First-order approximation methods are a standard technique for analyzing the local dynamics of dynamic stochastic general equilibrium (DSGE) models. Although linear methods yield quite accurate solutions for a broad class of DSGE models, some important economic issues (e.g., portfolio choice and welfare) cannot be adequately addressed by first-order methods. This paper provides yet another case when first-order methods may be inadequate for capturing the business cycle properties of a DSGE model. In particular, the authors show that increasing returns to scale (due to production externalities) may induce asymmetric business cycles and nonlinear income effects that are not fully captured by linear approximation methods. For example, hump-shaped output dynamics can emerge even when externalities are below the threshold level required for indeterminacy, and output expansion tends to be smoother and longer, whereas contraction tends to be deeper but shorter-lived, as observed in the U.S. economy. (JEL C63, E0, E32)

The standard approach to studying the business cycle implications of dynamic stochastic general equilibrium (DSGE) models is to focus on the models’ local dynamics near a steady state through linear (first-order) approximations (such as the log-linearization method of King, Plosser, and Rebelo, 1988). It is well known that for standard real business cycle (RBC) models with constant returns, first-order approximation methods provide quite accurate solutions and higher-order methods give almost identical predictions.

The central twist of this paper is the addition of increasing returns to scale (IRS) caused by production externalities. We show that this simple deviation from standard RBC models generates nontrivial nonlinearities that are not well captured by first-order methods. Importantly, these nonlinearities are increasing in the degree of externalities over parameter ranges that predict a unique bounded rational expectations equilibrium. The model does not rely on local indeterminacy to generate new and interesting dynamics, although one contribution of the paper is the documentation of model properties when the model gives rise to local indeterminacy.1

Specific results of interest are that technology shocks generate asymmetric effects on business cycles. These effects come from second-order components of the model, which are ignored by linear approximation methods. Conditional on a positive technology shock, hump-shaped impulse response functions are predicted for employment and output, consistent with much empirical evidence.

1 “Local indeterminacy” means that there are multiple rational expectations equilibrium paths that converge to the same steady state. Note that local indeterminacy is not the same thing as multiple steady-state equilibria. Indeterminacy can arise in a model with a unique steady state.
work. Smooth, prolonged dynamics are observed. In contrast, conditional on a negative technology shock, the model predicts sharp, less persistent dynamics. These predictions combined provide an explanation of the strong asymmetry of the business cycle in the U.S. economy.

The existing literature shows IRS are an important source of dynamics not only for endogenous growth (e.g., Romer, 1986), but also for the business cycle. Baxter and King (1991) show that incorporating production externalities into a standard RBC model generates a better overall fit of the model to U.S. data, especially under aggregate demand shocks. Benhabib and Farmer (1994), Farmer and Guo (1994), Wen (1998a), and Benhabib and Wen (2004) show that IRS can generate endogenous business cycles if externalities are large enough to make the model’s steady state locally indeterminate.\(^3\)

However, this segment of the literature has relied on first-order approximation methods to investigate the models’ dynamic behaviors. Although it has been shown that first-order approximation usually yields quite accurate results for standard RBC models (such as those of Kydland and Prescott, 1982, and King, Plosser, and Rebelo, 1988), it is less clear whether such methods retain their accuracy in describing a model’s local dynamics when market failures and non-convexities, such as externalities and IRS, are involved.

IRS can greatly amplify the impact of shocks (Baxter and King, 1991) and dramatically change a model’s topology around its steady state so that complicated dynamics, such as bifurcations, discontinuous jumps, and complex eigenvalues, may emerge (Benhabib and Farmer, 1994, and Coury and Wen, 2009). Therefore, it is interesting to investigate whether first-order approximation methods continue to yield accurate predictions of a model’s dynamics when production externalities are present.

In addition, one of the most important aspects of stochastic dynamic models—risk—cannot be captured by linear solution methods. For example, optimal portfolio decisions cannot be analyzed by linear approximation methods. For reasons such as this, second-order solution methods have been proposed and developed in the recent literature.\(^4\)

In this paper, we apply the second-order approximation method developed in the existing literature to analyze the local dynamics of an RBC model with externalities. We show that allowing for second-order terms not only improves the accuracy of approximations, but also changes the predicted local dynamics of the model dramatically when IRS exist. In particular, the magnitude of the impulse responses to positive technology shocks is significantly smaller and smoother under the second-order approximation method than the first-order method, and hump-shaped impulse responses can emerge even with degrees of externalities that are too small to trigger indeterminacy. This finding is in sharp contrast to the results of Benhabib and Wen (2004), who find that hump-shaped dynamics emerge under the first-order approximation method only when the degree of externalities is large enough that the model becomes locally indeterminate.

The new findings on the dynamic effects of externalities are driven by the fact that externalities introduce a large nonlinear income effect, which is captured by a large negative coefficient in front of the squared term for technology shocks and a large positive coefficient in front of the cross term for capital and technology shocks. This implies that technology shocks have the following asymmetric second-order effects on aggregate output dynamics: A positive technology shock generates smooth and hump-shaped output responses because the variance of technology shocks neutralizes the shock on impact, but the covariance of capital and technology shocks enhances (and thus propagates) the effects of the positive shock in the subsequent periods. Conversely, a negative technology shock generates much larger and sharper declines in output.
because the variance of technology shocks greatly amplifies the negative shock, while the covariance term reduces the shock’s negative impact in the subsequent periods; this causes the impulse responses of output to monotonically increase toward the steady state from below. Such an asymmetric nonlinear effect is similar to that observed in the U.S. economy: On the one hand, economic expansion tends to be smooth and gradual while economic contraction tends to be sharp and short-lived (see, e.g., Nefctçi, 1984; Sichel, 1993; Kim and Piger, 2002; McKay and Reis, 2008; and Morley and Piger, forthcoming). On the other hand, unexpected increases in oil prices (equivalent to negative total factor productivity shocks) tend to have a large adverse impact on aggregate output, whereas decreases in oil prices tend to have only a small or negligible impact on output (e.g., see Hamilton, 2003, and Mork, 1989).

