

Capacity Utilization under Increasing Returns to Scale

Yi Wen

*Department of Economics, Hong Kong University of Science & Technology,
Clear Water Bay, Kowloon, Hong Kong, Peoples Republic of China*
E-mail: yiwen@usthk.ust.hk

Received February 14, 1998

This paper overcomes an important objection against the empirical relevance of the Benhabib-Farmer model as a potential account of actual business cycle fluctuations. This is attributable to an elasticity effect and a returns-to-scale effect of capacity utilization. These effects are closely related to the empirical puzzles that capital appears to play an insignificant role in explaining cyclical movements in output and that the estimated labor elasticity appears to be larger than labor's share. Due to these effects, multiple equilibria and persistent fluctuations can easily occur in a growth model for externalities mild enough so that the aggregate-labor-demand curve is downward sloping. Analyses show that the propagation mechanism generated by capacity utilization under mild increasing returns is capable of explaining the periodic patterns of U.S. business cycles documented by Watson. *Journal of Economic Literature* Classification Numbers: E32, E22, C52.

© 1998 Academic Press

1. INTRODUCTION

Capacity utilization is potentially a powerful driving force behind business cycles. Although its relationship to business cycles has long been noticed by economists (e.g., Marris [39], Lucas [38], Taubman and Wilkinson [50]), explicit analyses of capacity utilization in a dynamic general equilibrium framework are more recent.¹ An important finding of this literature is that capacity utilization can greatly amplify business cycle shocks, since it provides an additional margin to adjust the level of output.

Empirical analyses show that ignoring capacity utilization decreases the measured equilibrium capital–output elasticity and increases the measured equilibrium labor–output elasticity (notably Shapiro [46], and Burnside,

¹ See, e.g., Kydland and Prescott [36, 37], Greenwood, Hercowitz, and Huffman [31], Finn [29], Greenwood, Hercowitz, and Krusell [32], Bils and Cho [17], Burnside and Eichenbaum [20], Cooley, Hansen, and Prescott [25], and DeJong, Ingram, Wen, and Whiteman [26].

Eichenbaum, and Rebelo [21]). This “elasticity effect” of capacity utilization arises if and only if capacity utilization co-moves with labor and counter-moves with capital, which would be the case if an intensified utilization of capacity accelerates the rate of capital depreciation so that the marginal gain of capacity utilization is increasing in the level of employment but decreasing in the level of capital stock (as in the model of Greenwood *et al.* [31]). In addition, ignoring capacity utilization tends to bias the observed returns-to-scale upward (e.g., Shapiro [46], Beaulieu and Shapiro [10], Basu [5], and Burnside [19]). This “returns-to-scale effect” of capacity utilization arises because capacity utilization may act as an independent factor of production, especially in the presence of productive externalities.

Theoretical literature has shown that with sufficiently large returns-to-scale, an otherwise standard Real Business Cycle (RBC) model can exhibit multiple convergence paths toward a steady state (e.g., Benhabib and Farmer [11]).² This type of model can better explain the cyclical properties of the U.S. time series than the standard RBC model type, which assumes constant returns-to-scale production technologies (e.g., see Farmer and Guo [28]). However, in order to generate multiple equilibria in a one-sector RBC model, the degree of increasing returns-to-scale must be large enough to imply that the aggregate labor demand curve should be upward-sloping and steeper than the labor supply curve (Benhabib and Farmer [11], and Schmitt-Grohe [45]). These implications are inconsistent with empirical estimates on returns-to-scale.³

This paper provides a possible reconciliation for the two types of literature represented by Shapiro *et al.* on the one hand and by Benhabib *et al.* on the other hand. One of the implications of the empirical literature on capacity utilization represented by Shapiro *et al.* is that a growth model ignoring capacity utilization may demand larger returns to scale than necessary to explain the business cycle features of the data. Indeed, it is shown in the paper that explicitly taking into account the effects of capacity utilization makes multiple equilibria and endogenous cycles easier to occur for mild enough increasing returns-to-scale such that the aggregate labor demand curve remains downward sloping.

² There exists a large body of literature studying the possibility of sunspot-driven business cycles in dynamic models with multiple equilibria. An incomplete list of important works includes Azariadis [3], Benhabib and Farmer [11, 12], Benhabib and Rustichini [15], Boldrin and Rustichini [18], Cass and Shell [22], Christiano and Harrison [24], Farmer [27], Farmer and Guo [28], Gali [30], Rotemberg and Woodford [44], Shell [48], Woodford [54, 55, 56], as well as many others. For a review of this fast growing literature, please see Benhabib and Farmer [13].

³ See, e.g., Norrbin [40], Bartlesman *et al.* [4], Burnside [19], Basu [5], and Basu and Fernald [6], as well as others.

The cyclical mechanism emerges as follows: Capacity utilization induces employment to co-move with consumption under mild externalities, because it can dramatically increase the marginal productivity of labor under the “elasticity effect” and the “returns-to-scale effect.” Thus a rise in consumption demand stimulates output expansion and capital accumulation, which renders consumers’ initial optimistic expectations of future income self-fulfilling. This generates a cumulative process of output expansion. As the expansion continues, capacity utilization and employment gradually reach their upper limit. The diminishing marginal product of capital dictates that the increase of aggregate output becomes less and less. This means a declining growth of output and investment. The result is that sooner or later consumption must fall. But a reduction in consumption demand then triggers a contraction. This endogenous multiplier–accelerator mechanism based on the consumption-labor co-movement is the key in generating dampened, expectation-driven business cycles in the model.

This endogenous propagation mechanism generated by capacity utilization under mild increasing returns-to-scale is capable of explaining the periodic patterns of U.S. growth cycles documented by Watson [51]. For example, the model can produce spectral densities that peak at the business cycle frequency as in the data, and it can explain 37–77% of the variance distribution of U.S. output growth even when the only source of uncertainty is consumers’ expectations of future income.

Benhabib and Farmer [21] and Perli [41] demonstrated that when more than one production sector is incorporated into the one-sector RBC model, the degree of increasing returns-to-scale required to generate multiple equilibria can also be substantially reduced.⁴ Perli [42] also showed that his model is capable of generating substantial movements around the business cycle frequency if in addition to sunspot shocks, a perfectly correlated and highly persistent technology shock is also allowed.

Burnside and Eichenbaum [20] and DeJong, Ingram, Wen, and Whiteman [26] analyzed the role of variable capital utilization rates in propagating shocks over the business cycle. They found that variable capital utilization rates can significantly amplify the impact of technology shocks. Although capacity utilization helps to magnify the propagation mechanisms generated by labor hoarding or human capital accumulation, it is not by itself a fundamental source of business cycle propagation in these models. Restricted by constant returns-to-scale technologies, these models still need to rely on persistent total-factor productivity shocks to

⁴ Benhabib and Nishimura [14] showed that indeterminacy can even arise in multisector models that have constant aggregate returns to scale and very small market imperfections.

generate significant serial correlations in aggregate output.⁵ The model presented here, in contrast, requires only i.i.d shocks (e.g., transitory demand shocks) to generate strong serial correlations in output to match with the data. And the required degree of increasing returns-to-scale in the model is mild enough (on the magnitude of 1.1) to be justified by most recent empirical studies.

In the remainder of the paper, we first present the model and analyze the potential roles that capacity utilization may play in propagating business cycles. Necessary and sufficient conditions for indeterminacy are discussed. Then we calibrate and simulate the model to see whether its dynamic implications fit post-war U.S. business cycles. A criterion proposed by Watson [51] is employed as the measure of fit. Separate contributions of supply and demand shocks to post-war U.S. business cycles are also analyzed. Finally, some concluding remarks are provided at the end of the paper.

2. THE MODEL

The model is based on Greenwood *et al.* [31]. A representative agent chooses sequences of consumption $\{c_t\}$, hours to work $\{n_t\}$, rate of capacity utilization $\{u_t\}$, and capital stock $\{k_t\}$ to solve:

$$\begin{aligned} \max E_0 \sum_{t=0}^{\infty} \beta^t \left(\log(c_t) - \frac{n_t^{1+\gamma}}{1+\gamma} \right) \\ \text{s.t. } c_t + k_{t+1} - (1 - \delta_t)k_t = A_t \bar{e}_t (u_t k_t)^\alpha n_t^{1-\alpha}; \\ \delta_t = \tau u_t^\theta, \quad 0 < \tau < 1, \quad \theta > 1; \\ \bar{e}_t = (\bar{u}_t \bar{k}_t)^{\alpha\eta} \bar{n}_t^{(1-\alpha)\eta}; \end{aligned} \quad (1)$$

where $\gamma \geq 0$, $0 < \alpha < 1$, $\eta \geq 0$; and A_t is technology, $u_t \in (0, 1)$ is the rate of capacity utilization, $\delta_t \in (0, 1)$ is the rate of capital depreciation defined as an increasing function of capacity utilization,⁶ and \bar{e}_t is the productive

⁵ For example, in the model of Burnside and Eichenbaum, the propagation mechanism is essentially generated by labor hoarding due to which the response of employment to a technology shock takes place one period after the shock. Consequently, the response of output displays a hump in the second period of the shock. But labor hoarding does not necessarily imply a hump in the response of output unless the initial impact of technology shock is expected to persist in the model.

⁶ Thus the speed of capital depreciation is endogenously determined in the model. This is a theory of depreciation in use. As will be seen shortly, this theory can nicely explain the empirical puzzles regarding the capital and labor elasticities stressed recently by Burnside, Eichenbaum, and Rebelo [21].

externality expressed as a function of the average economy-wide levels of productive capacity and labor. The parametric restriction $\theta > 1$ is meant to impose a convex cost structure on capacity utilization so that it has an interior solution in the steady state. The viewpoint of increasing costs in adjusting the utilization margin is particularly stressed by Shapiro [46, 47]. If $\theta \leq 1$, the optimal rate of capacity utilization is always one (100%) and the model is reduced to that of Baxter and King [9] and Benhabib and Farmer [11].

