Growth Accounting with Technological Revolutions

Andreas Hornstein

Generally, technological progress proceeds at a slow and measured pace, with only incremental improvements seen in existing products and technologies in the economy. At times, however, the pace accelerates, and the economy experiences a technological revolution during which radically new products and technologies are introduced. Recent discussions suggest that the world economy is currently experiencing just such a revolution, or paradigm shift, and that this revolution accounts for some of the observed decline and rebound of productivity growth. For example, David (1991) argues that the effect of information technologies on today's economy is comparable to the effects of the introduction of the dynamo and the subsequent availability of electric power in the late-nineteenth and early-twentieth centuries. It is important to understand the effects of technological progress as reflected in productivity growth because productivity growth determines the economy's long-run growth of output, consumption, and factor income such as wages.

In this article I consider one particular parable of a paradigm shift. This story builds on three assumptions: first, that technological change is associated with the introduction of new goods, in particular that new technologies are embodied in new machines; second, that production units learn about the newly introduced technologies, that is, new technologies do not immediately attain their full productivity potential, but instead productivity increases gradually for some time; and third, that the experience which production units have with existing technologies affects their ability to adopt new technologies.[1]

Andreas.Hornstein@rich.frb.org. I would like to thank Lawrence Santucci for research assistance and Mike Dotsey, Pierre Sarte, and Roy Webb for helpful comments. The views expressed are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] The ideas expressed in this article are based on work by Greenwood, Hercowitz, and Krusell (1997), Hornstein and Krusell (1996), Greenwood and Jovanovic (1998), and Greenwood and Yorukoglu (1997).

In the following pages I summarize the available evidence in support of these assumptions and then speculate on the possible implications of a paradigm shift for future output and productivity growth based on a parametric version of the standard neoclassical growth model. I find that all three assumptions together can account for a substantial and long-lasting decline in measured productivity and output growth during the initial stages of a technological revolution. This initial period is then followed by a long period of above-average long-run growth. Unfortunately, the results depend crucially on how experience with existing technologies affects the ability to adopt new technologies, a feature of the economy about which we know very little. An alternative parameterization of this feature of the economy predicts that the effects of a technological revolution on productivity and output growth might be negligible. Finally, I reconsider the evidence on the slowdown of measured productivity growth and find that it appears to be less dramatic if we calculate real output numbers using a more reliable price index.

1. SOME EVIDENCE ON TECHNOLOGICAL CHANGE

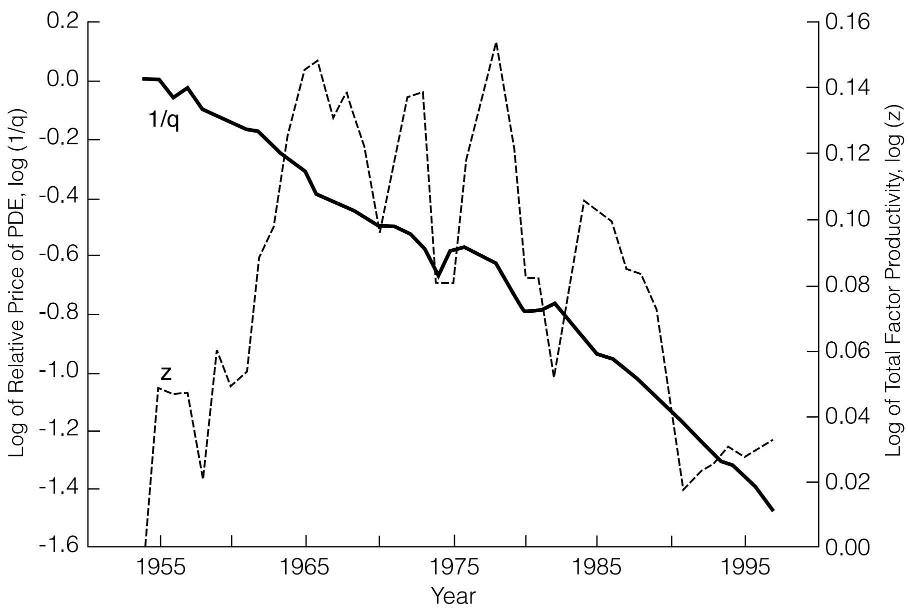
The Rate of Capital-Embodied Technological Change has Accelerated in the Early '70s

When people talk about a new technological revolution, they usually refer to the more widespread use of computers: the application of computers makes new products and services possible, it changes the way production processes are organized, and it is no longer limited to a small fraction of the economy. Unfortunately, many of these observations are anecdotal and provide only limited quantitative support for the impact of computers on the economy. There is one observation, however, that we all make and that might well be quantified; namely, that each new generation of PCs tends to do more things faster than the previous generation, yet we do not have to pay more for these higher-quality PCs. In short, for PCs the price-per-quality unit has been declining at a dramatic rate. This observation applies not only to PCs but to many other products, particularly producer-durable goods such as new capital goods.

While it is easy to say that new products are of better quality, it is difficult to actually measure and compare quality across different goods. In an extensive study, Gordon (1990) has constructed measures of the price of producer-durable equipment that account for quality changes. The line labeled $1/q$ in Figure 1 graphs the price of new producer-durable equipment relative to the price of nondurable consumption for the postwar U.S. economy.[2] I identify the rate of

[2] The series on the relative price of producer-durable equipment is from Greenwood, Hercowitz, and Krusell (1997). I have extrapolated the series from 1990 on using information on the price of producer-durable equipment from the National Income Accounts. Consumption covers nondurable goods and services, excluding housing services. Hornstein (1999) provides a complete description of the data used.

Figure 1 Measures of Embodied and Disembodied Technological Change



price decline with increased productivity in the capital goods producing sector that is embodied in the new capital goods.[3] In this figure it is apparent that producer-durable equipment goods have become cheaper over time relative to consumption goods and that the rate of price decline has accelerated in the mid-'70s from 3 percent before 1973 to 4.3 percent after 1977. A substantial part of the accelerated rate of price decline can be attributed to the fact that information technologies have gained more widespread application in the design of producer-durable equipment.

Learning-by-Doing is an Important Feature of Production

New products or new plants do not attain their full potential at the time they are introduced. Rather, we find that for some period of time productivity for

[3] In general, relative prices may change because the technology changes (shift of the production possibility frontier, PPF) or because of simple substitution between goods (movements along a PPF). Notice, however, that with an unchanged technology we would expect the relative price of a good to fall only if relatively less of the good is produced. Yet we have not observed a decline in the investment rate that should correspond to the decline in the relative price of capital. Work that tries to account for substitution effects finds even more acceleration in the rate of capital-embodied technological change (Hornstein and Krusell 1996).

a new good or plant is increasing. This increase in productivity is attributed to *learning-by-doing* (LBD); that is, firms acquire experience and improve their efficiency in resource use in the process of producing a good. One can think of this process as the accumulation of informational capital. This LBD phenomenon is so widespread and uniform across industries that the management literature summarizes it with the "20 percent rule," according to which labor productivity increases by 20 percent for every doubling of cumulative production (see, e.g., Hall and Howell [1985]).

One of the most frequently cited LBD examples is the case of the liberty ships of World War II. The more ships a navy yard built, the smaller was the labor input required for the next vessel it built (Figure 2). A more recent example of LBD is the production of dynamic random access memory (DRAM) chips in the semiconductor industry. Figure 3 displays the time paths for the average unit price and total shipments of successive generations of DRAM chips. This figure displays two common features of LBD. First, productivity improvements during the early stages of production are dramatic. Second, these improvements are attained within a short period of time, occurring within the first three to five years of production. Indeed, most of the productivity improvements have been made once shipments of a chip generation reach their peak. Notice also that during the first few years a new generation of chips is produced, the unit price is higher than the one of the previous generation.[4] The DRAM chip example also points to an important feature for my discussion of accelerated capital-embodied technological change: How much of the experience accumulated in the production of one generation of DRAM chips can be transferred to the production of the next generation of chips? More generally, how much of the experience accumulated for existing technologies can be applied to new technologies? The answer to this question is still open. Evidence from the semiconductor industry indicates that the transfer of experience is limited (Irwin and Klenow 1994).

