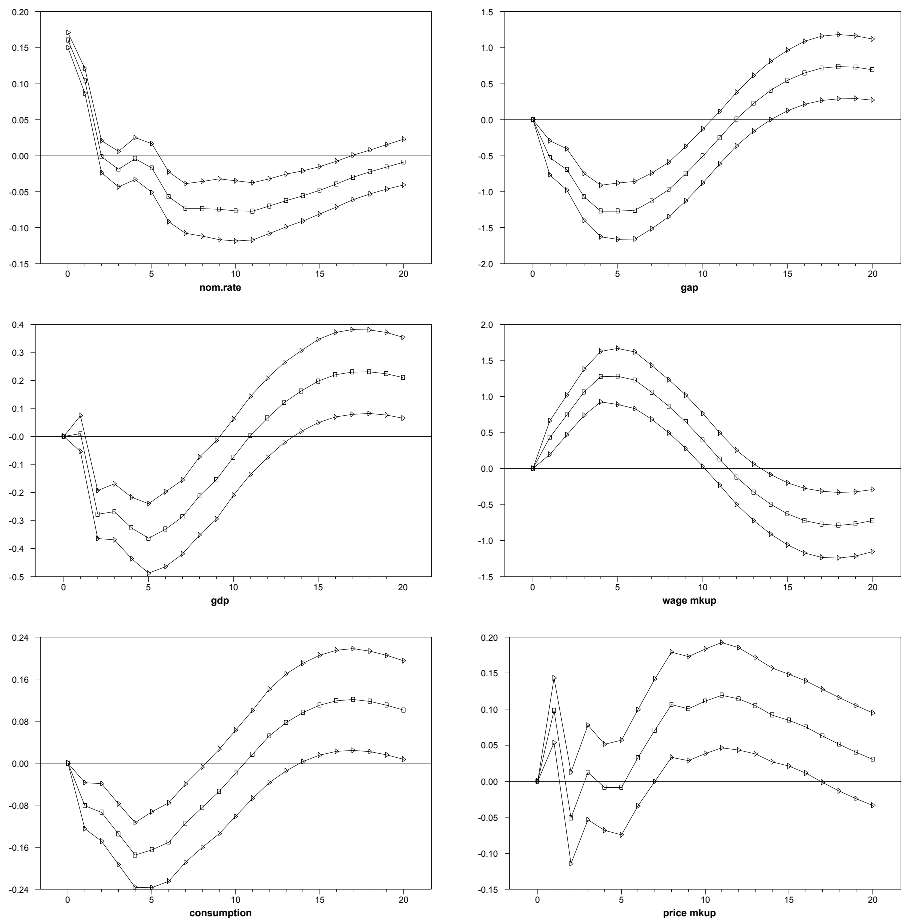

Figure 5. Theory vs. CBO based Output Gap

Baseline Calibration (σ=5 , φ=1)

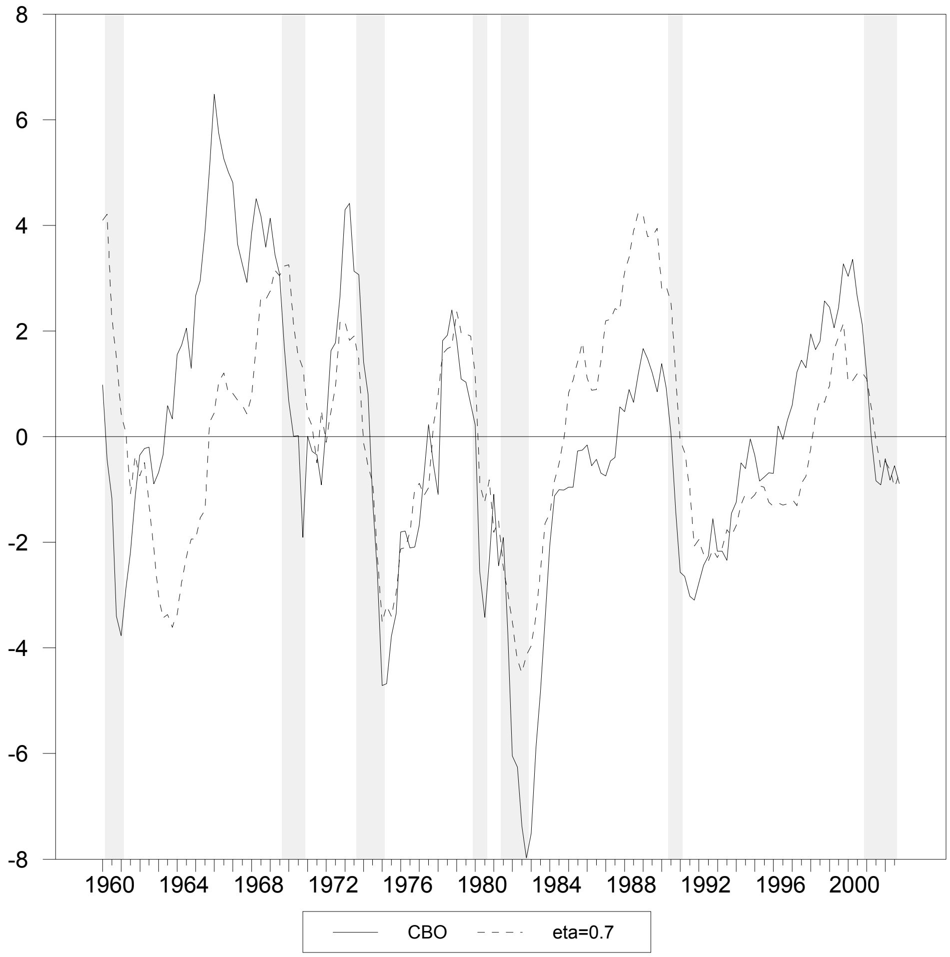

Figure 6. The Welfare Effects of Postwar U.S. Fluctuations

Baseline Calibration (σ=5, φ=1, μ=0.40)