To understand the role of capacity utilization in amplifying and propagating business cycles in this model economy, it is instructive to derive a reduced-form aggregate production function evaluated at the optimal rate of capacity utilization. The first order condition with respect to capacity usage u_t is⁷

$$\alpha \frac{y_t}{u_t} = u_t^{\theta-1} k_t, \quad (2)$$

where the LHS is the marginal output (y) gained by increasing the capacity utilization rate (u), and the RHS is the marginal loss in terms of capital depreciation due to the intensified usage of existing capital stock. Equation (2) can be rewritten as:

$$u_t = \left(\alpha \frac{y_t}{k_t} \right)^{1/\theta}, \quad (3)$$

which says that the optimal rate of capacity utilization is determined by the marginal product of capital. In other words, capital should be used more intensively during economic booms when its marginal product is high and less intensively during recessions when its marginal product is low. Using equilibrium conditions, one can use (3) to obtain an expression for the optimal capacity utilization rate in terms of aggregate capital and labor:

$$u_t = (\alpha A_t k_t^{\alpha(1+\eta)-1} n_t^{(1-\alpha)(1+\eta)})^{1/(\theta-\alpha(1+\eta))}. \quad (4)$$

Notice that capacity utilization is homogenous with degree zero in capital and labor only if the externality is zero. Otherwise, it is homogenous with a degree greater than zero. Finally, substituting (4) into the production function, we have

$$y_t = b A_t^{\tau n} k_t^{\alpha(1+\eta)\tau k} n_t^{(1-\alpha)(1+\eta)\tau n} \quad (5)$$

⁷ Since the parameter τ has no independent influence on the model's steady state and dynamics around the steady state, we simply set $\tau = 1/\theta$.

where the constants b , τ_k and τ_n are defined as

$$b = \alpha^{(\alpha(1+\eta)/(\theta - \alpha(1+\eta)))}, \quad \tau_k = \frac{\theta - 1}{\theta - \alpha(1 + \eta)}, \quad \tau_n = \frac{\theta}{\theta - \alpha(1 + \eta)}. \quad (6)$$

Expression (5) is the reduced-form aggregate production function, which indicates that capacity utilization effectively alters the equilibrium production function and amplifies technology shocks. These changes are mainly reflected by τ_k and τ_n . When the externality $\eta = 0$, it is easy to show that $\tau_k < 1$ and $\tau_n > 1$ since $\alpha < 1$ and $\theta > 1$.

Two effects are worth stressing. First, when the economy exhibits constant returns to scale ($\eta = 0$), capacity utilization has no effect on aggregate returns-to-scale (i.e., the factor elasticities sum to one):

$$\alpha\tau_k + (1 - \alpha)\tau_n = 1; \quad (7)$$

but it has significant effects on the distribution of factor elasticities—the capital elasticity decreases and the labor elasticity increases (because $\tau_k < 1$ and $\tau_n > 1$). For example, suppose the capital share of national income $\alpha = 0.3$ and the depreciation elasticity parameter $\theta = 1.4$ (which is the value calibrated by Greenwood *et al.* [31] according to the steady-state rate of capital depreciation $\delta = 0.025$ and the time discount factor $\beta = 0.99$), these then imply $\tau_k \approx 0.36$ and $\tau_n \approx 1.27$, which means that the effective capital-output elasticity is just about 0.1 while the actual capital-output elasticity is 0.3, and the effective labor-output elasticity is around 0.9 while its actual value is 0.7. This provides a possible explanation for the apparent empirical puzzle that the estimated capital elasticity is near zero and the estimated labor elasticity is near one (e.g., see Solow [49], Lucas [38], and Perry [43], among others).

This “elasticity effect” of capacity utilization is consistent with recent empirical findings of Shapiro [46] and Burnside, Eichenbaum, and Rebelo [21]. It arises because capacity utilization tends to co-move with labor and counter-move with capital (see Eq. 3). The reason is that the net marginal gain of capacity utilization is an increasing function of labor but a decreasing function of the capital stock at the steady state (due to the fact that capacity utilization accelerates the depreciation of existing capital stock). Thus, in addition to the direct multiplier effect of capacity utilization on amplifying technology shocks (from Eq. 5 one can see that this effect is τ_n), there is also an indirect multiplier effect resulting from the positive “elasticity effect” of capacity utilization on labor, which further amplifies technology shocks as it effectively increases the responsiveness of the production level to these shocks (remember that capital stock is fixed in the short term, so the adverse “elasticity effect” on capital does not matter).

Secondly, when the economy is subject to mild increasing returns to scale ($\eta > 0$), capacity utilization not only alters further the equilibrium

distribution of factor elasticities, but also has an effect on the aggregate returns-to-scale, since, if $\theta - \alpha(1 + \eta) > 0$,

$$\alpha(1 + \eta)\tau_k + (1 - \alpha)(1 + \eta)\tau_n > (1 + \eta). \quad (8)$$

This is called the “returns-to-scale effect” of capacity utilization. It can be shown that the “elasticity effect” of capacity utilization alone is not sufficient for explaining the fact that the estimated labor-output elasticity often exceeds one (e.g., Bernanke and Parkinson [16], and Burnside, Eichenbaum, and Rebelo [21]). The “elasticity effect” and the “returns-to-scale effect” combined together, however, are able to explain this well-known empirical puzzle. For example, given the previous parameterization, a mild degree of externalities in production, $\eta = 0.11$, would result in an observed labor-output elasticity around 1.02 while the actual elasticity is less than 0.78. In the following sections, we will see that these effects are the keys to understanding why capacity utilization is potentially a powerful driving force behind business cycles.

3. SOLVING THE MODEL

One nice property of the model is that it can be solved using reduced forms under conditions (3)–(5), so that its first-order conditions look almost identical to those without the variable of capacity utilization. These first order conditions are:

$$c_t = (1 - \alpha) \frac{y_t}{n_t^{(1+\gamma)}}, \quad (9)$$

$$\delta_t = \frac{\alpha}{\theta} \frac{y_t}{k_t}, \quad (10)$$

$$1 = \beta E_t \frac{c_t}{c_{t+1}} \left(\alpha \frac{y_{t+1}}{k_{t+1}} + (1 - \delta_{t+1}) \right), \quad (11)$$

$$c_t + k_{t+1} - (1 - \delta_t)k_t = y_t = bA_t^{\tau_n} k_t^{\alpha(1+\eta)\tau_k} n_t^{(-\alpha)(1+\eta)\tau_n}, \quad (12)$$

plus a transversality condition.⁸ The first equation determines the labor market equilibrium, the second equation determines the optimal rate of

⁸ It is easy to show that the maximization problem (1) is concave from the agent’s point of view, and the first-order conditions are thus sufficient for a maximum. Intuitively, notice that the reduced-form production function is homogeneous of degree one and concave from the agent’s point of view, and the reduced-form capital depreciation, $u^\theta k$, is proportional to output after substituting out capacity utilization using (4). Hence the model becomes the same as that of Benhabib and Farmer [11]. A more rigorous proof can be found in Greenwood *et al.* [31].

capacity utilization (or capital depreciation), the third one is the consumption Euler equation, and the last one is the consumer's budget constraint expressed at the optimal rate of capacity utilization using the reduced-form production function (5). From Eqs. (10) and (11), one can derive the following relationship in the steady state between the depreciation elasticity parameter and other structural parameters in the model:

$$\theta = \frac{(1 - \beta(1 - \delta))}{\beta\delta} \quad (13)$$

where δ is the steady-state rate of capital depreciation. It is easy to verify that $\theta > 1$ if $\delta \in [0, 1]$ and $\beta < 1$ (where β is the time discount factor).⁹

Equation system (9)–(12) does not have analytical solutions. Instead, we characterize the model's dynamics by linearizing these first order conditions around the steady state following King, Plosser, and Rebelo [35]. Using hat variables to denote the linearized variables, one can show that the system (9)–(12) can be reduced to the following linear dynamic systems under rational expectations:

$$\begin{pmatrix} \hat{k}_{t+1} \\ E(\hat{c}_{t+1} | t) \end{pmatrix} = B \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix}, \quad (14)$$

$$\begin{pmatrix} \hat{y}_t \\ \hat{i}_t \\ \hat{n}_t \\ \hat{u}_t \end{pmatrix} = \Pi \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix}, \quad (15)$$

where B and Π are real matrices, and technology A_t has been assumed to be constant for the purpose of exposition.

If the dynamic paths of capital and consumption can be solved from (14), dynamic paths of other model variables can then be solved using (15). In order to solve for (14), one must solve first for the decision rule of consumption in terms of capital stock, $c_t(k_t)$, since k_t is the only state variable

⁹ This suggests that in the Greenwood model θ is always greater than one and is not a free parameter. Moreover, the assumed functional form implies that the second additional parameter needed to calibrate this model, the elasticity of $\delta'(u_t)$, is also a function of θ so that the introduction of variable capacity utilization does not introduce an additional free parameter into the model. The model therefore does not nest the case where the depreciation rate and the utilization rate are constant. An alternative is to assume that $\delta_t = \delta_0 + \delta_1 u_t^\theta$ as in Basu and Kimball [8]. But their empirical results indicate that $\delta_0 = 0$, an assumption maintained in this paper. It is, however, important to point out that if we impose a value of δ_0 greater than zero then the required degree of externality for indeterminacy would be larger, other things equal.

given in the information set at time t . This requires that one of the eigenvalues of matrix B be explosive so that c_t can be uniquely solved forward as a function of k_t under rational expectations and the transversality condition. If both eigenvalues of B are stable, however, the model has multiple equilibria in the sense that c_t is indeterminate. In other words, any initial value of c_t is then consistent with the optimality conditions given by (9)–(12).

Also notice that if a saddle path exists (i.e., c_t can be solved forward in terms of k_t), then the only channel through which exogenous shocks can be propagated in the model economy is capital:

$$\hat{k}_{t+1} = \mu \hat{k}_t. \quad (16)$$

This means that the internal propagation mechanism of a standard RBC model with saddle path stability can never be richer than the law of motion of capital itself (characterized by an AR(1) process) unless additional endogenous state variables are introduced. Therefore, having c_t in the state space of the model (i.e., c_t is indeterminate) is a key for an model to exhibit complex dynamics, since the eigenvalues of B can then form a complex conjugate pair.