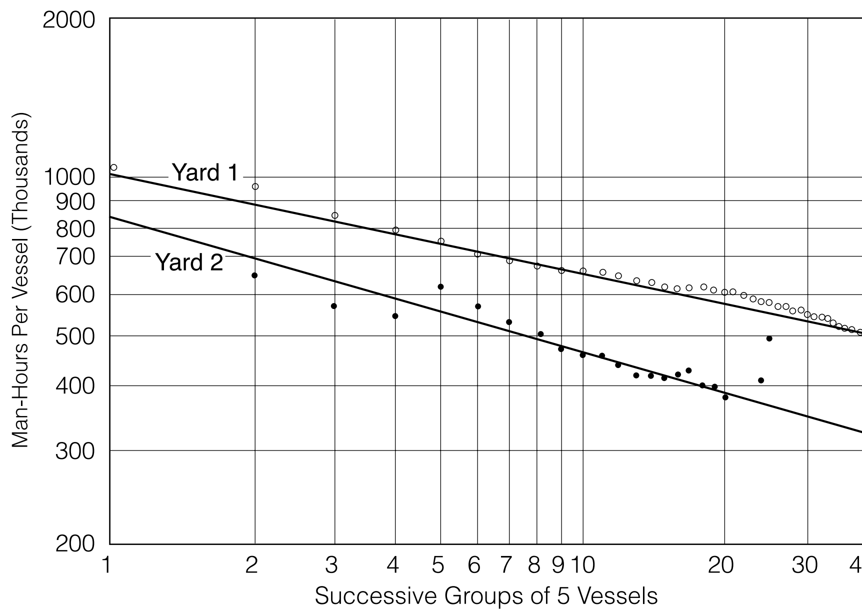
New Technologies Diffuse Slowly Through the Economy

When a radically new technology becomes available, not everybody in the economy will adopt this new technology simultaneously. For some time the use of the old and new technology will coexist while firms continue to make improvements in the old technology. This situation will occur since there are costs to adopting new technologies such as learning costs. Potentially, a new technology may be much more productive than the old technology, but initially users of the new technology have to start with a low experience level relative to that of old technologies.

---

[4] My interpretation that a decline in average unit price reflects an increase in productivity should be taken with a grain of salt since the market structure in the semiconductor industry is only approximately competitive.

Figure 2 Reductions in Man-Hours per Vessel with Increasing Production



Source: Lucas (1993).

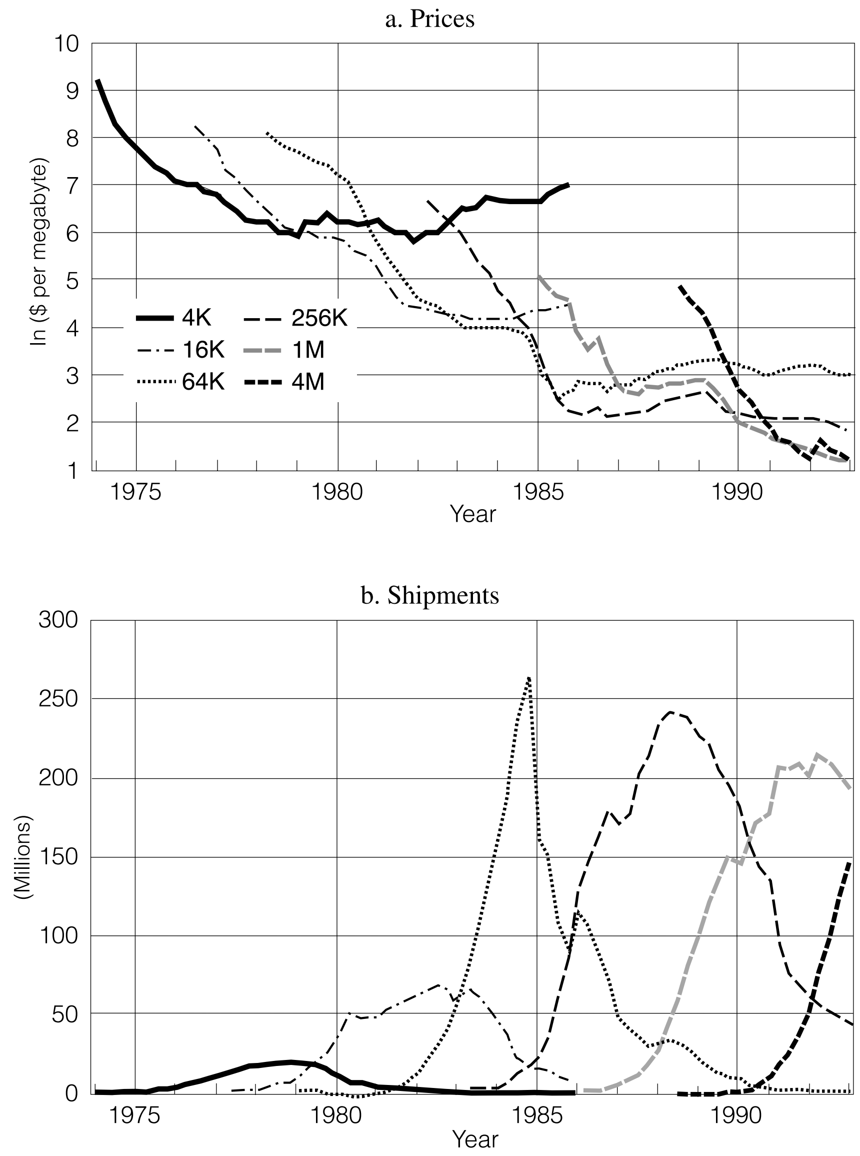
The idea that new technologies diffuse slowly through the economy relates to the observation that the use of new products diffuses slowly through the economy. Experts have made this observation for a wide variety of products from diesel locomotives to DRAM chips (Figures 3 and 4).[5] David (1991) makes a similar observation on the diffusion of the use of electrical power in the late-nineteenth and early-twentieth centuries.

2. TECHNOLOGY DIFFUSION IN A SIMPLE MODEL OF CAPITAL-EMBODIED TECHNOLOGICAL CHANGE AND LEARNING

The appearance of a new technology in the economy can significantly affect output and productivity growth during the transitional period when the new technology replaces the old technology. These effects come about because learning introduces another kind of capital that is not measured, *informational*

[5] Figure 3b plots shipments of DRAM chips from different generations, and Figure 4 plots the numbers of diesels in use as a fraction of the total number of locomotives.

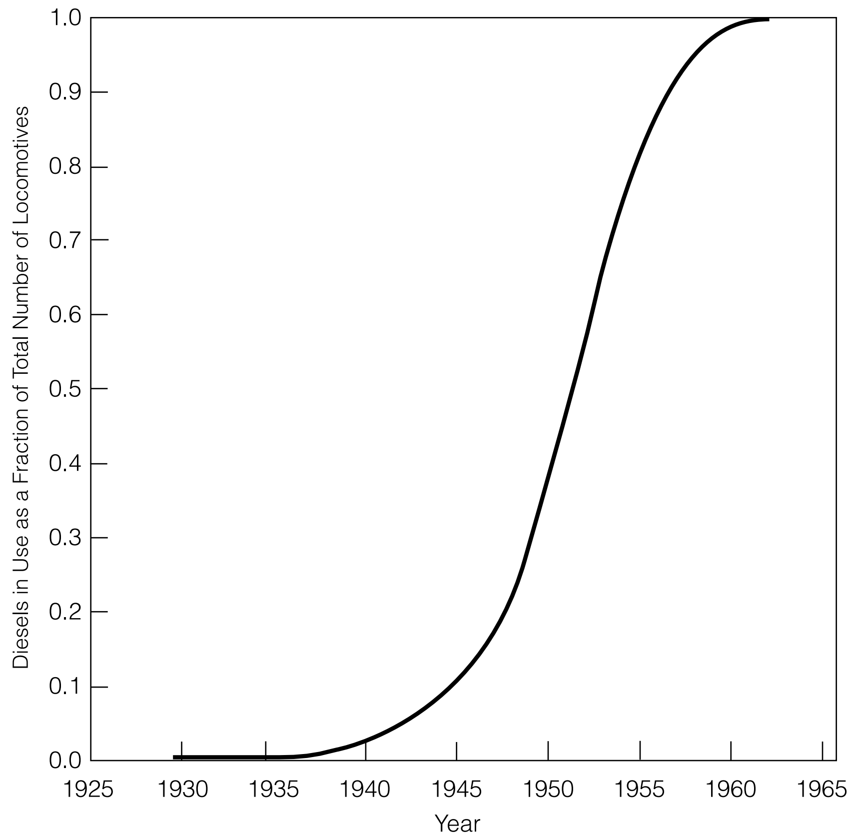
Figure 3 Dynamic Random Access Memory Semiconductors



Source: Irwin and Klenow (1994).

capital, and during the transitional period this capital stock can change significantly. This change in informational capital has real output growth effects, and it creates a problem for measuring productivity growth.

Figure 4 Diesel Locomotion in the U.S. Railroad Industry, 1925–66: Diffusion



Source: Jovanovic and McDonald (1994).

Informational capital represents the economy's experience with various vintages of capital goods, and it is not part of our standard measure of capital. Consequently, we do not measure changes of informational capital that occur during transitional periods. In particular, after we correct for depreciation, we assign the same value to capital from different vintages. So during transitional periods when substantial investment in new technologies with lower experience occurs, we tend to overestimate the contribution to output from investment in these new technologies. Because we overestimate capital accumulation, we underestimate total factor productivity growth. There is also a real effect of learning, since output growth slows down in the transitional period. A feature of this learning is that during the transitional period, production with new technologies is relatively less efficient than production with old technologies.

In the next section I will try to quantify the implications for output and productivity growth measurement when a new technology is introduced in a simple vintage capital model with learning. The structure of the model is very mechanical and many of the elements discussed above are taken as exogenous.