Our analysis also helps to explain a puzzle in the indeterminacy literature. For example, in the Benhabib-Farmer (1994) model, positive technology shocks are extremely expansionary before indeterminacy arises but suddenly become excessively contractionary when the model becomes locally indeterminate, thereby generating sharp falls in output, investment, and hours. Using a second-order method, we find that (i) positive technology shocks have a second-order negative impact on hours worked and (ii) this nonlinear income effect increases with the degree of IRS. Hence, as externalities increase, the second-order income effect gradually dominates the first-order substitution effect, and the initial responses of hours gradually become negative around the point of indeterminacy. Once the model becomes indeterminate, the coefficients of the second-order terms for labor become zero, so the strong income effects are captured instead by the first-order terms.5

An important caveat is that this paper’s main point is not to find frictions to improve the empirical fit of existing RBC models with the business cycle. Hence, moment matching is not our goal. Rather, we try to provide a new scenario that linear solution methods may yield inaccurate descriptions of the model’s dynamics. Moreover, given that the literature has shown that IRS may be important for understanding the business cycle based on first-order approximations, we are also interested in understanding whether IRS can introduce nonlinear dynamics not captured by first-order methods.

The literature has shown that linear approximation methods may result in quantitatively significant biases even for standard RBC models without externalities. For example, Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) study the econometrics of computed dynamic models and the consequences for inference of the use of approximated likelihoods. They find that second-order approximation errors in the policy function, which are completely ignored by first-order approximation methods, have first-order effects on the likelihood function.6 The problem of numerical simulations of dynamic economies with heterogeneous agents and economic distortions is also studied in the existing literature. For example, Feng et al. (2009) and Peralta-Alva and Santos (2010), among others, discuss problems related to the existence and computation of Markovian equilibria in economies with heterogeneous agents and market distortions, as well as convergence and accuracy properties of numerical solutions.7

The remainder of the paper is organized as follows: The next section presents the model with externalities and introduces the second-order perturbation method. A simple example is provided to illustrate the difference between first-order and second-order methods. We then examine the model’s dynamics with and without indeterminacy. Next, we compare the accuracy of linear and second-order approximations in the presence of IRS and then offer our conclusions.

5 See discussions in the subsection entitled “Second-Order Effects of Sunspots.”

6 However, Ackerberg, Geweke, and Hahn (2009) show by counterexample that this conclusion by Fernández-Villaverde, Rubio-Ramírez, and Santos (2006) is false and argue that second-order approximation errors in the policy function have at most second-order effects on parameter estimates.

7 More references to this literature can also be found in these two articles.
THE MODEL

Our model is similar to those of Baxter and King (1991) and Benhabib and Farmer (1994). There exists a continuum of identical agents in the unit interval [0,1]. A typical or representative agent chooses consumption \(c_t\), hours worked \(n_t\), and capital stock \(k_{t+1}\) to solve

\[
\max_{\{c_t, n_t, k_{t+1}\}} E_t \sum_{t=0}^{\infty} \beta^t \left( \frac{c_t^{1-\tau}}{1 - \tau} - a n_t^{1+\gamma} \right),
\]

subject to the resource constraint,

\[
c_t + k_{t+1} - (1 - \delta) k_t \leq A_t \left( F_t^{\alpha} \pi_t^{1-\alpha} \right) \eta k_t^{\alpha} n_t^{1-\alpha},
\]

where \(\{\tilde{k}_t, \tilde{n}_t\}\) denote, respectively, the average economy-wide capital stock and hours that are taken as given by individuals and \(A_t\) denotes aggregate technology shocks. The model exhibits IRS at the social level if the externality parameter \(\eta > 0\).

The equilibrium of the model is determined by the following necessary conditions:

1. \(an_t^\eta = c_t^{1-\tau} (1 - \alpha) A_t k_t^{\alpha[1+\eta]} n_t^{1-\alpha[1+\eta]}\)
2. \(c_t^{1-\tau} = \beta E_t \left\{ c_{t+1} \left( a A_{t+1} k_{t+1}^{\alpha[1+\eta]} n_{t+1}^{1-\alpha[1+\eta]} + 1 - \delta \right) \right\}\)
3. \(c_t + k_{t+1} - (1 - \delta) k_t = A_t k_t^{\alpha[1+\eta]} n_t^{1-\alpha[1+\eta]}\)
4. \(\log A_t = \rho \log A_{t-1} + \sigma \epsilon_t\)

where equation (1) is the optimal labor supply condition, equation (2) equates the marginal cost of reducing consumption in period \(t\) and the marginal gain of consumption in the next period, equation (3) is the aggregate resource constraint in equilibrium, and equation (4) specifies the dynamics of \(A_t\). The parameter \(\sigma\) controls the variance of the innovation \(\epsilon_t \sim N(0,1)\) and measures the level of uncertainty in the economy. Since agents are identical, in general equilibrium the individual variables are equal to their aggregate counterparts (e.g., \(k_t = \tilde{k}_t\) and \(n_t = \tilde{n}_t\)).