4. NECESSARY AND SUFFICIENT CONDITIONS FOR INDETERMINACY

A necessary and sufficient condition for indeterminacy is that both eigenvalues of B are less than one in modulus. This is the case if and only if the determinant and the trace of the transition matrix B satisfy

$$-1 < \det(B) < 1 \wedge -(1 + \det(B)) < \text{tr}(B) < (1 + \det(B)). \quad (17)$$

The determinant and the trace of B can be found to be (see the Appendix):

$$\det(B) = \frac{1}{\beta} \left(1 + \frac{\eta(1+\gamma)(1-\beta)\tau_n}{(1+\gamma) - \beta(1-\alpha)(1+\eta)\tau_n} \right), \quad (18)$$

$$\text{tr}(B) = 1 + \det(B) + \frac{(1+\gamma)(1-\beta)(\theta-\alpha)(1-\alpha(1+\eta)\tau_k) \frac{\delta}{\alpha}}{(1+\gamma) - \beta(1-\alpha)(1+\eta)\tau_n}, \quad (19)$$

where τ_k and τ_n are defined in (6). Notice that when an externality does not exist ($\eta = 0$), the determinant is simply $1/\beta$ (> 1), indicating a saddle-path stability as in a standard RBC model.

Notice that τ_k and τ_n are positive and finite when $\eta = 0$. To maintain this property in the presence of externalities, we restrict our analyses to cases where

$$\alpha(1 + \eta) < \theta. \quad (20)$$

The common denominator in expressions (18) and (19) suggests that when the externality parameter increases from zero, the model may go through a point of discontinuity at which $\det(B)$ and $\text{tr}(B)$ both change sign, passing from $+\infty$ to $-\infty$, if the conditions $\tau_n > 0$ and $\alpha(1 + \eta)\tau_k < 1$ still hold. These two conditions imply

$$\alpha(1 + \eta) < 1. \quad (21)$$

(Notice that $\tau_k > 0$ when (21) is satisfied and that (21) implies (20).) To reach the discontinuity point, however, we also need

$$\beta(1 - \alpha)(1 + \eta)\tau_n \geq (1 + \gamma). \quad (22)$$

This is an important necessary condition for indeterminacy. It is analogous to that derived by Benhabib and Farmer [11] in a continuous time model without variable capacity utilization, since the left-hand side of it is nothing but the *effective* labor-output elasticity evaluated at the optimal capacity utilization rate. This condition hence has a simple interpretation: The equilibrium wage-hours locus in the labor market need to be positively sloped and to cut the labor supply curve from below to generate indeterminacy. This condition implies:

$$\eta > \frac{\theta(1 + \gamma - \beta(1 - \alpha)) - (1 + \gamma)\alpha}{\beta(1 - \alpha)\theta + (1 + \gamma)\alpha}. \quad (23)$$

It needs to be verified that requirements (23) and (21) are compatible for certain parameter values of the model. These two requirements together imply:

$$\beta(1 - \alpha)(1 - \beta(1 - \delta)) > (1 + \gamma)(1 - \beta)\alpha. \quad (24)$$

It is obvious that there exist regions of the parameter space such that (24) is satisfied (e.g., for β close to one, or for α small enough).

If indeterminacy exists in the parameter regions specified above, the second term in $\det(B)$ and the third term in $\text{tr}(B)$ must increase from $-\infty$ to a finite number as η keeps increasing from the point of discontinuity. Therefore, we can limit our attention to the following simpler one-sided conditions as necessary and sufficient conditions for indeterminacy:

$$-1 < \det(B) \wedge -(1 + \det(B)) < \text{tr}(B), \quad (25)$$

since we are interested only in the smallest value of η that gives rise to indeterminacy. The condition $-1 < \det(B)$ implies

$$\eta > \bar{\eta} = \frac{\theta(1 + \gamma - \beta(1 - \alpha)) - (1 + \gamma)\alpha}{\beta(1 - \alpha)\theta + (1 + \gamma)\alpha - \frac{1 - \beta}{1 + \beta}(1 + \gamma)\theta}. \quad (26)$$

Comparing condition (26) with (23), there is an extra term in the denominator of the right-hand side of (26) making (26) more demanding for the size of η .¹⁰ The condition $-(1 + \det(B)) < \text{tr}(B)$ implies

$$\eta > \frac{\theta(1 + \gamma - \beta(1 - \alpha)) - (1 + \gamma)\alpha + \frac{(1 - \beta)}{(1 + \beta)}(1 + \gamma)\beta\delta(\theta - \alpha)\left(\frac{1 - \alpha\theta}{2\alpha}\right)}{\beta(1 - \alpha)\theta + (1 + \gamma)\alpha - \frac{(1 - \beta)}{(1 + \beta)}(1 + \gamma)\left(\theta - \frac{1}{2}(\theta - \alpha)(1 - \beta)\right)}. \quad (27)$$

Notice that condition (27) differs from (26) by a third term in both the denominator and the numerator, both of which vanish as $\beta \rightarrow 1$. Under a realistic value of β that is close to one, these two conditions are virtually the same. Using (26), it is easy to show that the following first derivatives hold in the permissible region for η :

$$\frac{\partial \bar{\eta}}{\partial \beta} < 0, \quad \frac{\partial \bar{\eta}}{\partial \delta} < 0, \quad \frac{\partial \bar{\eta}}{\partial \gamma} > 0, \quad \frac{\partial \bar{\eta}}{\partial \alpha} > 0. \quad (28)$$

That is, indeterminacy occurs more easily the larger the labor supply elasticity ($1/\gamma$) and the labor demand elasticity, $(1 - \alpha)(1 + \eta)\tau_n - 1$, are. (Note that τ_n is decreasing in both δ and β when β is near one).

The necessary and sufficient conditions for indeterminacy are therefore (27), (26), and (21). To show that the parameter region thus specified for η is not empty for realistic parameter values of the model, consider the parameterization of Farmer and Guo [28] in a quarterly model: $\gamma = 0$ (Hansen's [33] indivisible labor), $\alpha = 0.3$, $\beta = 0.99$, and $\delta = 0.025$ (implying $\theta = 1.4$). Figure 1 shows the regions of indeterminacy as functions of η when each of these four parameters is allowed to vary while holding the other three constant. It is seen there that conditions (27) and (26) are virtually the same at $\beta = 0.99$, and that the permissible regions for indeterminacy are very large.

¹⁰ But that term is insignificant if the time discount factor β is near one.

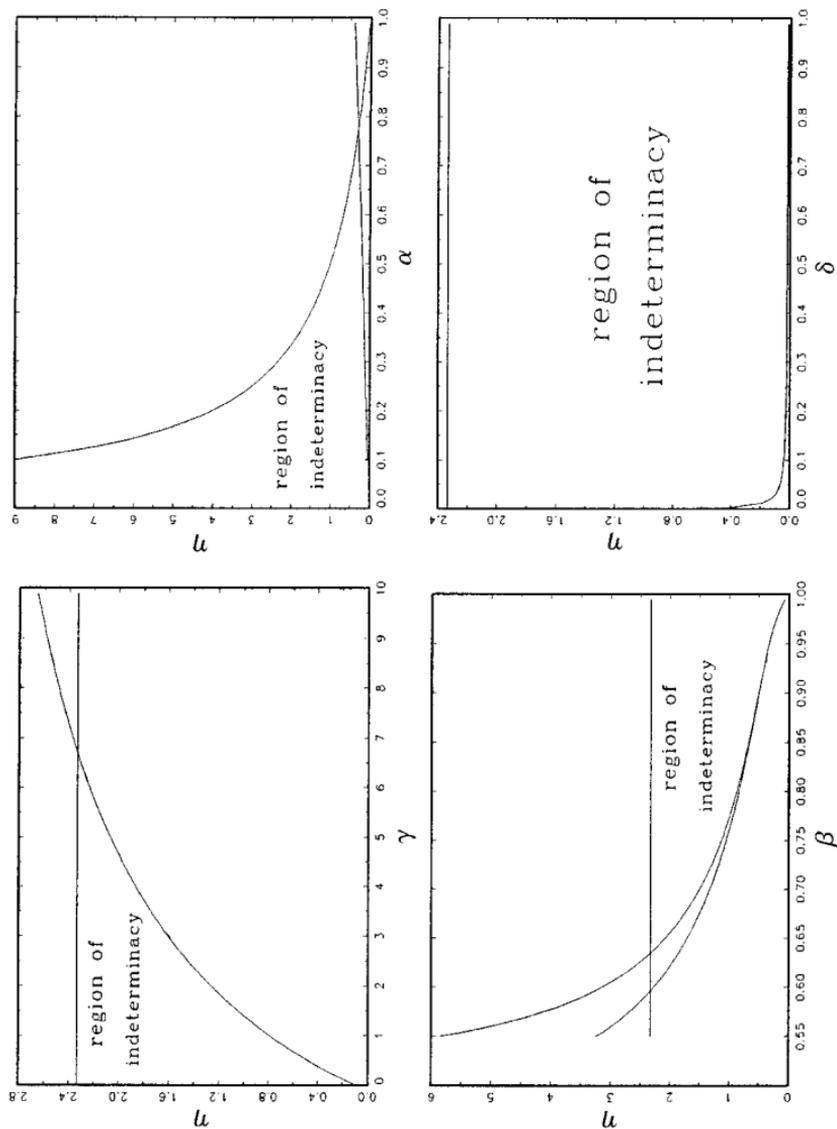


FIG. 1. Regions of indeterminacy (η is on the vertical axis). For each window, the upper line surrounding the region of indeterminacy is given by the constraint $(1 + \eta) < 1$ (condition (21) in the text), the bottom lines are given by constraints $\det(B) > -1$ and $\text{tr}(B) > -(1 + \det(B))$, which imply conditions (26) and (27) in the text. Notice that the bottom two lines in each window are hardly distinguishable except for window $[2, 1]$ when β is small.