The Solow Growth Model and Growth Accounting

I will start with the standard Solow growth model, which assumes a neoclassical production structure and a constant savings and investment rate. Each period, a homogeneous good y_t is produced using a constant-returns-to-scale technology with inputs capital k_t and labor n_t,

$$y_t = z_t k_t^{\alpha} n_t^{1-\alpha}, \tag{1}$$

where the elasticity of output with respect to capital satisfies $0 < \alpha < 1$. For simplicity I have assumed a Cobb-Douglas production function. Technological change is represented through changes in total factor productivity z_t and is disembodied; that is, with the same inputs, output increases when total factor productivity (TFP) increases. The economy's endowment of labor is fixed, $n_t = 1$. The output good can be used for consumption c_t or investment i_t:

$$c_t + i_t = y_t. \tag{2}$$

Investment is used to augment the capital stock and capital depreciates at a constant rate δ:

$$k_{t+1} = (1 - \delta)k_t + i_t, \tag{3}$$

and $0 < \delta < 1$. Expenditures on investment are assumed to be a constant fraction σ of output,

$$i_t = \sigma y_t, \tag{4}$$

and $0 < \sigma < 1$.

Assume that TFP grows at a constant rate, $z_{t+1} = \gamma_z z_t$ and $\gamma_z \geq 1$. It can be easily verified that an equilibrium exists for this economy where output, consumption, investment, and the capital stock all grow at constant rates. Such an equilibrium is called a *balanced growth path*. For the following let g_x denote the gross growth rate of the variable x: that is, $g_x = x_t/x_{t-1}$. From the savings equation (4), it follows that if both investment and output grow at a constant rate, then they must grow at the same rate, $g_y = g_i = g$. In turn, the resource constraint (2) shows that consumption must grow at that same rate $g_c = g$. Dividing the capital accumulation equation (3) by the capital stock k_t subsequently shows that if the capital stock grows at a constant rate, it must grow at the same rate as investment, $g_k = g$. Finally, the production function (1) relates the economy's output growth rate to the growth of inputs and the exogenous productivity growth rate $g = g_y = \gamma_z g_k^{\alpha} = \gamma_z g^{\alpha}$.[6] From this expression one can see that the economy's growth rate on the balanced growth path increases with the productivity growth rate and with the capital elasticity of output,

$$g = \gamma_z^{1/(1-\alpha)}. \tag{5}$$

We know that TFP in this economy is z_t, but how can we measure TFP if we do not observe z_t? In order to calculate the percentage change of TFP, take the log of equation (1), take the first difference,[7] and solve for the TFP growth rate $\hat{z}$,

$$\hat{z}_t = \hat{y}_t - \alpha\hat{k}_t - (1-\alpha)\hat{n}_t.$$

Here the measure of TFP growth requires observations on the growth rates of output and inputs and knowledge of the parameter α. Solow's (1957) important insight was that in a competitive economy α can be measured through observations on factor income shares. Suppose that all markets in this economy are competitive and that everybody has access to the technology represented by (1). Then consider a firm that maximizes profits, sells the output good at a price p_t, and hires labor (capital) services at the wage rate w_t (capital rental rate u_t). In order to maximize profits, the firm will hire labor (capital) services until the marginal revenue from the last unit of labor (capital) services hired equals its price:

$$p_t MPN_t = p_t\,(1-\alpha)\,z_t k_t^{\alpha} n_t^{-\alpha} = w_t \tag{6a}$$

$$p_t MPK_t = p_t \alpha z_t k_t^{\alpha-1} n_t^{1-\alpha} = u_t. \tag{6b}$$

Multiplying each side of the equation with $n_t/p_t y_t$ $(k_t/p_t y_t)$ shows that the labor (capital) coefficient in the production function equals the share in total revenues that goes to labor (capital):[8]

$$1-\alpha = w_t n_t/p_t y_t = s_{nt}$$

$$\alpha = u_t k_t/p_t y_t = s_{kt}.$$

[6] From equations (1), (3), and (4), it follows that a balanced growth path is associated with a particular level of the capital stock in the initial period k_0. One can show that the economy converges toward this balanced growth path if it starts with a different level of capital.

[7] In the following, a hat denotes the net growth rate of a variable: for example, $\hat{x}_t = (x_t - x_{t-1})/x_t$. For small changes in a variable, the first difference of the logs approximates the growth rate; for example, $\hat{x}_t = \ln x_t - \ln x_{t-1}$.

[8] Since the two coefficients sum to one, total payments to the two production factors capital and labor exhaust revenues; that is, there are zero profits. This is not specific to the assumption of a Cobb-Douglas production function. In general, profits are zero when production is constant returns to scale and all markets are competitive.

We can therefore measure productivity growth using observations on output growth, input growth, and factor income shares. This measure of TFP growth is the *Solow residual*:

$$\hat{z}_t^m = \hat{y}_t - s_{kt}\hat{k}_t - s_{nt}\hat{n}_t = \hat{z}_t. \tag{7}$$

The Solow residual provides an accurate measure of disembodied technological change not only for a Cobb-Douglas production structure but for any constant-returns-to-scale economy, as long as we are willing to assume that all markets are competitive. Finally, note that the wage and capital rental rate equations (6a and 6b) also imply that on a balanced growth path real wages w_t/p_t will grow at the economywide growth rate g, which is determined by the productivity growth rate, and that the real rental rate of capital is constant.

Capital-Embodied Technological Change

The secular decline of the relative price of producer-durable goods suggests that a substantial part of technological progress is embodied in new capital goods. A straightforward modification allows me to account for capital-embodied technological change in the Solow growth model. In the model described above the homogeneous output good can be used for consumption or investment, and the marginal rate of transformation between consumption and investment goods is fixed. In particular I have assumed that one unit of the consumption good can be transformed into q_t units of the investment good and $q_t = 1$. In order to show that over time the economy becomes more efficient in the production of capital goods, I simply assume that over time q_t grows at a constant rate, $q_{t+1} = \gamma_q q_t$ and $\gamma_q \geq 1$. The resource constraint for the output good is now

$$c_t + i_t/q_t = y_t. \tag{2a}$$

At the same time that the economy becomes more efficient in the production of capital goods, the relative price of capital goods $1/q_t$ declines. I continue to measure output in terms of consumption goods and assume that expenditures on investment goods in terms of consumption goods represent a constant fraction of income:

$$i_t/q_t = \sigma y_t. \tag{4a}$$

Analogous to the previous economy, there is a balanced growth path where output, consumption, investment, and capital all grow at constant rates:

$$g_y = g_c = (\gamma_z \gamma_q^{\alpha})^{1/(1-\alpha)} \text{ and } g_i = g_k = (\gamma_z \gamma_q)^{1/(1-\alpha)}. \tag{5a}$$

The measurement of TFP, that is, disembodied technological change, is affected in two ways by the presence of capital-embodied technological change. First, the capital stock measure is constructed as the cumulative sum of undepreciated past investment based on equation (3). Since changes in the quality of new capital goods are the hallmark of embodied technological change, we have to use an appropriate price index that accounts for these quality changes when we deflate nominal investment series to obtain real investment expenditures. Second, because the relative price between consumption and investment goods is changing over time, we have to decide whether we want to measure output in terms of consumption or investment goods. Since ultimate well-being in the economy depends on the availability of consumption goods, I decide to measure output in terms of consumption goods. The line labeled z in Figure 1 displays the measured TFP levels for the postwar U.S. economy. Here the measured capital stock is adjusted for embodied technological change using data on the relative price of durable goods.[9] Notice that contrary to capital-embodied technological change, which was positive for all of the postwar period, measured TFP does not represent a success story for the U.S. economy. Although TFP was increasing rapidly in the late ’50s and ’60s, TFP stagnated in the early ’70s and has actually declined since the mid-’70s when the rate of embodied technological change accelerated. Recently, starting in the ’90s, there has been a slight recovery of TFP, but the apparent negative trend in the ’70s and ’80s seems hard to rationalize.

Learning and Growth Accounting

The observed decline in measured TFP could simply be due to measurement error; that is, there never was a decline in actual TFP. To make sense of this explanation I provide a candidate for what has been mismeasured, and I argue why the measurement problem got worse in the mid-’70s and why we now observe a trend reversal. I suggest that the effective stock of capital has been mismeasured. In particular, I consider the possibility that standard measures of capital do not include informational capital in the economy. In the following I introduce informational capital into the Solow growth model through a simple model of learning. I show that even though measured capital does not include informational capital, there is no measurement problem on the balanced growth path; the measured capital stock may overestimate the effective capital stock during transitional periods when there are significant changes in the economy’s informational capital stock.