As shown by Benhabib and Farmer (1994), this model has a unique steady state that is a saddle if the degree of externalities is small and a sink if \(\eta\) is large enough. However, Coury and Wen (2009) show that this class of models may have multiple dynamic equilibria (such as stable \(n\)-period cycles) away from the steady state even if the steady state appears to be a saddle. Mindful of this fact, in the following analyses we choose the value of \(\eta\) sufficiently below the critical values found by Coury and Wen (2009) for \(n\)-period cycles.

Second-Order Taylor Expansion

The model’s equilibrium is solved by the second-order approximation method. The variables in the above equations can be grouped into two types: the state variables and the control variables. The state variables include the capital stock and the realized exogenous shock at the beginning of each period, denoted by the vector \(s_t = (k_t, A_t)\). The control variables include consumption and hours, denoted by \(z_t = (n_t, c_t)\).

Because of certainty equivalence, the equilibrium paths of economic variables determined by linear approximation methods are independent of the degree of uncertainty (\(\sigma\)) and the dynamic impulse responses of the model are symmetric with respect to the sign of the shocks (\(A_t\)). That is, the policy functions can be written as

\[
z_t = g(s_t),\]

and the state variables follow the law of motion

\[
s_{t+1} = h(s_t) + \zeta \sigma \epsilon_{t+1},\]

where the 2 \(\times\) 1 vector \(\zeta = [01]^\top\) since technological innovations do not directly affect the next-period capital stock.

---

8 We also studied the model of Wen (1998a) in which the degree of externalities required for indeterminacy is much smaller because of variable capacity utilization, and our conclusions remain robust. To simplify the analysis, we choose the simpler model of Benhabib and Farmer (1994) with fixed capacity utilization even though this model requires an implausibly large degree of IRS to generate a significant difference in local dynamics from those of standard RBC models.

9 Benhabib and Farmer (1994) prove that this representative-agent model with externalities is equivalent to a decentralized market economy with increasing returns at the firm level.

10 If indeterminacy occurs, period-\(t\) consumption can be treated as a state variable since the steady state is a sink.
However, under second-order approximations, second-order terms such as
\[ E_t \hat{A}_{t+1}^2 = \frac{\sigma^2}{1-\rho^2} \text{ and } E_t \hat{A}_{t+1} \hat{K}_{t+1} \]
emerge, so the policy functions \( g(\cdot) \) and equilibrium paths of state variables \( h(\cdot) \) are dependent on \( \sigma \). Accordingly, the general policy functions are characterized by
\[ z_t = g(s_t, \sigma) \]
\[ s_{t+1} = h(s_t, \sigma) + \zeta^\sigma \sigma_{t+1}. \]

Note that when the standard deviation (SD) of the shock \( \sigma = 0 \), there is no uncertainty in the model and all variables remain in the non-stochastic steady state. Therefore, the impulse responses of the system to a technology shock obtained by first-order and second-order methods converge to each other as \( \sigma \) approaches zero.

Define \( \hat{z}_t = \log z_t - \log z_0 \), \( \hat{s}_t = \log s_t - \log s_0 \). Following the literature (see, e.g., Schmitt-Grohé and Uribe, 2004), the policy functions \( g(s, \sigma) \) can be approximated by a second-order Taylor expansion,
\[ \hat{z}_t = \begin{bmatrix} g_s & g_\sigma \end{bmatrix} \sigma \hat{s}_t + \begin{bmatrix} \hat{s}_t \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \hat{s}_t \sigma \\ \sigma \end{bmatrix} \begin{bmatrix} g_{ss} & g_{s\sigma} \\ g_{s\sigma} & g_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \hat{s}_t \sigma \\ \sigma \end{bmatrix}, \]
\[ \hat{s}_{t+1} = \begin{bmatrix} h_s & h_\sigma \end{bmatrix} \sigma \hat{s}_t + \begin{bmatrix} \hat{s}_t \sigma \\ \sigma \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \hat{s}_t \sigma \\ \sigma \end{bmatrix} \begin{bmatrix} h_{ss} & h_{s\sigma} \\ h_{s\sigma} & h_{\sigma\sigma} \end{bmatrix} \begin{bmatrix} \hat{s}_t \sigma \\ \sigma \end{bmatrix}, \]
where the first derivatives \( g_s \) and \( g_\sigma \) are known from the solution to the first-order system. By taking second-order derivatives of the first-order conditions with respect to \( s \) and \( \sigma \), we can obtain a linear equation system with the unknown elements in cross-derivative coefficients \( g_{ss\sigma}, h_{ss\sigma}, g_{s\sigma\sigma}, \) and \( h_{\sigma\sigma} \). The cross terms are symmetric and equal to zero, \( g_{ss\sigma} = g_{s\sigma\sigma}, h_{ss\sigma} = h_{s\sigma\sigma} = 0 \), as shown by the literature (e.g., see Schmitt-Grohé and Uribe, 2004). The second-order coefficient matrices are three dimensional. For example, since we have two elements in \( g(s, \sigma) \) and two elements in \( s, \sigma \), \( g_{ss} \) is thus a \( 2 \times 2 \times 2 \) matrix.

The steady state under second-order methods differs from that under linear methods. For example, if the initial state is zero,
\[ \hat{s}_0 = 0, \text{ then } \hat{s}_1 = \frac{1}{2} \sigma^2; \]
consequently, the system will evolve and rest at another steady state. Moreover, at the non-stochastic steady state where \( \hat{z}_t \) and \( \hat{s}_t \) are zero, the policy functions differ from those under linear solutions by a constant term proportional to \( \sigma^2 \).