From the necessary condition (22), it is evident that this model requires a smaller degree of increasing returns than the corresponding model with fixed capacity utilization (Benhabib-Farmer [11]) to induce indeterminacy due to the presence of the term τ_n , which arises because of the effect of capacity utilization. For example, at the above parameterization, the required value of η for inducing indeterminacy is 0.1036, implying a downward sloping aggregate labor demand curve (slope = $(1 - \alpha)(1 + \eta) - 1 = -0.23$) and a mild increasing return-to-scale ($1 + \eta \approx 1.1$). If the capacity utilization were fixed as in the model of Farmer and Guo, however, the minimum degree of externality required for generating indeterminacy would be 0.4927, implying a substantially larger return-to-scale ($1 + \eta \approx 1.5$) and an upward sloping aggregate labor demand curve.

The insight is that variable capacity utilization increases the elasticity of output with respect to labor and that, in the presence of mild external increasing returns to scale, this effect can be sufficient to push the labor elasticity of output above one. In equilibrium the marginal product of labor is thus increasing in labor rather than decreasing as in the standard model. And as shown by Benhabib and Farmer [11], this implies that the rational-expectations equilibrium is indeterminate.

The intuition can also be understood using the labor supply and demand curves (Aiyagari [2]). Since the capacity utilization rate can respond to changes in consumption level at the impact period, an upward shift of the

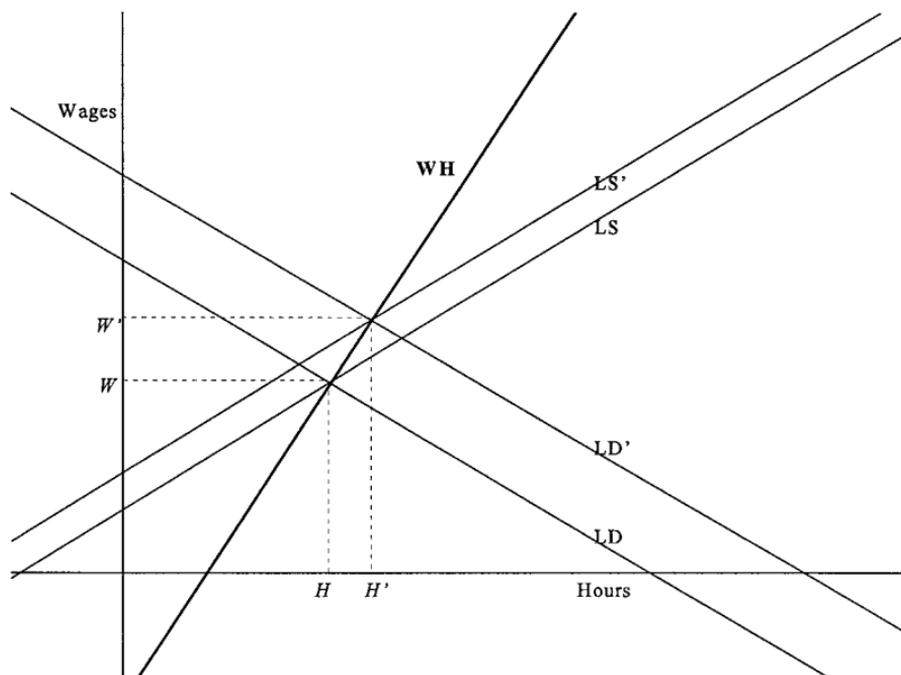


FIG. 2. Indeterminacy and the labour market.

labor supply curve caused by an increase in the initial consumption level can at the same time trigger an upward shift of the aggregate labor demand curve (because of increases in the capacity utilization rate). If this shift in the labor demand curve is large enough, then equilibrium labor and real wages will both increase (indicating a rise in the permanent income), which substantiates the initial increase in consumption (see Fig. 2, where LS represents aggregate labor supply, LD represents aggregate labor demand, and WH represents the equilibrium wage-hour loci). If capacity utilization were fixed, however, an upward shift of the labor supply curve would result in a decrease in equilibrium labor and output unless the aggregate labor demand curve was upward sloping and was steeper than the labor supply curve (as in the model of Benhabib and Farmer [11]). Thus, capacity utilization explains why multiple equilibria may emerge in a one-sector growth model with a downward sloping aggregate labor demand curve.

5. A SIMULATION EXERCISE

The dynamic properties of the model can be better appreciated in a simulation exercise, for we can get a quantitative sense about how the impulse responses of the economy depend on the underlying structures of the model. We adopt a parameterization standard in the literature: $\alpha = 0.3$, $\gamma = 0$ (Hansen's indivisible labor), $\beta = 0.99$, and $\delta = 0.025$ (which implies $\theta = 1.4$ according to Eq. 13). In order to set the value for η , we utilize the dynamic properties of capacity utilization in the data. Capacity utilization in the U.S. manufacturing sector can be reasonably characterized as a stationary AR(2) process, where the first lag coefficient is 1.4 and the second is -0.54 (the R^2 of this autoregression is 0.88 and the Durbin-Watson statistic is 1.95). These coefficients imply that the time series of capacity utilization has a pair of complex roots $(0.7 \pm 0.23i)$, indicating dampened cycles at frequency of about 0.05 (cycles per quarter). In the current model, an externality of 0.11 would imply a dynamic system (characterized by the eigenvalues of the transition matrix B in Eq. 14) that roughly matches the dynamics of the U.S. manufacturing sector's capacity utilization rate in terms of cycle frequency.¹¹ Simulations carried out in the following, therefore, use 0.11 as a benchmark value for externality, which can be viewed as a calibrated value using capacity utilization. The degree

¹¹ The frequency of cycles for a given pair of complex roots, $\alpha \pm \beta i$, is determined by

$$\cos^{-1} \left(\frac{\alpha}{\sqrt{\alpha^2 + \beta^2}} \right) \frac{1}{2\pi}.$$

of aggregate returns-to-scale in the range of 1.05 to 1.15 is empirically plausible even judged by most recent empirical estimates (e.g., see Basu and Fernald [7], and Burnside, Eichenbaum, and Rebelo [21]).¹²

Figure 3 shows the dynamic responses of output, investment, and consumption to an unexpected one percentage increase in the initial consumption level due to the agent's optimistic expectations about future income. The initial response of output is about 55%, and that of investment is about 253%. Thus the model correctly predicted the comovements among these variables. However, the extremely large multiplier effect seems to be empirically implausible. This is attributable to the extremely volatile labor in the model. At the optimal rate of capacity utilization, the consumption elasticity of labor can be determined from the log-linearized first-order condition for labor:

$$\hat{n}_t = \frac{1}{(1-\alpha)(1+\eta)\tau_n - (1+\gamma)} \hat{c}_t, \quad (29)$$

where the capital stock is ignored since it does not respond to shocks at the impact period. Given the parameterization adopted above, Eq. (29) implies a 54% increase in labor in responding to the 1% increase in consumption demand. This is so since the effective labor elasticity of output evaluated at the optimal rate of capacity utilization is near one ($= (1-\alpha)(1+\eta)\tau_n = 1.0186$), and thus the slope of the reduced-form labor demand curve evaluated at the optimal rate of capacity utilization is near zero ($((1-\alpha)(1+\eta)\tau_n - 1 = 0.0186)$). Combined with a flat labor supply curve (at $\gamma = 0$), it means that a slight shift in either of the two curves can cause a huge movement in equilibrium labor. This also means that the real wage is nearly constant, as is the consumption level.¹³

A very smooth consumption path relative to that of income then implies a very volatile investment path, since investment is the perfect buffer for

¹² At the calibrated externality ($\eta = 0.11$), the elasticity effect and the returns-to-scale effect of capacity utilization are indicated by $\tau_k = 0.377$ and $\tau_n = 1.311$, which means that under the optimal rate of capacity utilization the effective capital elasticity of output is just 0.126 while the effective labor elasticity is around 1.0186. As was mentioned before, this provides an explanation for the apparent empirical puzzles that the estimated capital elasticity is nearly zero and that of labor is greater than one when capacity utilization is omitted from the production function, and is consistent with empirical findings of Shapiro [46] and Burnside, Eichenbaum, and Rebelo [21]. Although their empirical findings do not reject the hypothesis of constant returns-to-scale increasing returns-to-scale around the degree of 1.1 is perfectly inside the permissible range of their empirical estimates.

¹³ The dynamic path of consumption in the model is the same as that of the real wage when other fundamental shocks are not present.

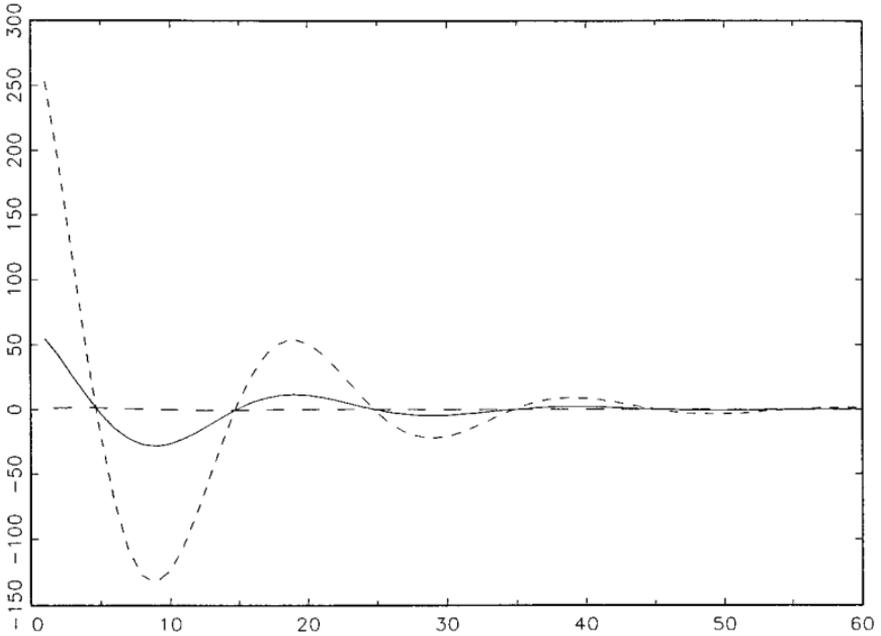


FIG. 3. Impulse responses of consumption (long dashed lines), output (solid lines), and investment (short dashed lines) to a 1% standard deviation in consumption demand at the steady state.

consumption smoothing. Quantitatively speaking, the standard deviations of consumption, investment and output almost satisfy

$$(1 - s)\sigma_c + s\sigma_i = \sigma_y, \quad (30)$$

(where s is the steady-state saving ratio) under the assumption that investment is procyclical and is almost perfectly correlated with output.¹⁴ Hence, the standard deviation of output is approximately a convex combination of that of consumption and investment. For example, suppose that the relative volatility of consumption to output is very small and the steady-state savings ratio is approximately 0.2, then Eq. (24) implies a relative volatility of investment to output to be approximately 5.