Assume that new capital goods do not immediately attain their full potential, but in the process of producing goods, more is learned about each capital good and the efficiency with which it is used increases over time. We now have to distinguish between different vintages of capital goods because a producer has less experience with a capital good that is newly introduced than with a capital good that has been around for some time. Let $k_{t,a}$ denote a capital good

[9] The measure of TFP is based on work by Greenwood, Hercowitz, and Krusell (1997) as extended in Hornstein (1999). For a more detailed description see either of the two references.

that is a years old at time t. If this capital good is employed with $n_{t,a}$ units of labor, output $y_{t,a}$ is

$$y_{t,a} = z_t e_{t,a} k_{t,a}^{\alpha} n_{t,a}^{1-\alpha}, \tag{1a}$$

where $e_{t,a}$ is the experience index of a capital good that is a years old. For simplicity I assume that maximal experience is one and convergence to it is geometric at rate λ:

$$1 - e_{t+1,a+1} = \lambda(1 - e_{t,a}) \text{ for } a = 1, 2, \ . \ . \ . \ , \tag{8}$$

starting from some initial experience level $0 \leq e_{t,1} \leq 1$, and $0 < \lambda < 1$. I continue to assume that capital depreciates at rate δ:

$$k_{t+1,a+1} = (1 - \delta) k_{t,a}. \tag{3a}$$

Total output, employment, and investment are

$$y_t = \sum_{a=1}^{\infty} y_{t,a}, n_t = \sum_{a=1}^{\infty} n_{t,a}, \text{ and } i_t = k_{t+1,1}, \tag{9}$$

and I continue to assume that the markets for output, labor, and the different capital vintages are all competitive. An attractive feature of this model is the existence of an exact aggregate capital index. We can write aggregate output as a Cobb-Douglas function of total employment and the aggregate capital index $\bar{k}$:[10]

$$y_t = z_t \bar{k}_t^{\alpha} n_t^{1-\alpha}, \text{ and } \bar{k}_t = \sum_{a=1}^{\infty} e_{t,a}^{1/\alpha} k_{t,a}. \tag{9a}$$

From this expression one can see how informational capital, $e_t = \{e_{t,a} : a = 1, 2, \ . \ . \ . \}$, affects aggregate output. Note that the usual measure of the

[10] The aggregate capital index can be derived as follows. A profit-maximizing competitive firm using vintage a capital goods hires labor until it equates the marginal revenue of labor with its marginal cost, $p(1 - \alpha) z e_a k_a^{\alpha} n_a^{-\alpha} = w$. Solving this expression for n_a defines the demand for labor by firms using vintage a capital, $n_a = [(1 - \alpha) z e_a / (w/p)]^{1/\alpha} k_a$. The real wage w/p then adjusts such that the total demand for labor is equal to the supply of labor:

$$n = \sum_{a=1}^{\infty} n_a = [(1 - \alpha) z/(w/p)]^{1/\alpha} \sum_{a=1}^{\infty} e_a^{1/\alpha} k_a = [(1 - \alpha) z/(w/p)]^{1/\alpha} \bar{k}.$$

One can solve this expression for the equilibrium real wage, substitute it in the labor demand equation, and obtain the output of firms using vintage a capital as $y_a = z e_a^{1/\alpha} k_a (n/\bar{k})^{1-\alpha}$. Total output is then

$$y = \sum_{a=0}^{\infty} y_a = z \bar{k}^{\alpha} n^{1-\alpha}.$$

economy's capital stock as the sum of undepreciated past investment does not take into account the informational capital

$$k_t^m = \sum_{a=1}^{\infty} k_{t,a} = (1-\delta)k_{t-1}^m + i_{t-1}. \tag{9b}$$

To close the model I identify what determines initial experience with a new capital good. I assume that there is an externality, and experience with older capital goods is partially transferrable to new capital goods according to the following expression:

$$e_{t+1,1} = \frac{\theta}{1-\rho} \sum_{a=1}^{\infty} \rho^{a-1} n_{t,a} e_{t,a}, \tag{8a}$$

with $0 < \rho < 1$ and $\theta > 0$. This formulation of the learning externality follows Lucas (1993). The factor ρ^a measures the extent to which experience with vintage a contributes to initial experience with new capital goods. The larger that ρ is, the more important is experience with existing capital goods. Since $\rho < 1$, experience with older vintages is less important for the initial experience with a new capital good. Notice also that I have assumed the contribution of vintage a is weighted by how intensively this vintage is used, whereby I measure the intensity of use by the share in employment.

The balanced growth path of this economy is very similar to the path of the previous economy. Output, consumption, investment, and capital grow at the same rates, and the initial experience e_1 is constant. Because the initial experience is constant, the informational capital does not change, $e_{ta} = e_a$, and the exact aggregate capital index (9a) and the measured capital stock (9b) grow at the same rate. Therefore, the Solow residual accurately reflects true growth of TFP. If the economy is not on the balanced growth path, three things happen. Initial experience and the informational capital changes over time, changes in the measured capital stock do not accurately reflect changes in the exact aggregate capital index, and the Solow residual mismeasures true TFP growth.

The economy may not be on its balanced growth path for various reasons. Here I consider the possibility that the acceleration of capital-embodied technological change in the mid-'70s was associated with a qualitative change in the kind of technology used. Furthermore, the adoption of this new technology proceeded gradually. To be more specific assume that at some time t_0 this new qualitatively different technology becomes available. From this point on I distinguish between vintages belonging to the old technology, $i = 1$, and vintages belonging to the new technology, $i = 2$. This means that in any time period t output, capital, employment, and experience are now indexed by the type of technology i and its vintage a, $\{y_{t,a}^i, k_{t,a}^i, e_{t,a}^i, n_{t,a}^i\}$. I assume that the new technology is potentially better because capital-embodied technological progress proceeds at a higher rate for the new technology $\gamma_q^2 > \gamma_q^1$ and $q_{t_0}^2 = q_{t_0}^1$. At first, however, the new technology may be worse because the economy has less experience with it. Since the new technology may be initially inferior, I assume that the new technology diffuses slowly. In particular, only a fraction ψ_t of total investment expenditures is used for the purchase of capital goods with the new technology, and

$$\psi_t \begin{cases} = 0 & \text{for } t < t_0, \\ \in (0,1) & \text{for } t = t_0 + 1, \ldots, \ t_0 + T, \\ = 1 & \text{for } t > t_0 + T, \end{cases} \tag{10}$$

and ψ_t increases monotonically. As before, initial experience e^i for a new vintage of a technology i depends on the existing experience with older vintages of that technology,

$$e_{t+1,1}^i = \frac{\theta}{1-\rho} \sum_{a=1}^{\infty} \rho^{a-1} n_{t,a} e_{t,a}^i. \tag{8b}$$

For completeness assume that the experience of a new technology vintage that never existed is zero; that is, $e_{ta}^2 = 0$ for $t - a < t_0$.[11]

In order to consider the quantitative implications of the diffusion of a new technology, I select parameter values for the economy that are consistent with observations on long-run growth, the evidence on the accelerated embodied technological change, learning in the economy, and the diffusion of new technologies.

In the postwar U.S. economy, the average annual depreciation rate is about 10 percent, the average investment rate is about 20 percent, and the average capital income share is about 30 percent. I assume that there is no disembodied technological change such that we can interpret the output and measured TFP growth rates as possible losses/gains due to the diffusion of a new technology. I also assume that the new technology is implemented beginning in 1974 and that it will take 40 years for all new investment to take the form of the new technology. This means that we have passed the midpoint of the diffusion process. The parameterization of the diffusion process (T, ψ_t) is consistent with observations as discussed in Section 2. The rate of capital-embodied technological change for the old and new technology corresponds to the average rate of decline for the relative price of equipment before and after 1974. The parameterization of the internal learning process $(\lambda, e_{t_0,1}^1)$ is based on Bahk and Gort (1993). We know the least about learning externalities (ρ, θ). I simply assume that $\rho = 0.8$ and that in the years before 1974 the economy was on its balanced growth path. With this observation I can recover the value of θ.

[11] The assumption that experience is not transferable across technologies is extreme, but allowing for partial transferability changes the results insignificantly.