Accordingly, to generate impulse response functions and time-series comparable to those under linear methods, we can find the stochastic steady state by numerical simulation and then introduce shocks to generate time-series or impulse responses relative to the stochastic steady-state values.\(^{11}\) Furthermore, we should note that direct use of equations (9) and (10) can generate an exploding path since second-order terms result in unnecessary higher-order terms in consecutive iteration. The pruning process proposed by Kim et al. (2008) is useful for overcoming this problem. In fact, this process uses only first-order parts of the response to generate the second-order terms in the recursive computation.

**A Simple Example**

Suppose an economy is described by the following nonlinear equation:
\[ y_t = E_t A_{t+1} y_{t+1}, \quad |\theta| < 1; \]
where \( y_t \) is an endogenous jump variable and \( A_{t+1} \) an exogenous driving process satisfying
\[ \log A_t = \rho \log A_{t-1} + \sigma \varepsilon_t \]
and \( \varepsilon_t \sim N(0,1) \), where \( \sigma \) denotes the SD of \( \varepsilon_t \). The model has two steady states: \( \bar{y} = 0 \) and \( \bar{y} = 1 \). We

---

\(^{11}\) We depart from Schmitt-Grohé and Uribe (2004) and Kim et al. (2008) by adjusting the steady state so that the impulse responses return to zero in the long run.
consider the second steady state because it is “saddle stable.” To facilitate Taylor expansions around a steady state, denote \( \hat{y}_t = \log \hat{y}_t - \log \bar{y} \) and rewrite equation (11) as

\[
(13) \quad e^{\log y_t} = E_t e^{\log \hat{A}_{t+1} + \theta \log y_{t+1}}.
\]

First-Order Method. The first-order Taylor expansion of equation (13) around the steady state \( \log \bar{y} = 0, \log \bar{A} = 0 \) is given by

\[
(14) \quad \hat{y}_t = E_t \left[ \hat{A}_{t+1} + \theta \hat{y}_{t+1} \right].
\]

Using the method of undetermined coefficients, we can guess the solution \( \hat{y}_t = \gamma_A \hat{A}_t + \gamma_\sigma \sigma \) and substitute the solution into equation (14) to obtain

\[
(15) \quad \gamma_A \hat{A}_t + \gamma_\sigma \sigma = E_t \left[ \hat{A}_{t+1} + \theta \left( \gamma_A \hat{A}_{t+1} + \gamma_\sigma \sigma \right) \right].
\]

Applying the law of motion (equation 12) and comparing the coefficients of \( \{\hat{A}_t, \sigma\} \) on both sides of equation (15) gives

\[
\gamma_A = \frac{\rho}{1 - \theta \rho} \quad \text{and} \quad \gamma_\sigma = 0.
\]

Therefore, the first-order accurate solution is given by

\[
(16) \quad \hat{y}_t = \frac{\rho}{1 - \theta \rho} \hat{A}_t.
\]

The impulse response function of \( \hat{y}_t \) to a 1-SD shock in \( \varepsilon_t \) can be generated from the state-space representation of the model,

\[
(17) \quad \begin{pmatrix} \hat{y}_t \\ \hat{A}_t \end{pmatrix} = \begin{pmatrix} 0 & \rho \gamma_A \\ 0 & \rho \end{pmatrix} \begin{pmatrix} \hat{y}_{t-1} \\ \hat{A}_{t-1} \end{pmatrix} + \begin{pmatrix} \gamma_A \\ 1 \end{pmatrix} \sigma \varepsilon_t,
\]

by setting \( \hat{y}_0 = 0, \hat{A}_0 = 0, \sigma = 1, \varepsilon_1 = 1, \) and \( \varepsilon_t = 0 \) for all \( t > 1 \).

Second-Order Method. The second-order Taylor expansion of equation (13) around the steady state \( \log \bar{y} = 0, \log \bar{A} = 0 \) is given by

\[
(18) \quad \hat{y}_t + \frac{1}{2} \hat{y}_t^2 = E_t \left[ \hat{A}_{t+1} + \theta \hat{y}_{t+1} + \frac{1}{2} \left( \hat{A}_{t+1} + \theta \hat{y}_{t+1} \right)^2 \right].
\]

We can guess a second-order solution with undetermined coefficients:

\[
(19) \quad \hat{y}_t = \gamma_A \hat{A}_t + \gamma_{AA} \hat{A}_t^2 + \gamma_\sigma \sigma + \gamma_{o\sigma} \sigma^2.
\]

Notice that using this second-order solution as a substitute in equation (18) would generate higher-order terms such as \( \{\hat{A}_t^x, \hat{y}_t^x, \sigma^x\} \) with \( x \geq 3 \). Since these higher-order terms are irrelevant for our second-order solution, they can be ignored. That is, we can substitute the second-order terms in equation (18) with the first-order solution in equation (16) while keeping the first-order terms as they are in equation (18). Substituting the linear solution (equation (16)) into the second-order terms in equation (18) yields

\[
(20) \quad \hat{y}_t = E_t \left[ \hat{A}_{t+1} + \theta \hat{y}_{t+1} + \frac{1}{2} \left( \hat{A}_{t+1} + \frac{\theta \rho}{1 - \theta \rho} \hat{A}_{t+1} \right)^2 \right] - \frac{1}{2} \left( \frac{\rho}{1 - \theta \rho} \hat{A}_1 \right)^2.
\]