Table I reports some of the basic RBC statistics of the model, assuming that the expectation shocks are serially uncorrelated. All variables are positively correlated with output. The relative variability of consumption to output is 0.04, that of investment to output is 4.6, that of labor to output is 0.99, and that of capacity utilization to output is 0.76. Although no data

¹⁴ To derive (30), log-linearize the income identity, $c_t + i_t = y_t$, we get $(1 - s)\hat{c}_t = \hat{y}_t - s\hat{i}_t$, which implies $(1 - s)^2\sigma_c^2 \approx \sigma_y^2 - s^2\sigma_i^2 - 2s\sigma_y\sigma_i$, under the assumption that i_t and y_t are almost perfectly correlated. Rearranging terms by factorization then gives the result.

TABLE I

Sample and Population Moments

Var	U.S. Sample 1948:1-1994:2			Model Population Moments		
	σ_x/σ_y	Cor(xy)	Autocar	σ_x/σ_y	Cor(x, y)	Autocar
y	1.00	1.00	0.95	1.00	1.00	0.79
c	0.74	0.88	0.97	0.04	0.37	0.93
inv	2.63	0.87	0.92	4.63	0.99	0.79
n	0.88	0.62	0.92	0.99	0.99	0.79
u	—	—	—	0.76	0.96	0.80
y/n	0.83	0.55	0.90	0.04	0.37	0.93

Note. Variables (y, c, inv, n, u, y/n) stand for output, consumption, investment, labor, capacity utilization, and productivity, respectively. The U.S. data were detrended by a quadratic time trend.

are available for the aggregate U.S. economy's capacity utilization rate, data from the manufacturing sector indicate that capacity utilization is about as volatile as the output. Although the model generates a productivity that is perhaps too smooth (in the absence of technology shocks), it is nevertheless procyclical ($\text{cor}(y/n, y) = 0.37$). Procyclical productivity is one of the key motivations for the technology-shock-based explanation of business cycle fluctuations. Here, despite the source of shock being from the demand side, productivity is procyclical. This is attributable to the "elasticity effect" and the "returns-to-scale effect" of capacity utilization.

Perhaps one of the most striking features of the model is the persistent oscillations shown in Fig. 3. In responding to a purely transitory consumption shock, the model displays cycles with a half-life of nearly eight quarters. An economic expansion is followed by a recession, which in turn is followed by another expansion. The cyclical mechanism arises due to an intrinsic multiplier-accelerator mechanism: A rise in consumption demand stimulates output expansion and capital accumulation; the diminishing marginal product of capital, however, dictates that capacity utilization and employment must fall as the capital stock rises, which leads to negative growth of output and consumption. As output and consumption decline, investment (as well as the capital stock) falls and the marginal product of capital rises eventually to a point where increasing capacity utilization and employment become optimal. This triggers another period of expansion.

Capacity utilization plays an important role in giving rise to this oscillation mechanism: The elasticity effect and the returns-to-scale effect of capacity utilization create short-run increasing returns to labor under mild externalities, which induce hours to comove with consumption demand.

Without such comovement between consumption and labor hours, the multiplier-accelerator control mechanism would not have arisen.¹⁵

6. MEASURING THE FIT

Although it was shown by Farmer and Guo [28] that models with indeterminacy driven solely by i.i.d expectation shocks can perform at least as well as a standard RBC model driven by persistent technology shocks in explaining the basic business cycle facts of the data and can outperform a standard RBC model in explaining the propagation mechanisms of the average U.S. business cycles, it is not clear whether this type of model can generate sufficient fluctuations around the business cycle frequency. For example, the parameterization adopted by Farmer and Guo [28] in their model as an attempt to match the second moments of the data produces cycles at frequency of 0.0163 (cycles per quarter), implying a periodicity of 61 quarters per cycle. This obviously does not match the average U.S. business cycle (with periodicity of 20 quarters per cycle at frequency 0.05 cycles per quarter). Furthermore, it generates essentially the same spectral density functions as those implied by the King–Plosser–Rebelo [35] model (see Wen [53]).

This motivates a formal assessment of the fit of the current model using criteria proposed by Watson [51]. Watson showed that a standard RBC model with constant returns-to-scale and driven by persistent technology shocks cannot generate enough movements around the business cycle frequencies to replicate the spectra of the U.S. data. The spectral density function reveals the distribution pattern of the variance of a time series across different frequencies. The time-domain statistics can only tell us the total variance of a time series, but cannot tell us from which cyclical frequencies this variance comes. For example, the variance of the U.S. GNP growth rate is not distributed evenly across frequencies. Its spectrum shows a distinctive peak around the five-year cycle frequency and indicates that about 60% of its total variance is contributed by movements around this business cycle frequency. The variance of output growth implied by a standard RBC model, however, is distributed evenly across frequencies and can only explain about 27% of the data's volatility around the business cycle frequency, although the total variance of the model may match that of the data very well (see Wen [52]). Since the spectrum is a weighted sum of

¹⁵ In a standard RBC model with constant returns-to-scale and fixed capacity utilization rate, employment falls in response to a rise in consumption demand. This is because an expansion of labor effort, given the short-run fixed supply of capital at the full capacity utilization rate, causes labor's productivity to decline. Consequently, intertemporal substitution induces agents to increase leisure and cut labor supply.

the autocovariance functions under the Fourier transform, the cyclical dynamics implied by the standard RBC model are thus very different from those implied by the data. Therefore, although the standard RBC model can match some second moments of the data in the time domain, it nevertheless fails the Watson criterion dramatically in the frequency domain. In this section, we examine whether a model of variable capacity utilization driven by expectation shocks can outperform a standard RBC model under the Watson criterion.¹⁶

To assess the predictive power of belief shocks propagated through the mechanism of capacity utilization under increasing returns-to-scale, we rewrite Eq. (14) as:

$$\begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = B \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_{t+1} \end{pmatrix}, \quad \varepsilon_t \sim i.i.d(0, \sigma^2), \quad (26)$$

where ε is a random variable with zero mean conditional on time t information and standard deviation σ , and can be interpreted as shocks to autonomous consumption demand or consumer sentiment.¹⁷ The corresponding spectral density function is:

$$f_s(e^{-i\omega}) = (I - B e^{-i\omega})^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} (I - B' e^{i\omega})^{-1}, \quad (27)$$

where ω is the frequency and is measured in cycles per quarter with $-\pi \leq \omega \leq \pi$. The spectral density functions of other model variables in equation (15) can be determined by

$$f_y(e^{-i\omega}) = \Pi f_s(e^{-i\omega}) \Pi'. \quad (28)$$

Using parameter values specified before and choosing the standard deviation of the demand shock (σ) so that the total variance of output growth in the model matches that of the data (this gives $\sigma = 0.00013$), Fig. 4 plots the variance distributions (spectral density functions) of both

¹⁶ Wen [52] showed that increasing returns-to-scale, combined with habit formation on leisure, can substantially improve the empirical fit of a standard RBC model using the Watson criterion. That model, however, does not allow indeterminacy to arise due to the fact that adjustment costs in labor supply can insulate an economy from expectation-driven fluctuations. In addition, that model requires a substantially larger aggregate return to scale (in the magnitude of 1.5) to match the data.

¹⁷ The current literature commonly names it as belief shocks or simply sunspots.