Table 1 Parameter Values

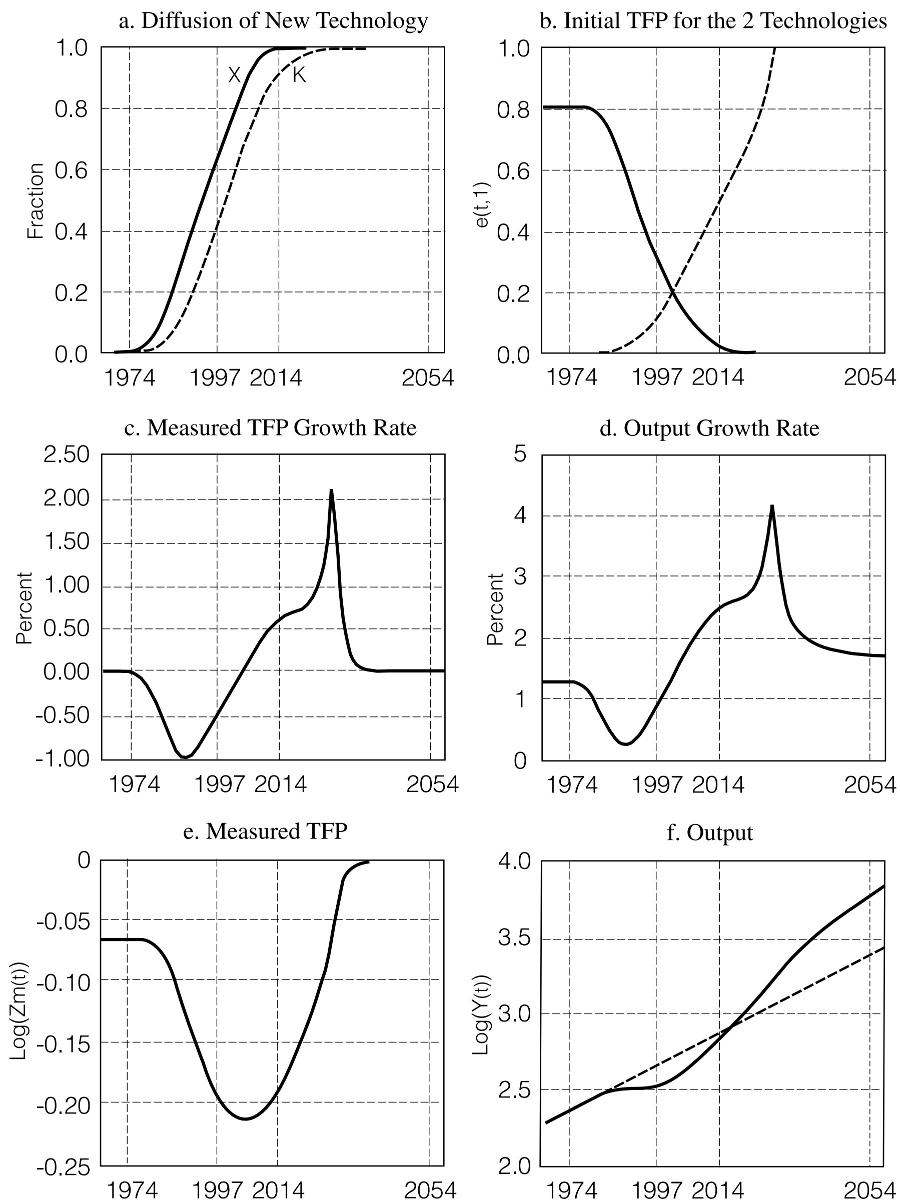
Solow growth model	$\alpha = 0.3$, $\delta = 0.1$, $\sigma = 0.20$
Disembodied technological change	$\gamma_z = 1.00$
Capital-embodied technological change	$\gamma_q^1 = 1.03$ and $\gamma_q^2 = 1.04$
Learning	$\lambda = 0.7$ and $e^1_{t_0,1} = 0.8$
Learning externality	$\rho = 0.8$ and $\theta = 12.11$
Diffusion	$T = 40$ and ψ_t follows an S-shaped diffusion (third-order polynomial)

The results are displayed in Figure 5. Panel a shows the gradual diffusion of the new technology for investment and the capital stock. Since investment adds to the existing capital stock, the diffusion of the new technology in the total capital stock proceeds at a slower rate than it does relative to investment. Panels c and d display measured TFP growth rates and output growth rates, and we observe a long-lasting and substantial decline in measured TFP growth and output growth (1 percentage point). This decline bottoms out in the mid-'80s, and we are now in a recovery phase. According to this simulation we can expect a considerable increase of the trend growth rates for measured TFP and output for the next 20 years. Panel f shows that the effects of the lower output growth are quite substantial in the sense that another 15 years have to pass before the level of output catches up with the initial balanced growth path.[12]

Why do we get these big effects during the transitional period when the new technology is adopted? The simple answer lies in the graph of initial experience for the two technologies (panel b of Figure 5). Notice that initial experience in the old technology is declining during the transitional phase. The decline occurs because according to the specification of the learning externality (8a), the contribution of a vintage is weighted by employment in that vintage. During the transitional phase employment shifts from old to new technologies and, with this learning specification, the economy tends to "forget" about the old technology. Sizeable changes in output and measured TFP growth do not occur, however, because investment in new technologies starts out with a low experience; these changes make up only a small fraction of total investment after all. Rather the big changes in output and measured TFP growth occur because initial experience is falling for investment in old technologies, and this investment contributes the most to total capital accumulation.

[12] The results are sensitive with respect to technology spillovers ρ. If spillovers are unimportant ($\rho = 0.5$), then the decline in measured TFP growth is much more persistent, and output growth does not overshoot very much.

Figure 5 A Transition Path with Big Effects



We can evaluate the importance of this effect by changing the specification of the learning externality (8a) such that we do not weight experience by employment; that is, how much the experience of an old vintage contributes to the initial experience of a new vintage is independent of how intensively the old vintage is used:

$$e^i_{t+1,1} = \frac{\theta}{1-\rho} \sum_{a=1}^{\infty} \rho^{a-1} e^i_{t,a}. \tag{8c}$$

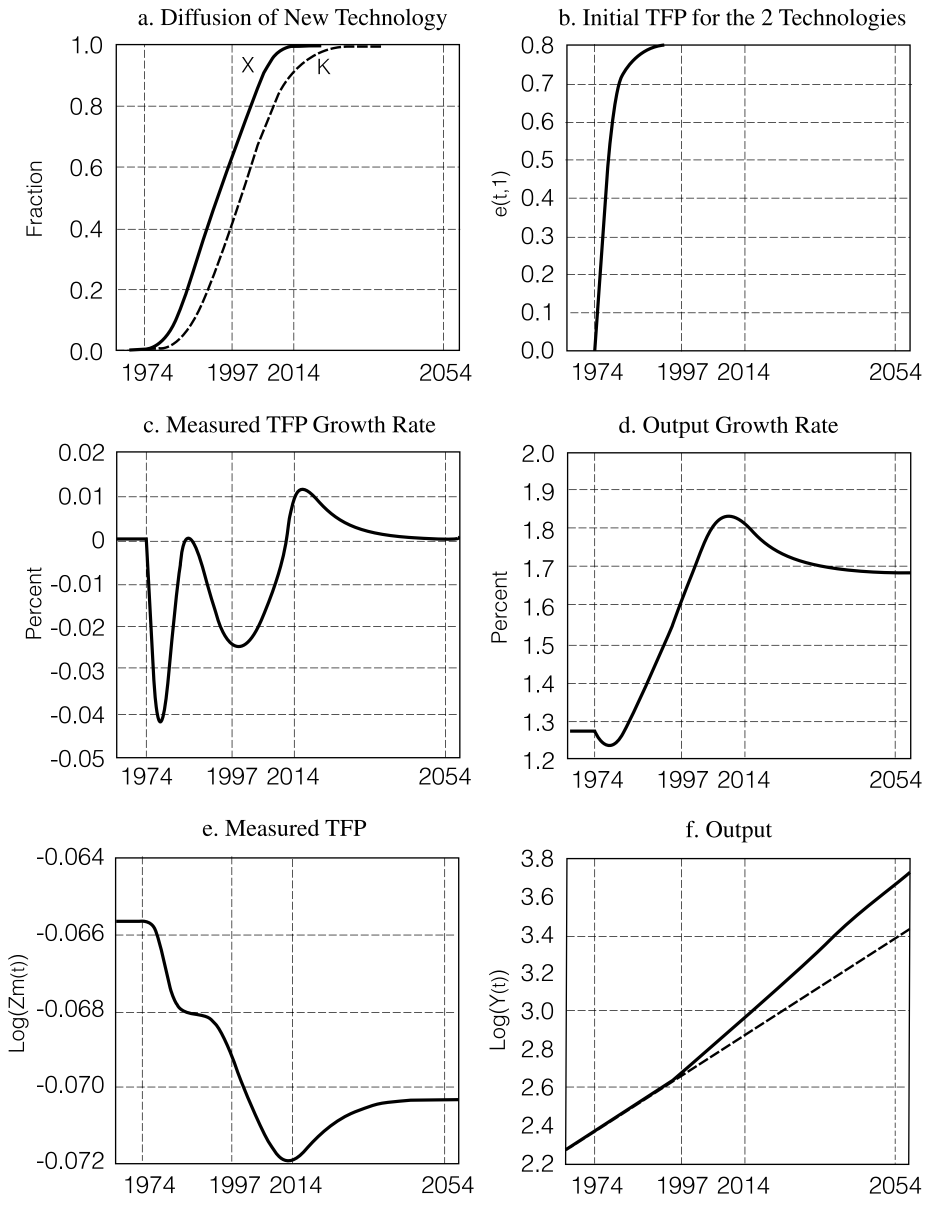
The results of this change are displayed in Figure 6. Note that with this specification initial experience with the old technology remains constant at 0.8. As we can see, the maximal reduction in measured TFP growth is now only 0.04 percentage points, as opposed to 1 percentage point previously, and there is almost no decline in output growth; the maximal increase corresponds to the balanced growth increase of about 0.5 percentage points.

I am not aware of any empirical work that has studied the quantitative properties of the transfer of knowledge in the economy and that would allow us to pick between the two learning specifications (8a) and (8c). Although I find specification (8a) reasonable—in the sense that intensity of use should matter for the transfer of knowledge—and although it is quite possible that an economy "forgets" about old technologies if they are not used, I do not believe that the process occurs as fast as implied by the specification above. If, as I believe, the economy is not that forgetful, then specification (8c) may be a good short- to medium-term approximation, and I would have to conclude that the possible effects of a technological revolution are limited.