Since \( E_t \hat{A}_{t+1}^2 = \rho^2 \hat{A}_t^2 + \sigma^2 \), the above equation reduces to

\[
(21) \quad \hat{y}_t = \theta E_t \hat{y}_{t+1} + \rho \hat{A}_t + \frac{1}{2(1 - \theta \rho)^2} \sigma^2.
\]

Substituting the conjectured second-order solution (19) into equation (21) and simplifying gives

\[
(22) \quad \rho (1 + \theta \gamma_A) \hat{A}_t + \theta \gamma_\sigma \sigma + \theta^2 \gamma_{AA} \hat{A}_t^2 + \left( \theta \gamma_{AA} + \theta \gamma_{o\sigma} + \frac{1}{2(1 - \theta \rho)^2} \right) \sigma^2.
\]

Comparing the coefficients on both sides gives

\[
\gamma_A = \frac{\rho}{1 - \theta \rho}, \quad \gamma_\sigma = 0, \quad \gamma_{AA} = 0, \quad \text{and}
\]

\[
\gamma_{o\sigma} = \frac{1}{2(1 - \theta)(1 - \theta \rho)^2}.
\]

Hence, the second-order rational expectations equilibrium is given by

\[
(23) \quad \hat{y}_t = \frac{\rho}{1 - \theta \rho} \hat{A}_t + \frac{1}{2(1 - \theta)(1 - \theta \rho)^2} \sigma^2.
\]
Note that the second-order solution conforms to the general solution in equation (9). However, in the special model the second-order effect from $\hat{A}_t^2$ does not exist and the solution differs from the first-order solution only by a constant proportional to $\sigma^2$. Thus, the corresponding impulse response function is also the same as that of the first-order solution (up to a constant term in the steady state).

**Calibration and Eigenvalues**

We calibrate the model based on the existing literature. In particular, we set the following: the discount factor $\beta = 0.99$, the elasticity of intertemporal substitution $\tau = 1$, the inverse elasticity of labor supply $\gamma = 0$ (indivisible labor), capital’s share $\alpha = 0.3$, and the persistence of shock $\rho = 0.9$.

The eigenvalues of the model are given by the linear terms and thus are not affected by higher-order terms. Hence, the region of local indeterminacy is not influenced by the variance of technology shocks under second-order expansion. Excluding the exogenous driving process, the system has two eigenvalues. When externalities are small, only one of the eigenvalues lies inside the unit circle; thus, the steady state is a saddle. When externalities are large enough that $\eta > \eta^* = 0.4935$, both eigenvalues lie inside the unit circle and the system is indeterminate (Benhabib and Farmer, 1994). Figure 1 plots the two eigenvalues as functions of the externality parameter $\eta$. The critical value for indeterminacy is $\eta^* = 0.4935$.\(^{12}\)

The dashed lines in Figure 1 correspond to the region of complex eigenvalues. The figure clearly shows that the eigenvalues undergo dramatic changes as the degree of externalities ($\eta$) increases. The explosive root jumps from positive infinity to negative infinity, while the stable root goes through similar topological changes near the critical value $\eta^*$. Both eigenvalues become stable and form a complex conjugate pair (dashed lines) for the region $\eta > \eta^*$. Such dramatic changes in eigenvalues indicate that (i) the topology of the model near the steady state is significantly altered by externalities and (ii) linear approximation may not be accurate enough to capture the curvature of the equilibrium path near the steady state.

\(^{12}\) Wen (1998a) showed that introducing capacity utilization can reduce the critical value $\eta^*$ to about 0.1, which is more consistent with empirical estimates. Our results apply to the model of Wen (1998a), but we choose the simpler Benhabib-Farmer (1994) model for exposition purposes in this paper.


**DYNAMIC ANALYSIS**

**Impulse Responses**

The impulse responses of the economy to a technology shock with SD $\sigma = 0.3$ are graphed in Figure 2.\(^{13}\) For comparison, the left panels show the impulse responses under the first-order approximation method and the right panels show them under the second-order method. The top panels show the results when externalities are absent ($\eta = 0$), and the bottom panels show them when externalities are large enough ($\eta = 0.4$) but well below the critical value of 0.4935 needed for indeterminacy.\(^{14}\)

It is clear from the top panels in Figure 2 that the first-order and second-order methods yield very similar results when externalities are not present, confirming the literature’s findings that linear solution methods provide reasonably good approximations for standard RBC models. For example, the initial impulse response of output is about 0.7 under first-order approximation and about 0.68 under second-order approximation.

However, when externalities exist, the two solution methods yield dramatically different results. The bottom panels in Figure 2 indicate that the initial impulse response of output is about 1.6 under first-order approximation but only about 0.75 under second-order approximation. More importantly, while the responses of output and labor remain monotonic under the first-order method, they all become hump-shaped under the second-order method, suggesting a much richer internal propagation mechanism.