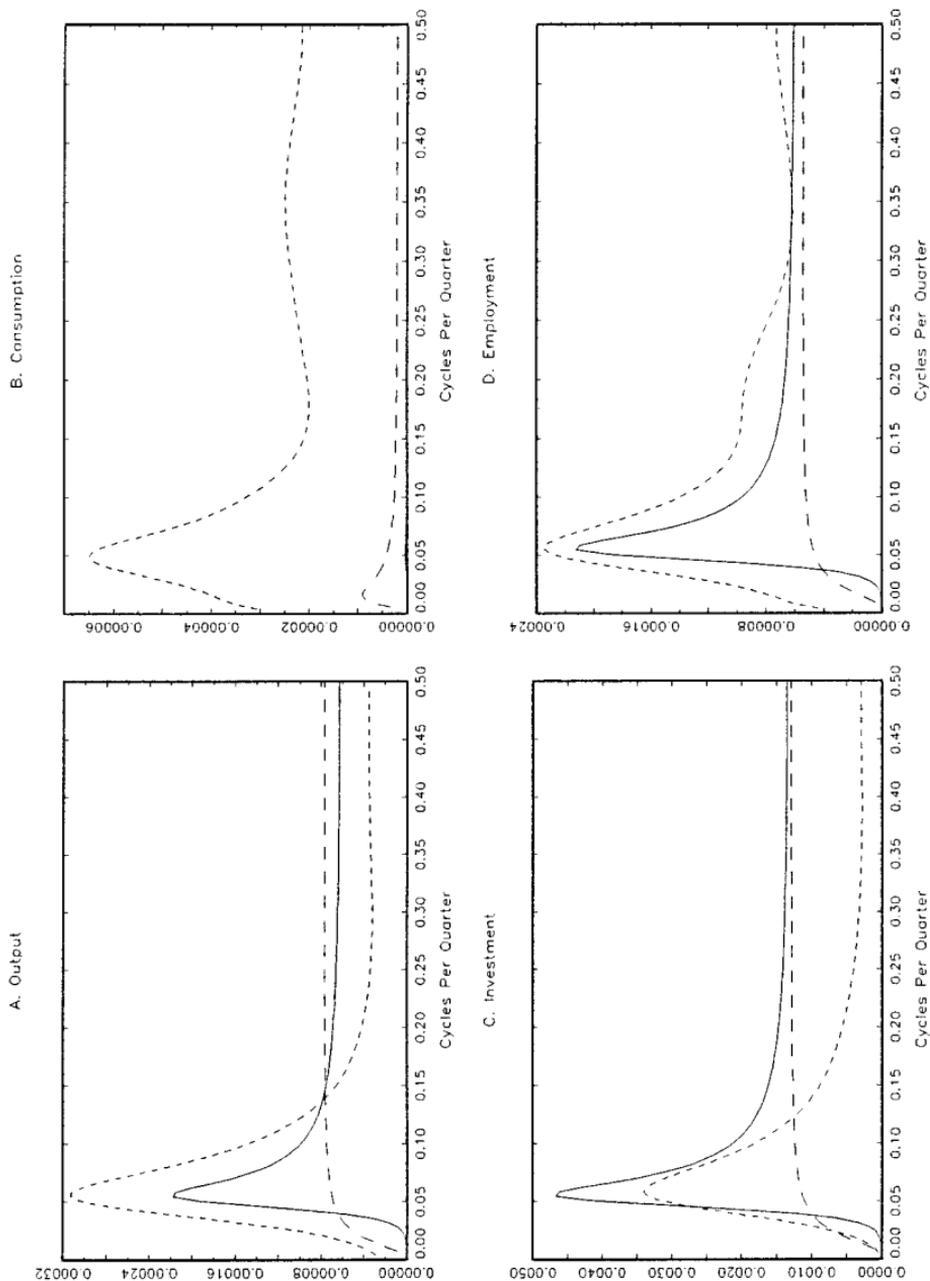


FIG. 4. Spectra of data (short dashed lines), current Model (solid lines), and KPR model (long dashed lines).

the model and the data in terms of growth rates.¹⁸ To facilitate comparison, the spectra generated from a standard RBC model (King, Plosser and Rebelo, KPR [35]) are also graphed.¹⁹

From the spectral shape alone, it is clear that the current model is apparently doing a better job in replicating the dynamics of the data than the KPR model, especially around the business cycle frequency. Specifically, the model is capable of generating sufficient movements around the business cycle frequency with a distinctive spectral peak similar to that in the data. In contrast, the spectra of the KPR model are essentially flat across different frequencies.

Next we address Watson's question: How much stochastic error would have to be added to the current model so that the autocovariances of the augmented model are equal to the autocovariances of the data? If the variance of the required error is large, then the discrepancy between the model and data is large, and the converse is true if the variance of the error is small. Table II presents the relative mean square approximation error (RMSAE) needed to reconcile the spectra of the model and data. Also reported in Table II are the RMSAEs of the KPR model. The RMSAE statistics are analogous to $1 - R^2$ statistics in regression analysis—smaller is better.

The performance of the current model in explaining consumption is not good: Consumption is simply too smooth in the model (its spectrum cannot even be detected from the graph), and its RMSAE is nearly 100%, meaning that the model explains practically nothing for consumption. The model does a better job in characterizing the remaining series. The RMSAE

¹⁸ The spectra of the model's growth rates were obtained by taking the first difference of Eq. (24). To compute the spectra of the data, a VAR was estimated with four lags included for output, consumption, investment, and employment. The quarterly data used here are from CITIBASE, 1948:1–1994:2. The consumption measure is total real consumption expenditure on nondurables, and the measure of investment is the total real fixed investment. The measure of output is defined as the sum of consumption and investment. The measure of employment is total labor hours in private nonagricultural establishments. All variables are expressed in per capita terms using the total civilian, noninstitutional population over the age of 16. With the CITIBASE labels, the precise variables used were *gcnq* for consumption and *gigq* for investment. The measure of total labor was constructed as $(lhem-lpgov)(lhch)$. The population series was *p16*. The variables are logged before estimation, so collinearity was not a problem in the regression. The spectra of growth rates were then obtained by applying the first different filter when taking the Fourier transform on the estimated VAR.

¹⁹ Since the capacity utilization model of Greenwood *et al.* [31] with constant returns behaves very much like the KPR model in terms of endogenous propagation mechanism (see Eq. 16), we compare the current model with the KPR model only. In obtaining the spectra of the KPR model, the structural parameters $(\alpha, \beta, \delta, \gamma)$ were set at the same value as those in the current model, the persistence parameter of technology shocks was set to 0.9, and the standard deviation of technology shocks was also chosen such that the total variance of output growth in the model matches that of the data exactly.

TABLE II
Relative Mean Square Approximation Error

Variable	All Frequencies		Business Cycle Frequencies	
	Our Model	KPR Model	Our Model	KPR Model
Output	0.29	0.34	0.12	0.21
Cons.	0.98	0.83	0.97	0.77
Invst.	0.38	0.36	0.05	0.10
Labor	0.71	0.72	0.39	0.47

Note. Numbers shown in the table are the RMSAE statistics ($1-R^2$) of different models, representing the relative mean square approximation error required for a model to explain the spectra of the data. In all cases, relative equal weights (normalized by variables' relative standard deviation to output) are assigned when calculating the RMSAEs.

for output is 0.29 over all frequencies, and 0.12 over business cycle frequencies (representing cycles of 6–32 quarters); in contrast, the KPR model gives 0.34 and 0.21, respectively. The RMSAE for investment is 0.38 over all frequencies and 0.05 over business cycle frequencies; in comparison, the KPR model gives 0.36 and 0.10, respectively. The RMSAE for labor is 0.71 over all frequencies and 0.39 over business cycle frequencies; while they are 0.72 and 0.47 for the KPR model. Excepting consumption, the current model performs better than the KPR model around the business cycle frequency.

The extremely smooth path of real wage is responsible for the excessive smoothness of the consumption path in the current model. This has partially to do with the fact that the KPR model is driven by highly persistent technology shocks that have a huge direct impact on the real wage, while the current model is driven by serially uncorrelated belief shocks that have only an indirect impact on the real wage through productive externality. This indicates that in order to better match the data, shocks to fundamentals are necessary.

6. HOW IMPORTANT ARE SUNSPOT SHOCKS?

How important are sunspot shocks to post-war U.S. business cycles relative to fundamental shocks such as technology shock and preference shock? This question is particularly interesting here since the empirical literature on capacity utilization seems to suggest that productivity shocks may become less important in explaining output fluctuations when capacity utilization is explicitly taken into account (Shapiro [46] and Burnside, Eichenbaum, and Rebelo [21]). And Cooley *et al.* [25], on the other

hand, seems to suggest that demand shocks may have played only a small role in explaining cyclical fluctuations of the U.S. economy.

It is generally quite difficult to evaluate the relative contributions of different shocks to business cycles in a theoretical model, since we have few ideas about the true time series properties of those shocks unless they are empirically observable. This section tries to address that issue by allowing three independent shocks into the model: a permanent Hicks-neutral technology shock, a transitory preference shock in a way suggested by Baxter and King [9],²⁰ and an i.i.d sunspot shock. We propose to calibrate the variance of these three shocks in the following way. The innovations in the three shocks are assumed to be independently distributed; the first-order autocorrelation coefficients for the three shocks are assumed to be 1.0, 0.5, and 0.0 for technology, preference, and sunspot, respectively; and the relative standard deviations of the three shocks are then calibrated such that the model generates a correlation between productivity and labor that is zero. The Dunlop–Tarshis empirical observation that the productivity–labor correlation is near zero is an important testimony for economic theories. Christiano and Eichenbaum [23] showed that both demand shocks and supply shocks are needed in an RBC in order to explain this correlation. We therefore choose this correlation as the benchmark for calibrating the relative standard deviations of the three shocks.²¹ In addition, we also impose the constraint that the model-generated spectrum for output growth matches that of the data as closely as possible. More precisely, we calibrate the standard deviation of the three shocks such that they minimize the model’s productivity–labor correlation (in absolute value) and the model’s RMSAE with respect to output growth using the Watson criterion, subject to the constraint that the model’s spectral density is bounded above by that of the data at each frequency. This semi-estimation procedure helps to pin down the standard deviations of the three shocks by forcing the model to mimic both the productivity–labor correlation and the spectral density of output growth in the data. From this calibration, we shall know how good this model is in explaining the total variance distribution of U.S. output growth under multiple shocks. We

²⁰ Namely, the preference over consumption is now specified as $\ln(c - \Delta)$, where Δ is a shock with its mean specified as 10% of the steady state consumption level. A positive increase in Δ therefore signifies an urgent need to consume (e.g., as a result of the rise in demand for the necessities). The preference shock differs from the sunspot shock in at least two aspects. First, as a forecasting error under rational expectations, the sunspot shock must be an i.i.d random variable, while the preference shock can be serially correlated. Second, the sunspot shock affects only the intertemporal consumption Euler equation, while the preference shock affects also the labor-market equilibrium condition. As a result, consumption’s volatility is the same as that of the real wage under sunspot shocks, but is greater than that of the real wage under preference shocks.

²¹ This idea was first proposed by Aiyagari [1].

then analyze the contribution of each individual shock to the U.S. output growth by turning off the other two shocks each time in turn.²²

The results are presented in Fig. 5. Figure 5A (solid line) represents the spectrum of output growth implied by the model when all three shocks are active. The model is seen capable of explaining most of the variations in U.S. output growth (dashed line). The ratio between the areas underneath the two spectral density functions is 0.72, indicating that only 28% of the U.S. output growth is left unexplained by the model (due to mis-specification of the model, to the omission of other important shocks such as monetary shocks, or to sampling errors). The minimized productivity–labor correlation is -0.03, and the RMSAE for all frequencies is 0.06 (indicating that the R^2 of the model could be as high as 0.94 in the sense of Watson [51]). This is a substantial improvement over the previous case when sunspot shock was the only source of uncertainty (where the RMSAE was 0.29).