3. RECONSIDERING THE MEASUREMENT OF PRODUCTIVITY GROWTH

This article reviews the possible implications of a technological revolution for the measurement of the U.S. economy's productivity performance. I have shown evidence for the acceleration of capital-embodied technological change and at the same time a substantial decline of TFP, which represents disembodied technological change. I have argued that part of the decline in TFP can in principle be attributed to a measurement problem associated with accumulating informational capital during a technological revolution. Unfortunately, the process by which informational capital is accumulated in an economy is not well understood, and any exercise that studies this aspect of the economy has to be somewhat speculative in nature. I would like to conclude my discussion of the U.S. economy's productivity performance with one more observation. Although this observation makes the description of productivity behavior even more ambiguous, it seems to indicate that the performance of the U.S. economy has not been as bad as Figure 1 suggests.

Figure 6 A Transition Path with Small Effects



My discussion of the implications of a technological revolution has focused on problems associated with the measurement of capital in a broad sense. Part of the measurement problem is accounting for changes in the quality of producer-durable goods, but for this part I have taken the view that Gordon's (1990) price index does account for most of the quality changes that occur for producer-durable goods. I have also identified embodied technological progress with the rate of decline of the price of producer-durable goods relative to consumption goods. At this point I should note that the quality of consumption goods also changes over time, a process that in principle is no different from that of producer-durable goods. But this means that for the construction of a consumer price index one also has to be careful how one accounts for quality change in new consumer goods. To the extent that our consumer price index does not capture quality changes in goods, we will overestimate the rate of price increase and underestimate the growth in real consumption.[13]

The diffusion of information technologies has certainly affected the quality of consumer goods we are now able to purchase, an observation that is most evident for consumer services. Take, for example, the services provided by the financial sector: we are now able to obtain cash at conveniently located automatic teller machines, we can access our bank accounts and make transactions from home, we can trade shares directly on the Internet without going through a broker, etc. It has always been recognized that accounting for quality changes is relatively more difficult for services than it is for commodities, a problem that has probably been exacerbated through increasingly widespread use of the new information technologies.[14]

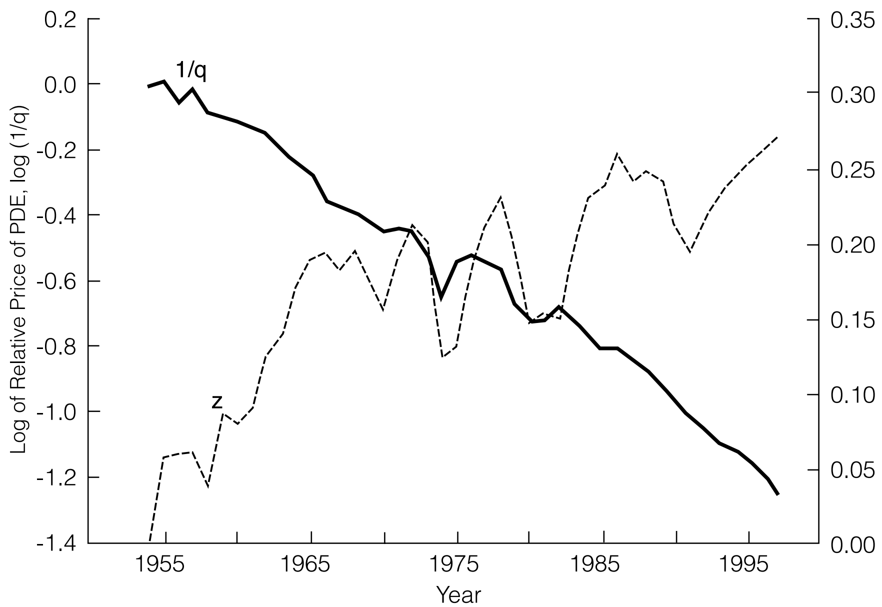
A price index that overestimates the rate of price increase for consumer goods has two implications for the productivity growth measures I have discussed in this article. First, since the rate of decline for the price of producer-durable goods relative to the price of consumer goods is overestimated, the rate of embodied technological change is overestimated. Second, because output as measured in terms of consumption goods is actually growing faster than the consumption price index seems to indicate, the rate of disembodied technological change is underestimated. Can we say anything about the potential magnitude of this measurement problem?

I have argued that the measurement problem is probably more relevant for the consumption of services rather than the consumption of goods. If services made up only a small fraction of consumption, the potential bias would probably be small, but today expenditures on services excluding housing are about 50 percent higher than expenditures on nondurable goods. Since the price index for nondurable consumption goods appears to be less subject to measurement

[13] For a discussion of the potential biases in the consumer price index, see Boskin et al. (1996).

[14] See Griliches (1994) on the quality of output and price indexes for different industries.

Figure 7 Measures of Embodied and Disembodied Technological Change Reconsidered



error than the price index for services, I recalculated the estimates for embodied and disembodied technological change using the price index for nondurable consumption goods only, rather than the price index for nondurable goods and services (excluding housing) as shown in Figure 1. The revised productivity series are graphed in Figure 7.

The alternative measure of real output mainly effects the measure of disembodied technological change as opposed to the measure of embodied technological change. Embodied technological change now proceeds at a slower rate, and it does not accelerate as much in the mid-'70s.[15] The effect on the measure of disembodied technological change as reflected in TFP growth is more dramatic. With the new measure of real output, TFP growth still stagnates starting in the '70s, but there is no longer a secular decline. Notice also the strong recovery of TFP since the early '90s, although it remains to be seen whether this is a purely cyclical upswing or whether it represents a change in the long-run growth path for TFP. In conclusion, as is evident from Figures 1 and 7, the productivity

[15] The rate of price decline now accelerates from 2.7 percent before 1973 to 3.5 percent after 1977.

performance of the U.S. economy appears to be consistent with a wide range of views, from pessimistic to guardedly optimistic. Clearly more work has to be done.

REFERENCES

Bahk, Byong-Hyong, and Michael Gort. "Decomposing Learning by Doing in New Plants," *Journal of Political Economy,* vol. 101 (August 1993), pp. 561–83.

Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson. "Toward a More Accurate Measure of the Cost of Living." Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index, 1996.

David, Paul. "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror," in *Technology and Productivity: The Challenge for Economic Policy.* Paris: Organization for Economic Cooperation and Development, 1991, pp. 315–47.

Gordon, Robert J. *The Measurement of Durable Goods Prices.* Chicago: The University of Chicago Press, 1990.

Greenwood, Jeremy, and Boyan Jovanovic. "Accounting for Growth." Manuscript. 1998.

Greenwood, Jeremy, Zvi Hercowitz, and Per Krusell. "Long-Run Implications of Investment-Specific Technological Change," *American Economic Review,* vol. 87 (June 1997), pp. 342–62.

Greenwood, Jeremy, and Mehmet Yorukoglu. "1974," in *Carnegie-Rochester Conference Series on Public Policy,* vol. 46 (June 1997), pp. 49–95.

Griliches, Zvi. "Productivity, R&D, and the Data Constraint," *American Economic Review,* vol. 84 (March 1994), pp. 1–23.

Hall, G., and S. Howell. "The Experience Curve from the Economist's Perspective," *Strategic Management Journal*, vol. 6 (1985), pp. 197–212.

Hornstein, Andreas. "Reconsidering Investment-Specific Technological Change." Manuscript. 1999.

__________, and Per Krusell. "Can Technology Improvements Cause Productivity Slowdowns?" in Ben Bernanke and Julio Rotemberg, eds., *NBER Macroeconomics Annual 1996.* Cambridge, Mass.: MIT Press, 1996, pp. 209–59.

Irwin, Douglas A., and Peter J. Klenow. "Learning-by-Doing Spillovers in the Semiconductor Industry," *Journal of Political Economy,* vol. 102 (December 1994), pp. 1200–27.

Jovanovic, Boyan, and Glenn M. MacDonald. "Competitive Diffusion," *Journal of Political Economy,* vol. 102 (February 1994), pp. 24–52.

Lucas, Robert E., Jr. "Making a Miracle," *Econometrica,* vol. 61 (March 1993), pp. 251–72.

Solow, Robert. "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics,* vol. 70 (1957), pp. 65–94.

Two Approaches to Macroeconomic Forecasting

Roy H. Webb

Following World War II, the quantity and quality of macroeconomic data expanded dramatically. The most important factor was the regular publication of the National Income and Product Accounts, which contained hundreds of consistently defined and measured statistics that summarized overall economic activity. As the data supply expanded, entrepreneurs realized that a market existed for applying that increasingly inexpensive data to the needs of individual firms and government agencies. And as the price of computing power plummeted, it became feasible to use large statistical macroeconomic models to process the data and produce valuable services. Businesses were eager to have forecasts of aggregates like gross domestic product, and even more eager for forecasts of narrowly defined components that were especially relevant for their particular firms. Many government policymakers were also enthusiastic at the prospect of obtaining forecasts that quantified the most likely effects of policy actions.