When the level of externalities increases, the hump-shaped responses become even more prominent. For $\eta$ close to the critical value $\eta^*$, even a very small value of $\sigma$ can give rise to hump-shaped dynamics under the second-order method. It is thus evident that production externalities dramatically change the topology of the model near the steady state. Benhabib and Wen (2004) show that under the linear approximation method, hump-shaped impulse responses and oscillating cycles emerge when the externalities are large enough to make the model indeterminate. Here, we show that such nonlinear dynamics may already exist when the degree of externalities is below the critical value required for indeterminacy, but they can be captured only by higher-order terms.\(^{15}\)

The above results are obtained because the coefficients of the second-order terms in a standard RBC model without externalities are generally very small (close to zero); hence linear methods usually yield quite accurate solutions. However, once externalities or increasing returns are allowed, the second-order terms can become non-negligible and very large. For example, when $\sigma = 0.3$ and $\eta = 0.4$, we have the following decision rules for capital, consumption, and labor:

\[
\begin{bmatrix}
\hat{k}_{t+1} \\
\hat{c}_t \\
\hat{n}_t
\end{bmatrix} =
\begin{bmatrix}
0.9124 & 0.5484 \\
0.4325 & 0.9101 \\
-0.6272 & 4.4930
\end{bmatrix}
\begin{bmatrix}
\hat{k}_t \\
\hat{A}_t
\end{bmatrix} +
\begin{bmatrix}
0.0463 \\
0.0143 \\
-0.7155
\end{bmatrix}
\begin{bmatrix}
\hat{\sigma}_t \\
9.0494 \\
-19.6185
\end{bmatrix}
\]

\[\text{(24)} \begin{bmatrix}
1 \\
2
\end{bmatrix}
\begin{bmatrix}
2.4241 \\
-0.4161
\end{bmatrix}
\begin{bmatrix}
\hat{A}_t^2
\hat{k}_t \hat{A}_t
\end{bmatrix}
\]

Note that in the labor decision rule the coefficient of $\hat{A}_t^2$ is $-19.6185$ and that of the cross term $\hat{k}_t \hat{A}_t$ is 9.0494, but these coefficients are close to zero when $\eta = 0$. The coefficients of $\hat{A}_t^2$ and $\hat{k}_t \hat{A}_t$ for hours worked increase in absolute value as $\eta$ increases toward the critical value $\eta^*$, but both become zero again as soon as $\eta > \eta^*$. This is shown in Figure 3. Labor has no second-order

---

\(^{13}\) If the variance of the shock $\sigma$ is too small, the first-order and second-order methods yield very similar results. However, second-order methods may no longer be accurate if $\sigma$ is too large. Hence, we choose a large-enough $\sigma$ so that (i) the second-order terms are significant and (ii) the second-order solution is more accurate than linear methods. If we use Wen’s (1998a) model with variable capacity utilization, the values of $\sigma$ can be made much smaller: on the order of $\sigma = 0.03$.

\(^{14}\) This value ($\eta = 0.04$) is also sufficiently below the threshold value for generating the $n$-period cycles found by Coury and Wen (2009).

\(^{15}\) Coury and Wen (2009) show that the Benhabib-Farmar (1994) model has global indeterminacy even when the model’s steady state appears to be a saddle judged by eigenvalues (i.e., for externalities below the critical level required for local indeterminacy based on the first-order approximation method).
Figure 2

Impulse Responses of the Economy to a Positive Technology Shock

First-Order Method

Second-Order Method

σ = 0.3, η = 0

σ = 0.3, η = 0

σ = 0.3, η = 0.4

σ = 0.3, η = 0.4

NOTE: SD, σ = 0.3. See text for details.
terms under indeterminacy because hours worked are a log-linear function of consumption and capital, both of which are state variables under indeterminacy.

Hence, the second-order method predicts more subdued and hump-shaped output responses because externalities introduce a large negative coefficient in front of the squared term for the technology shock, $\hat{A}_t^2$, and a large positive term in front of the cross term for the capital stock and technology, $\hat{k}_t \hat{A}_t$. This means that the effect of a positive technology shock on labor and output is neutralized by the squared term in the impact period but enhanced by the cross term in the subsequent periods, giving rise to the hump-shaped dynamic pattern.

Such nonlinear second-order income effects on hours worked also imply that the economy’s dynamic responses to technology shocks are not symmetric. Under a negative technology shock, the squared term $\hat{A}_t^2$ reinforces the negative shock on impact, while the cross term $\hat{k}_t \hat{A}_t$ offsets the effect of the negative shock in subsequent periods, thus making the impulse response of output to a negative technology shock more contractionary but less persistent than that predicted by the linear method (Figure 4). For example, Figure 4 shows that the initial drop in output is about $-1.6$ by linear approximation and about $-2.5$ by second-order approximation. The predicted half-life is about seven quarters under the linear method and less than four quarters under the second-order method.

Such an asymmetric property is similar to what is observed in the U.S. economy. Expansions tend to be more gradual and long-lived, while contractions tend to be sharper and short-lived (see, e.g., Neftçi, 1984, and Sichel, 1993). Also, Mork (1989) and Hamilton (2003) point out that the economy’s response to oil price shocks is highly asymmetric: A sudden increase in oil prices (an adverse productivity shock) tends to depress the economy, while a sudden decrease in oil prices tends to have little effect. Aguiar-Conraria and Wen (2007) argue that IRS large enough to trigger indeterminacy can explain the large negative impact of oil price increases on the U.S. economy. Here we show that smaller externalities can produce a similar effect under second-order approximation.

Given the same level of $\sigma$, as $\eta$ increases, the impulse responses of output under second-order expansion become increasingly hump-shaped and the initial value can even become negative (which is then followed by a positive hump). This dynamic pattern offers one explanation of a
Figure 4
Responses to a Negative Technology Shock

Figure 5
Responses to a Positive Technology Shock
puzzle in the indeterminacy literature: Technology shocks suddenly become contractionary once the model becomes locally indeterminate.