Table III reports the standard RBC statistics generated by the current model, in order to compare with the previous case (Table I). The improvement can be seen in almost all aspects. For example, the relative volatility of consumption to output increases substantially from 0.04 to 0.49, a 12-fold magnification, and the relative volatility of productivity to output also increases substantially from 0.04 to 0.48. The relative volatility of investment to output decreases from 4.63 to 3.98, a number that is closer to the data. The relative volatility of labor to output now matches that of the data exactly. The correlation between productivity and output also improves from 0.37 to 0.48 (the U.S. data is 0.55).

Figures 5B, C, and D (solid lines) represent the model's spectra of output growth when technology shock, preference shock, and sunspot shock are individually active. Surprisingly, the technology shock contributes virtually nothing to the variance of U.S. output growth (its spectral density can hardly be seen in the graph). The preference shock and the sunspot shock, however, contribute greatly to the variance of U.S. output growth (35 and 37%, respectively). The contribution of preference shock (35%) is mostly to variations around the business cycle frequency while the contribution of sunspot shock (37%) is mostly to high frequency noises in output growth.²³ Therefore, out of the total 72% of U.S. output growth

²² Since the impulse responses of the model to fundamental shocks are indeterminate at the impact period, we arbitrarily fix the initial investment level at its steady state in accordance with the idea of time-to-build. This practice should have little effect on the second moments of the model reported in Fig. 4 and Table III.

²³ The RMSAE under preference shock only is 0.26 and under sunspot shock only is 0.23, indicating that the upper bound of explanatory power is 74% for preference shock and 77% for sunspot shock in the sense of Watson [51].

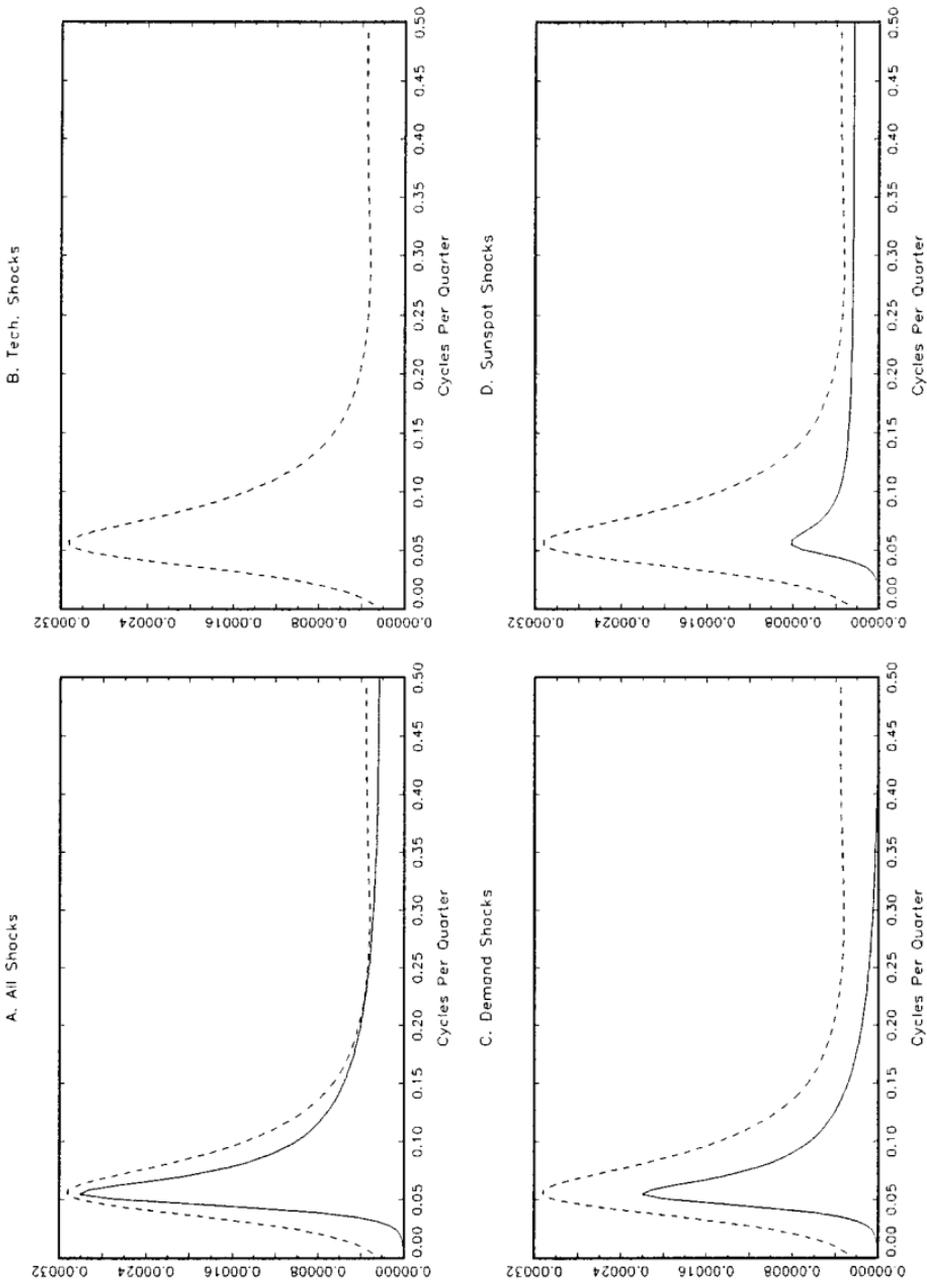


FIG. 5. Spectral decomposition of different shocks' independent contributions to U.S. output growth. In all windows, dashed lines represent data, solid lines represent model.

TABLE III
Sample and Population Moments

Var	U.S. Sample 1948:1-1994:2			Model Population Moments		
	σ_x/σ_y	Cor(xy)	Autocar	σ_x/σ_y	Cor(x, y)	Autocor
y	1.00	1.00	0.95	1.00	1.00	0.90
c	0.74	0.88	0.97	0.49	0.55	0.95
inv	2.63	0.87	0.92	3.98	0.93	0.86
n	0.88	0.62	0.92	0.88	0.86	0.87
u	—	—	—	0.68	0.86	0.86
y/n	0.83	0.55	0.90	0.51	0.48	1.0

Note. Variables (y , c , inv, n , u , y/n) stand for output, consumption, investment, labor, capacity utilization, and productivity, respectively. The U.S. data were detrended by a quadratic time trend.

explained by the model, demand-side shocks (preference shock and sunspot shock) account for nearly all of it, and technology shock accounts for virtually nothing. And among the two most significant demand-side shocks, preference shock is mainly responsible for variations at the business cycle frequency.

7. CONCLUSIONS

This paper has shown that the “elasticity effect” and the “returns-to-scale effect” of capacity utilization can dramatically change the dynamics of a growth model. Under these effects, indeterminacy can easily occur for externalities mild enough so that the aggregate labor demand curve is downward sloping. These effects provide a powerful propagation mechanism of the business cycle and are consistent with the empirical findings that the estimated elasticity of capital is near zero and that of labor is near or greater than one. In particular, this paper has found that business cycles propagated under variable capacity utilization and mild increasing returns-to-scale can better explain the growth dynamics of U.S. aggregates at the business cycle frequency than models with constant returns-to-scale. Moreover, demand-side shocks are found to be far more important than supply-side shocks in explaining U.S. business cycles under the current specification of the model.²⁴

²⁴ An important caveat is that the quantitative results obtained in the paper should not be viewed as robust to different specifications of the model. They serve only as illustrative exercises under the current specification of the model. For example, the reason that technology shock appears unimportant may simply be due to our pre-requirement that the model should deliver a zero productivity–labor correlation.

This paper has studied capacity utilization without explicitly distinguishing its intensive margin and its extensive margin. An extension of the model to allow capacity to vary along both margins (in ways suggested by Bils and Cho [17] and Cooley, Hansen, and Prescott [25]) is expected to be able to reduce further the required degree of externalities for generating indeterminacy and to enrich further the model's dynamics. This is left for future research.

APPENDIX

In equation systems (9)–(12), substituting the first two equations for consumption (c_t) and depreciation rate (δ_t) into the last two equations (the Euler equation and the budget constraint), and then log-linearizing around the steady state gives:

$$(1 + \beta(a^* - 1))\hat{k}_{t+1} + (\beta(b^* - (1 + \gamma)))\hat{n}_{t+1} = a^*\hat{k}_t + (b^* - (1 + \gamma))\hat{n}_t \quad (a)$$

$$\hat{k}_{t+1} = \hat{k}_t + \delta \left(\frac{\theta}{\alpha} - 1 \right) (1 + \gamma)\hat{n}_t$$

where

$$a^* \equiv \alpha(1 + \eta)\tau_k, \quad (b)$$

$$b^* \equiv (1 - \alpha)(1 + \eta)\tau_n.$$

Equation system (a) can be rewritten in matrix form:

$$M_1 \begin{pmatrix} \hat{k}_{t+1} \\ \hat{n}_{t+1} \end{pmatrix} = M_2 \begin{pmatrix} \hat{k}_t \\ \hat{n}_t \end{pmatrix}, \quad (c)$$

where M_1 and M_2 are 2x2 real matrices given by:

$$M_1 \equiv \begin{pmatrix} 1 + \beta(a^* - 1) & \beta b^* - (1 + \gamma) \\ 1 & 0 \end{pmatrix}, \quad (d)$$

$$M_2 \equiv \begin{pmatrix} a^* & b^* - (1 + \gamma) \\ 1 & ((\theta/\alpha) - 1)(1 + \gamma)\delta \end{pmatrix}.$$

Since labor can be expressed as a linear function of capital and consumption using the first equation in (9)–(12), Eq. (c) can be written as:

$$M_1 \begin{pmatrix} 1 & 0 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} \hat{k}_{t+1} \\ \hat{c}_{t+1} \end{pmatrix} = M_2 \begin{pmatrix} 1 & 0 \\ \alpha_1\alpha_2 & \end{pmatrix} \begin{pmatrix} \hat{k}_t \\ \hat{c}_t \end{pmatrix}, \quad (e)$$

which does not change the determinate and the trace of the Jacobian of the system (c). It is easy to show that the Jacobian is

$$\begin{pmatrix} 1 & (\theta - \alpha)(1 + \gamma) \frac{\delta}{\alpha} \\ \frac{(1 - \beta)(1 - a^*)}{1 + \gamma - \beta b^*} & \frac{1 + \gamma - b^* + (1 + \beta(a^* - 1))(\theta - \alpha)(1 + \gamma) \delta/\alpha}{1 + \gamma - \beta b^*} \end{pmatrix}. \quad (f)$$

Hence, the determinate and trace of B in (14) are given by:

$$\det(B) = \frac{\det(M_2)}{\det(M_1)} = \frac{(1 + \gamma) - b^* + a^*(\theta - \alpha)(1 + \gamma) \frac{\delta}{\alpha}}{(1 + \gamma) - \beta b^*}, \quad (g)$$

$$\text{tr}(B) = 1 + \det(B) + \frac{(1 - \beta)(1 - a^*)(\theta - \alpha)(1 + \gamma) \frac{\delta}{\alpha}}{(1 + \gamma) - \beta b^*}. \quad (h)$$

Simplifying and rearranging terms then gives (18) and (19) in the text.