In the 1960s large Keynesian macroeconomic models seemed to be natural tools for meeting the demand for macroeconomic forecasts. Tinbergen (1939) had laid much of the statistical groundwork, and Klein (1950) built an early prototype Keynesian econometric model with 16 equations. By the end of the 1960s there were several competing models, each with hundreds of equations. A few prominent economists questioned the logical foundations of these models, however, and macroeconomic events of the 1970s intensified their concerns. At the time, some economists tried to improve the existing large macroeconomic models, but others argued for altogether different approaches. For example, Sims (1980) first criticized several important aspects of the large models and then suggested using vector autoregressive (VAR) models for macroeconomic forecasting. While many economists today use VAR models, many others continue to forecast with traditional macroeconomic models.

Roy.Webb@rich.frb.org. The views expressed are the author's and not necessarily those of the Federal Reserve Bank of Richmond or the Federal Reserve System.

This article first describes in more detail the traditional and VAR approaches to forecasting. It then examines why both forecasting methods continue to be used. Briefly, each approach has its own strengths and weaknesses, and even the best practice forecast is inevitably less precise than consumers would like. This acknowledged imprecision of forecasts can be frustrating, since forecasts are necessary for making decisions, and the alternative to a formal forecast is an informal one that is subject to unexamined pitfalls and is thus more likely to prove inaccurate.

1. TRADITIONAL LARGE MACROECONOMIC MODELS

These models are often referred to as Keynesian since their basic design takes as given the idea that prices fail to clear markets, at least in the short run. In accord with that general principle, their exact specification can be thought of as an elaboration of the textbook IS-LM model augmented with a Phillips curve. A simple version of an empirical Keynesian model is given below:

$$C_t = \alpha_1 + \beta_{11}(Y_t - T_t) + \varepsilon_{1,t} \tag{1}$$

$$I_t = \alpha_2 + \beta_{21}(R_t - \pi^e_{t+1}) + \varepsilon_{2,t} \tag{2}$$

$$M_t = \alpha_3 + \beta_{31} Y_t + \beta_{32} R_t + \varepsilon_{3,t} \tag{3}$$

$$\pi_t = \alpha_4 + \beta_{41} \frac{Y_t}{Y^p_t} + \varepsilon_{4,t} \tag{4}$$

$$\pi^e_{t+1} = \theta_{51}\pi_t + \theta_{52}\pi_{t-1} \tag{5}$$

$$Y \equiv C_t + I_t + G_t. \tag{6}$$

Equation (1) is the consumption function, in which real consumer spending C depends on real disposable income $Y - T$. In equation (2), business investment spending I is determined by the real interest rate $R - \pi^e$. Equation (3) represents real money demand M, which is determined by real GDP Y and the nominal interest rate R.[1] In equation (4), inflation is determined by GDP relative to potential GDP Y^p; in this simple model, this equation plays the role of the Phillips curve.[2] And in equation (5), expected inflation π^e during the

[1] The same letter is used for GDP and personal income since in this simple model there are no elements such as depreciation or indirect business taxes that prevent gross national product from equaling national income or personal income.

[2] In this article the role of the Phillips curve is to empirically relate the inflation rate and a measure of slack in the economy. In a typical large Keynesian model, the Phillips curve would be an equation that relates wage growth to the unemployment rate, with an additional equation that relates wage growth to price changes and another relating the unemployment rate to GDP relative to potential.

next period is assumed to be a simple weighted average of current inflation and the previous period's inflation. Equation (6) is the identity that defines real GDP as the sum of consumer spending, investment spending, and government spending *G*. In the stochastic equations, ε is an error term and α and β are coefficients that can be estimated from macro data, usually by ordinary least squares regressions. The Θ coefficients in equation (5) are assumed rather than estimated.[3]

One can easily imagine more elaborate versions of this model. Each major aggregate can be divided several times. Thus consumption could be divided into spending on durables, nondurables, and services, and spending on durables could be further divided into purchases of autos, home appliances, and other items. Also, in large models there would be equations that describe areas omitted from the simple model above, such as imports, exports, labor demand, and wages. None of these additions changes the basic character of the Keynesian model.

To use the model for forecasting, one must first estimate the model's coefficients, usually by ordinary least squares. In practice, estimating the model as written would not produce satisfactory results. This could be seen in several ways, such as low R^2 statistics for several equations, indicating that the model fits the data poorly. There is an easy way to raise the statistics describing the model's fit, however. Most macroeconomic data series in the United States are strongly serially correlated, so simply including one or more lags of the dependent variable in each equation will substantially boost the reported R^2 values. For example, estimating equation (2) above from 1983Q1 through 1998Q4 yields an R^2 of 0.02, but adding the lagged dependent variable raised it to 0.97. What has happened is that investment has grown with the size of the economy. The inclusion of any variable with an upward trend will raise the reported R^2 statistic. The lagged dependent variable is a convenient example of a variable with an upward trend, but many other variables could serve equally well. This example illustrates that simply looking at the statistical fit of an equation may not be informative, and economists now understand that other means are necessary to evaluate an empirical equation or model. At the time the Keynesian models were being developed, however, this point was often not appreciated.

Once the model's coefficients have been estimated, a forecaster would need future time paths for the model's exogenous variables. In this case the exogenous variables are those determined by government policy—*G*, *T*, and *M*—and potential GDP, which is determined outside the model by technology. And although the money supply is ultimately determined by monetary policy,

[3] The coefficients are assumed, rather than estimated, due to the problematic nature of existing data on actual expectations of inflation.

the Federal Reserve's policy actions immediately affect the federal funds rate. Thus rather than specifying a time path for the money supply, analysts would estimate the money demand equation and then rearrange the terms in order to put the interest rate on the left side. The future time path for short-term interest rates then became a key input into the forecasting process, although its source was rarely well documented.

Next, one could combine the estimated model with the recent data for endogenous variables and future time paths for exogenous variables and produce a forecast. With most large Keynesian models that initial forecast would require modification.[4] The reason for modifying the forecast is to factor in information that was not included in the model. For example, suppose that the model predicted weak consumer spending for the current quarter, but an analyst knew that retail sales grew rapidly in the first two months of the quarter. Or suppose that the analyst observes that consumer spending had been more robust than the model had predicted for the last several quarters. Also, the model's forecast might display some other property that the analyst did not believe, such as a continuously falling ratio of consumer spending to GDP. These are all examples of information that could lead an analyst to raise the forecast for consumer spending above the model's prediction. To change the forecast an analyst would use "add factors," which are additions to the constant terms in the equations above. Thus if one wanted to boost predicted consumer spending by $100 billion in a particular quarter, the analyst would add that amount to the constant term for that quarter. In the model given above, there are four constant terms represented by the α coefficients. To forecast ahead eight quarters, one could consider 32 possible add factors that could modify the forecast. Add factors have long been a key part of the process that uses Keynesian models to produce forecasts and are still important. For example, an appendix to a recent forecast by Data Resources, a leading econometric forecasting service that uses a Keynesian model, lists over 10,000 potential add factors.

2. CRITICISMS OF KEYNESIAN MODELS FOR FORECASTING

One of the most critical components of an economywide model is the linkage between nominal and real variables. The Phillips curve relation between wage or price growth and unemployment rates provided that key linkage for Keynesian macroeconomic models. The Phillips curve was discovered, however, as an empirical relationship. Thus when it was first incorporated in Keynesian

[4] Not every Keynesian model required modification, however. Fair (1971), for example, presented a model that has evolved over time but has not made use of the add factors defined below.

models, it did not have a firm theoretical foundation in the sense that it was not derived from a model of optimizing agents. Milton Friedman (1968) criticized the simple Phillips curve, similar to equation (5), at the time that it appeared to be consistent with the unemployment and inflation rates that had been observed in the 1950s and the 1960s. His concern was that the Phillips curve may at times appear to give a stable relation between the amount of slack in the economy and the inflation rate. But suppose that the Federal Reserve were to ease monetary policy in an attempt to permanently raise output above potential. The model above ignores the fact that people would eventually figure out the new policy strategy, and thus, according to Friedman's logic, an expectations formation equation such as (5) would no longer hold. In the long run, he argued, an attempt to hold output above potential would fail; expectations would fully adjust to the new policy and output would return to potential, but inflation would be permanently higher.

Friedman's verbal exposition was very influential, but it did not contain a fully specified analytical model. Using a formal model that captured Friedman's insight, Lucas (1972) introduced rational expectations to macroeconomic analysis as a key element for constructing a dynamic macro model. Among the important conclusions of that paper, he demonstrated that a Phillips curve could fit previously observed data well but would not be valid if the monetary policy process were to change. The book that contained the Lucas paper also contained several papers that presented long-run Phillips curves from leading Keynesian models; a representative result of those models was that a 4 percent rate of unemployment corresponded to 3.5 percent inflation and that higher inflation would give lower unemployment (Christ 1972).

Propositions in economics are rarely tested decisively. In this case, though, it was soon clear that the simple Phillips curve was not a stable, dependable relation. In the fourth quarter of 1972 the unemployment rate averaged 5.4 percent and consumer inflation over the previous four quarters was 3.3 percent. By the third quarter of 1975, unemployment had risen to 8.9 percent; the inflation rate, however, did not fall but instead rose to 11.0 percent.