The left panel in Figure 5 shows the first-order approximation of the model under a positive technology shock when the steady state is indeterminate (i.e., \( \eta = 0.6 \)),\(^{16}\) and the right panel shows the second-order approximation of the model under a positive technology shock when \( \eta = 0.48 \), which is below the critical value for indeterminacy. The left panel shows that technology shocks are contractionary under indeterminacy, which is puzzling because the left panels in Figure 2 indicate that externalities amplify technology shocks in the positive direction. The existing indeterminacy literature has not provided an explanation for this puzzling phenomenon.

However, the right panel in Figure 5 suggests that technology shocks are contractionary from a second-order viewpoint even before the model becomes indeterminate. Therefore, we believe that the puzzle is caused by a large income effect from IRS on hours worked. When the marginal product of labor is high, it is optimal to reduce hours worked and increase leisure under the income effect but increase hours worked under the substitution effect. However, when externalities are below the critical value \( \eta^* \), the income effect is captured only by second-order terms, while the substitution effect is captured by first-order terms. Therefore, the first-order method will show positive impulse responses under technology shocks. Once the model becomes indeterminate, if consumption is treated as a state variable under indeterminacy, the decision rule of labor has only first-order terms because the optimal first-order condition of labor supply (equation (1)) implies that it is a log-linear function of the model’s state space. Hence, in this case there are no second- or higher-order terms in labor; consequently, the strong income effect of a technology shock on hours can be captured only by first-order terms, which explains the puzzle in the left panel of Figure 5.

---

\(^{16}\) The initial consumption level is fixed at the steady state, \( \tilde{c}_t = 0 \).

---

**Investment Behavior**

Based on linear methods, it has become well known in the RBC literature that investment is extremely volatile under technology shocks (see, e.g., Kydland and Prescott, 1982). However, this is not necessarily the case from a second-order perspective (Figure 6).

Because of the asymmetric second-order income effect, investment is very volatile only under a negative technology shock but not under a positive technology shock. In fact, if the elasticity of the labor supply is large enough or the technology shock is not highly persistent, the initial response of savings (investment) to a positive technology shock can be very mild or even negative, as shown in Figure 6. Notice that in each panel of Figure 6, second-order approximation always yields a lower level of investment than first-order approximation. This result is due to an extremely large negative coefficient of the squared term \( \hat{A}_t^2 \) in the investment decision rule,

\[
\begin{bmatrix}
\hat{i}_t \\
\end{bmatrix} = (-2.5029 \quad 21.9379) \begin{bmatrix}
\hat{k}_t, \hat{A}_t \\
\end{bmatrix} + \\
\frac{1}{2} \begin{bmatrix}
-10.1104 & 139.2464 & -439.1138 \\
\end{bmatrix} \begin{bmatrix}
\hat{k}_t^2, \hat{k}_t, \hat{A}_t^2 \\
\end{bmatrix} + 96.9648 \\
\frac{\sigma^2}{2}.
\]

By the same token, however, investment will always appear to be more volatile than that predicted by the linear method under a negative technology shock.

**Second-Order Effects of Sunspot Shocks**

Since the linear solution is used to obtain the coefficients of second-order terms (see Schmitt-Grohé and Uribe, 2004, and Lombardo and Sutherland, 2007), the zone of indeterminacy remains unchanged if indeterminacy arises in the model.\(^{17}\) Farmer and Guo (1994) simulate the Benhabib-Farmer (1994) model under indeterminacy.

---

\(^{17}\) When recursive equilibria fail to exist, the perturbation methods used in the literature and this paper cannot be applied (see Peralta-Alva and Santos, 2010, for the issues involved). Here we follow Benhabib and Farmer (1994) and the existing literature by taking as given the existence of recursive equilibria in the region of indeterminacy.
nacy by the linear method and show that sunspot shocks can generate business cycle comovements among output, consumption, investment, and hours. The effects of positive and negative sunspot shocks under the linear method are symmetric and monotonic (see the left panels in Figure 7, where the upper row pertains to a positive shock and the lower row pertains to a negative shock).

However, as shown in the right panels of Figure 7, under second-order approximation, positive and negative sunspot shocks have asymmetric effects on the economy. In particular, under a negative sunspot shock (the lower-right panel), the economy has hump-shaped impulse responses, while under a positive sunspot shock (the upper-right panel), the economy has no initial hump but tends to overshoot its steady state from above. The hump-shaped impulse responses of output to negative sunspot shocks are in contrast to the analysis of Schmitt-Grohé (2000), where the author argues (under the linear solution method) that sunspot shocks cannot generate hump-shaped output dynamics and forecastable comovements.
Figure 7
Responses to Sunspot Shocks

First-Order Method

Positive Shock

Second-Order Method

Negative Shock

Output Consumption Labor

\( \sigma = 0.3, \eta = 0.6 \)

\( \sigma = 0.3, \eta = 0.6 \)

\( \sigma = 0.3, \eta = 0.6 \)

\( \sigma = 0.3, \eta = 0.6 \)
among output, consumption, hours, and investment. Here we show that this may not necessarily be the case if the second-order effects are considered.