ACKNOWLEDGMENTS

I thank Rao Aiyagari, David Cook, Roger Farmer, Takashi Kamihigashi, Danyang Xie, and especially Jess Benhabib, Jang-Ting Guo, and an anonymous referee for discussions and very helpful comments. I also thank Camilla Erskine and Virginia Anne Unkefer for editorial assistance.

REFERENCES

1. R. Aiyagari, On the contribution of technology shocks to business cycles, *Fed. Res. Bank Minn. Quart. Rev.*, Winter (1994), 22–34.
2. R. Aiyagari, The econometrics of indeterminacy: A comment, *Carnegie-Rochester Conf. Ser. Public Policy* **43** (1995), 273–284.
3. C. Azariadis, Self-fulfilling prophecies, *J. Econ. Theory* **25** (1981), 380–396.
4. E. Bartlesman, R. Caballero, and R. Lyons, Customer- and supplier-driven externalities, *Amer. Econ. Rev.* **84** (1994), 1075–1084.
5. S. Basu, Procyclical productivity: increasing returns or cyclical utilization?, *Quar. J. Econ.* **111** (1996), 719–951.
6. S. Basu and J. Fernald, Are apparent productive spillovers a figment of specification error?, *J. Monet. Econ.* **36** (1995), 165–188.
7. S. Basu and J. Fernald, Returns to scale in U.S. production: Estimates and implications, *J. Political Economy* **105** (1997), 249–283.
8. S. Basu and M. Kimball, “Cyclical Productivity with Unobserved Input variation,” Mimeo, Univ. of Michigan (1995).

9. M. Baxter and R. King, "Productive Externality and Cyclical Volatility," Working paper 245, Univ. of Rochester (1990).
10. J. Beaulieu and M. Shapiro, "Capital Utilization, Factor Productivity, and Returns to Scale: Evidence from Plant-level Data," Manuscript, Board of Governors of the Federal Reserve System, December, 1995.
11. J. Benhabib and R. Farmer, Indeterminacy and increasing returns, *J. Econ. Theory* **63** (1994), 19–41.
12. J. Benhabib and R. Farmer, Indeterminacy and sector-specific externalities, *J. Monet. Econ.* **37** (1996), 421–443.
13. J. Benhabib and R. Farmer, "Indeterminacy and Sunspots in Macroeconomics," Manuscript, New York University (1997).
14. J. Benhabib and K. Nishimura, Indeterminacy and sunspots with constant returns, *J. Econ. Theory* **81** (1998), 58–96.
15. J. Benhabib and A. Rustichini, Introduction to the symposium on growth, fluctuations, and sunspots: Confronting the data, *J. Econ. Theory* **63** (1994), 1–18.
16. B. Bernanke and M. Parkinson, Procyclical productivity and competing theories of the business cycle: Some evidence from interwar U.S. manufacturing industries, *J. Polit. Econ.* **99** (1991), 439–459.
17. M. Bils and J.-O. Cho, Cyclical factor utilization, *J. Monet. Econ.* **33** (1994), 319–354.
18. M. Boldrin and A. Rustichini, Indeterminacy of equilibria in models with infinitely lived agents and external effects, *Econometrica* **62** (1994), 323–342.
19. C. Burnside, Production function regressions, returns to scale, and externalities, *J. Monet. Econ.* **37** (1996), 177–201.
20. C. Burnside and M. Eichenbaum, Factor hoarding and the propagation of business cycle shocks, *Amer. Econ. Rev.* **86** (1996), 1154–1174.
21. C. Burnside, M. Eichenbaum, and S. Rebelo, "Capital Utilization and Returns to Scale," Working Paper 402, Rochester Center for Economic Research, 1995.
22. D. Cass and K. Shell, Do sunspots matter? *J. Polit. Econ.* **91** (1983), 193–227.
23. L. Christiano and M. Eichenbaum, Current real-business-cycle theories and aggregate labor-market fluctuations, *Amer. Econ. Rev.* **82** (1992), 430–450.
24. L. Christiano and S. Harrison, "Chaos, Sunspots, and Automatic Stabilizers," NBER Working Paper 5703 (1997).
25. T. Cooley, G. Hansen, and E. Prescott, Equilibrium business cycles with idle resources and variable capacity utilization, *Econ. Theory* **6** (1995), 35–49.
26. D. DeJong, B. Ingram, Y. Wen, and C. Whiteman, "Cyclical Implications of the Variable Utilization of Physical and Human Capital," Manuscript, The University of Iowa (1996).
27. R. Farmer, "The Macroeconomics of Self-Fulfilling Prophecies," MIT Press, Cambridge, MA, (1993).
28. R. Farmer and J. T. Guo, Real business cycles and the animal spirits hypothesis, *J. Econ. Theory* **63** (1994), 42–73.
29. M. Finn, "Energy Price Shocks and Variance Properties of Solow's Productivity Residual," Manuscript, Federal Reserve Bank of Richmond (1992).
30. J. Gali, Monopolistic competition, business cycles, and the composition of aggregate demand, *J. Econ. Theory* **63** (1994), 73–96.
31. J. Greenwood, Z. Hercowitz, and G. Huffman, Investment, capacity utilization, and the real business cycle, *Amer. Econ. Rev.* **78** (1988), 402–417.
32. J. Greenwood, Z. Hercowitz, and P. Krusell, "Macroeconomic Implications of Investment-Specific Technological Change, Manuscript," University of Rochester, (1992).
33. G. Hansen, Indivisible labor and the business cycle, *J. Monet. Econ.* **16** (1985), 309–325.
34. G. Hansen and R. Wright, The labor market in real business cycle theory, *Fed. Res. Bank Minn. Quart. Rev.* Spring (1992), 2–12.

35. R. King, C. Plosser, and S. Rebelo, Production, growth and business cycles: I. The basic neoclassical model, *J. Monet. Econ.* **21** (1988), 195–232.
36. F. Kydland and E. Prescott, The workweek of capital and its cyclical implications, *J. Monet. Econ.* **21** (1988), 343–360.
37. F. Kydland and E. Prescott, Hours and employment variation in business cycle theory, *Econ. Theory* **1** (1991), 63–81.
38. R. E. Lucas, Capacity, overtime, and empirical production functions, *Amer. Econ. Rev.* **60** (1970), 23–27.
39. R. Marris, “The Economics of Capital Utilisation,” Cambridge University Press, Cambridge, UK, 1964.
40. S. Norrbin, The relation between prices and marginal cost in US industry: A contradiction, *J. Polit. Econ.* **101** (1993), 1149–1164.
41. R. Perli, Indeterminacy, home production, and the business cycle: A calibrated analysis, *J. Monet. Econ.*, forthcoming.
42. R. Perli, Home production and persistence of business cycles, *J. Econ. Dynam. Control*, forthcoming.
43. G. Perry, Capacity in manufacturing, *Brookings Pap. Econ. Act.* **3** (1973), 701–742.
44. J. Rotemberg and M. Woodford, Dynamic general equilibrium models with imperfectly competitive product markets, in “Frontiers in Business Cycle Research” (T. Cooley, Ed.), Princeton Univ. Press, Princeton, NJ, 1995.
45. S. Schmitt-Grohe, Comparing four models of aggregate fluctuations due to self-fulfilling expectations, *J. Econ. Theory*, forthcoming.
46. M. Shapiro, Cyclical productivity and the workweek of capital, *Amer. Econ. Rev.* **83** (1993), 229–233.
47. M. Shapiro, Macroeconomic implications of variation in the workweek of capital, *Brookings Pap. Econ. Act.* **2** (1996), 79–133.
48. K. Shell, “Monnaie et Allocation Intertemporelle,” Mimeo, Seminaire d’Econometrie Roy-Malinvaud, Centre National de la Recherche Scientifique, Paris, November 1977).
49. R. Solow, “Draft of Presidential Address: On the Short-Run Relation of Employment and Output,” manuscript, MIT 1964.
50. P. Taubman and M. Wilkinson, User cost, capital utilization and investment theory, *Int. Econ. Rev.* **11** (1970), 209–215.
51. M. Watson, Measures of fit for calibrated models, *J. Polit. Econ.* **101** (1993), 1011–1041.
52. Y. Wen, “Can a Real Business Cycle Model Pass the Watson Test?” *J. Monet. Econ.*, forthcoming.
53. Y. Wen, “Capacity Utilization under Increasing Returns to Scale,” Manuscript, The Hong Kong University of Science & Technology, 1996.
54. M. Woodford, Stationary sunspot equilibria in a finance constrained economy, *J. Econ. Theory* **40** (1986), 128–137.
55. M. Woodford, Expectations, finance and aggregate instability, in “Finance Constraints, Expectations and Macroeconomics” (M. Kohn and S. C. Tsing, Eds.), Oxford Univ. Press, London, 1988.
56. M. Woodford, Self-fulfilling expectations and fluctuations in aggregate demand, in “New Keynesian Economics” (G. Mankiw and D. Romer, Eds.), Vol. 2, MIT Press, Cambridge, MA.