Notice in Figure 7 that both the first-order and second-order methods yield exactly the same initial magnitude of responses for all variables under a positive sunspot shock. The results differ only in the subsequent periods. This occurs because with indeterminacy, consumption is a state variable and the coefficients of second-order terms for labor are zero; thus, in the first period the responses of consumption and labor are due only to first-order effects while the capital stock stays unchanged. Consequently, a sunspot shock has only a first-order effect on the responses of output in the impact period. For example, the decision rules under sunspot shocks take the following form when \( \sigma_s = 0.3 \) and \( \eta = 0.6 \):

\[
\begin{pmatrix}
\hat{k}_{t+1} \\
\hat{c}_{t+1} \\
\hat{n}_t
\end{pmatrix}
= \begin{pmatrix}
0.5070 & 1.0000 \\
-1.0134 & 1.2229 \\
-4.0000 & 8.3333
\end{pmatrix}
\begin{pmatrix}
\hat{k}_t \\
\hat{c}_t
\end{pmatrix}
\]

\[
(25) \quad \frac{1}{2}
\begin{pmatrix}
2.5900 & -9.7502 & 9.1002 \\
0.0327 & 0.1184 & -0.2984 \\
0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\hat{k}_{t}^2 \\
\hat{c}_{t}^2 \\
\hat{k}_{t} \hat{c}_{t}
\end{pmatrix}
\]

where \( \hat{A}_t = 0, \sigma^2 \) is the variance of \( \epsilon_{st} = \hat{c}_t - E_{t-1} \hat{c}_t \), which is the one-period-ahead forecasting error of consumption (sunspots). Clearly, starting from the steady state where \( \hat{k}_t = \hat{k}_{t-1} = \hat{c}_{t-1} = 0 \), the effects of a sunspot shock on consumption in period \( t \) are determined only by the forecasting error \( \epsilon_{st} \) and not by any higher-order terms in the state space. In contrast, the coefficients of the second-order terms are zero in the decision rule for hours worked. In addition, output is a linear function of capital and labor. Therefore, both methods yield the same initial impulse responses for consumption, hours, and output, but in the subsequent periods the responses diverge significantly. In particular, because the cross term \( \hat{k}_t \hat{c}_t \) has a large negative effect and the squared term \( \hat{c}_t^2 \) has a large positive effect on the capital stock, the impulse responses of output and labor become U-shaped under a negative sunspot shock. However, under a positive sunspot shock, they decrease monotonically and overshoot the steady state from above.

## ACCURACY TESTING

As noted by Jin and Judd (2002), second-order approximation is not necessarily better than linear methods in terms of solution accuracy, depending on the models used and the parameter regions. Judd (1998) proposes using the Euler equation error (EEE) as a criterion for non-local accuracy testing. The EEE is expressed as the logarithm of the Euler equation residual

\[
EEE = \log_{10}
\left|
\frac{1}{c_t}
- \frac{1}{\beta E_t} \left[ (\alpha A_t k_{t+1}^{\alpha(1+\eta)} n_t^{(1-\alpha(1+\eta))} - 1 + \delta) \right]^{\frac{1}{1-\alpha}}
\right|
\]

where \( \{c_{t+1}, c_t, n_{t+1}, k_{t+1}\} \) are determined by the second-order policy rules discussed previously. Note that EEEs are generally negative, and a smaller (i.e., more negative) EEE implies improved accuracy. For example, the EEE is negative infinity (\( -\infty \)) when the solution is exact (i.e., 100 percent accurate).

Figure 8 plots the EEEs of the model under the first-order and second-order methods when \( \sigma = 0.3 \) and \( \eta = 0.04 \). The vertical axis represents the EEE, the right-front axis represents deviations of \( \hat{A}_t \) from its steady state 0 in both positive and negative directions, and the left-front axis represents the deviations of capital from its steady state \( \hat{k} = 0 \). Since the first-order EEE lies everywhere above the second-order EEE both at the steady state and when the model is significantly away from the steady state, the second-order solution dominates the linear method in terms of accuracy under the current calibrations.

Alternatively, if we set the technology level to its steady state (\( \hat{A} = 0 \)) or the capital stock to its steady state (\( \hat{k} = 0 \)) and let the degree of exter-
nalities vary, the top panels in Figure 9 show that the degree of accuracy deteriorates as $\eta$ increases toward $\eta^* = 0.4935$. In particular, the second-order method is not necessarily better than the linear method around the critical point of indeterminacy, $\eta^*$. However, for most values of $\eta$, the second-order method dominates the first-order method. The bottom panel in Figure 9 is a two-dimensional graph of the top panels at the point where $\hat{k} = 0$ and $\hat{A} = 0$. It shows that the accuracy of both the first-order and the second-order approximation methods deteriorates as the externality $\eta$ increases toward the critical point of indeterminacy, but the second-order method is more accurate than the linear method except around the point of indeterminacy.

CONCLUSION

This paper shows that externalities and increasing returns to scale can have important implications for business cycle dynamics that have not been fully appreciated by first-order approximation methods. In particular, externalities induce a strong second-order asymmetric income effect on leisure so that hump-shaped output dynamics may emerge even if externalities are sufficiently below the critical value required for local indeterminacy. Similarly, i.i.d. sunspot shocks under indeterminacy may generate forecastable comovements of output, hours, consumption, and investment, in contrast to the conclusion reached by Schmitt-Grohé (2000). Moreover, this asymmetric second-order income effect can generate business cycle dynamics that are qualitatively consistent with the asymmetric nature of the business cycle observed in the U.S. economy. Whether similar results can also be found in models with other types of frictions, such as borrowing constraints, remains to be investigated.
Figure 8
Accuracy with Respect to $\eta$

$\hat{A}_t = 0$

$\hat{k}_t = 0$

$\hat{A}_t = \hat{k}_t = 0$

---

First-Order Method

Second-Order Method
REFERENCES