In retrospect, one can identify many problems with the Keynesian models of that period. Some could be resolved without making wholesale change to the models. For example, most models were changed to incorporate a natural rate of unemployment in the long run, thereby removing the permanent trade-off between unemployment and inflation. Also, most large Keynesian models were expanded to add an energy sector, so that exogenous oil price changes could be factored in. But some of the criticisms called for a fundamental change in the strategy of building and using macroeconomic models.

One of the most influential was the Lucas (1976) critique. Lucas focused on the use of econometric models to predict the effects of government economic policy. Rather than thinking of individual policy actions in isolation, he defined policy to mean a strategy in which specific actions are chosen in order to achieve well-defined goals. As an example of this meaning of policy, consider the possibility that the Federal Reserve changed interest rates during the early 1960s in order to keep GDP close to potential and inflation low. That behavior could be represented as a reaction function such as equation (7):

$$R_t = R_{t-1} + \beta_{61}\frac{Y_t}{Y_t^p} + \beta_{62}\pi_t + \varepsilon_{6,t}. \tag{7}$$

Now suppose that the reaction function changed in the late 1960s and that less importance was placed on achieving a low rate of inflation. One can imagine replacing equation (7) with the new reaction function; however, Lucas argued that even with the new reaction function, a model would not give reliable policy advice. The reason is that the parameters of all the other equations reflect choices that were made when the previous policy rule was in effect. Under the new policy rule the parameters could well be significantly different in each equation above. This result is easiest to see in equation (6), which describes the formation of expectations of inflation in a manner that might be reasonable for a period when the monetary authority was stabilizing inflation. Individuals could do better, though, if the monetary policy strategy was in the process of changing substantially. During that period an analyst who wanted to produce reliable conditional forecasts would need to replace equation (6), even if the model as a whole continued to provide useful short-term forecasts of overall economic activity. As Lucas (1976, p. 20) put it, "the features which lead to success in short-term forecasting are unrelated to quantitative policy evaluation, . . . [T]he major econometric models are (well) designed to perform the former task only, and . . . simulations using these models can, in principle, provide *no* useful information as to the actual consequences of alternative economic policies."

This critique presented a difficult challenge for macroeconomic model builders. Every macroeconomic model is a simplification of a very complex economy, and the Keynesian models are no exception. One of the key elements of Keynesian models is that prices do no adjust instantaneously to equate supply and demand in every market. The reasons underlying sluggish price adjustment are not usually modeled, however. Thus the models cannot answer the question of to what extent, in response to a policy change, the sluggishness of price adjustment would change. The Lucas critique challenged the reliability of policy advice from models that could not answer such a basic question.

Analysts continue to offer policy advice based on Keynesian models and also other macroeconomic models that are subject to the Lucas critique. These analysts are in effect discounting the relevance of the possibility that their estimated coefficients could vary under the type of policy change analyzed by Lucas. For a succinct example of the reasoning that would allow the use of Keynesian models for policy analysis, consider the counterargument given by Tobin (1981, p. 392), "Lucas's famous 'critique' is a valid point . . . [but]

the critique is not so devastating that macroeconomic model-builders should immediately close up shop. The public's perception of policy regimes is not so precise as to exclude considerable room for discretionary policy moves that the public would see neither as surprises nor as signals of a systematic change in regime. Moreover, behavioral 'rules of thumb,' though not permanent, may persist long enough for the horizons of macroeconomic policy-makers." Sims (1982) gave a lengthier defense of traditional policy analysis.

Authors such as Lucas and Sargent (1979) and Sims (1980) also criticized Keynesian models for not being based on intertemporal optimizing behavior of individuals. At the time they recommended different strategies for model building. Since that time, however, there have been notable improvements in the economic theory embodied in Keynesian models. For example, in the Federal Reserve Board's FRB/US model, it is possible to simulate the model under the assumption that the expectations of individuals are the same as the entire model's forecasts (Brayton et al. 1997). And many modelers have successfully derived individual equations from optimizing dynamic models. Still, Keynesian models continue to be based on unmodeled frictions such as sluggish price adjustment. It is therefore not surprising that economists have explored alternative methods of forecasting and policy analysis. One important method was proposed by Sims (1980) and is discussed in the next section.

3. VAR MODELS

VAR models offer a very simple method of generating forecasts. Consider the simplest reasonable forecast imaginable, extrapolating the recent past. In practice, a reasonably accurate forecast for many data series from the United States over the past half century can be made by simply predicting that the growth rate observed in the previous period will continue unchanged. One could do better, though, by substituting a weighted average of recent growth rates for the single period last observed. That weighted average would be an autoregressive (AR) forecast, and these are often used by economists, at least as benchmarks. Only slightly more complicated is the idea that, instead of thinking of an autoregressive forecast of a single variable, one could imagine an autoregressive forecast of a vector of variables. The advantage of such a VAR relative to simpler alternatives would be that it allowed for the possibility of multivariate interaction. The simplest possible VAR is given below in equations (8) and (9), with only two variables and only one lagged value used for each variable; one can easily imagine using longer lag lengths and more variables:

$$R_t = a_{11}R_{t-1} + a_{12}p_{t-1} + u_{1,t} \tag{8}$$

$$p_t = a_{21}R_{t-1} + a_{22}p_{t-1} + u_{2,t}. \tag{9}$$

Because of the extreme simplicity of the VAR model, it may seem unlikely to produce accurate forecasts. Robert Litterman (1986), however, issued a series of forecasts from small VAR models that incorporated from six to eight variables. The results, summarized in Table 1, are root mean squared errors (RMSEs), that is, $e = \sqrt{\sum_t (A_t - P_t)^2}$, where e is the RMSE, A is the actual value of a macroeconomic variable, and P is the predicted value. One caveat is that the data summarized in this table cover a relatively short time period, and thus it is a statistically small sample. Over that period, in comparison with forecasts from services using large Keynesian models, the VAR forecasts were more accurate for real GNP more than one quarter ahead, less accurate for inflation, and of comparable accuracy for nominal GNP and the interest rate.

In another study, Lupoletti and Webb (1986) also compared VAR forecasts to those of commercial forecasting services over a longer time period than in the previous comparison. A different caveat applies to their results, shown in Table 2. They studied simulated forecasts versus actual forecasts from the forecasting services. While the details[5] of the simple model were not varied to obtain more accurate forecasts, it is inevitable in such studies that if the VAR forecasts had been significantly less accurate, then the results probably would not have seemed novel enough to warrant publication. That said, their five-variable VAR model produced forecasts that, for four and six quarters ahead, were of comparable accuracy to those of the commercial forecasting services. The commercial services predicted real and nominal GNP significantly more accurately for one and two quarters ahead, which probably indicates the advantage of incorporating current data into a forecast by using add factors.[6]

The VAR model studied by Lupoletti and Webb has five variables, each with six lags. With a constant term, each equation contains 31 coefficients to be estimated—a large number relative to the length of postwar U.S. time series. Although there are methods to reduce the effective number of coefficients that need to be estimated, the number of coefficients still rises rapidly as the number of variables is increased. Thus as a practical matter, any VAR model will contain only a fairly small number of variables. As a result, a VAR model will always ignore potentially valuable data. How, then, is it possible for them

[5] In this case the authors could have changed the start date of the regressions used to estimate the VAR model's coefficients, the choice of variables (monetary base versus M1 or M2, for example), or the number of lag lengths. In addition, this model was unrestricted, whereas most VAR forecasters use restrictions to reduce the effective number of estimated coefficients; experimenting with methods of restricting parameters would have lowered the average errors of the VAR forecasts.

[6] For example, an analyst might note that labor input, measured as employee hours, was increasing rapidly in a quarter in which GDP was forecast to rise slowly. The unexpected increase in employee hours could indicate that labor demand had risen due to unexpectedly rapid GDP growth. If other data were consistent with that line of reasoning, the analyst would then increase the constant terms in the equations determining GDP for the current quarter and quite possibly the next quarter as well. Since statistical agencies release important new data every week, there are many such opportunities for skilled analysts to improve forecast accuracy by informally incorporating the latest data.

Table 1 Average Forecast Errors from Forecasts Made in the Early 1980s

Variable: Forecast Horizon (Quarters)	Chase	DRI	WEFA	BVAR
Real GNP:				
1	2.4	2.0	3.1	2.8
2	2.6	2.3	2.6	2.1
4	2.7	2.5	2.4	1.9
8	2.0	2.0	1.7	1.3
GNP deflator:				
1	1.4	1.4	1.9	2.5
2	1.0	1.1	1.5	2.5
4	1.4	1.4	1.7	3.3
8	2.2	2.2	2.4	4.1
Nominal GNP:				
1	3.2	2.7	3.7	3.6
2	3.2	2.7	3.6	3.3
4	3.6	3.2	3.8	4.0
8	3.6	3.6	2.4	3.5
Treasury bill rate:				
1	0.2	0.1	0.4	0.1
2	1.8	1.9	1.8	1.7
4	3.3	3.2	3.2	2.9
8	2.9	3.7	1.1	3.7

Notes: Data are root mean squared errors (RMSEs) from postsample forecasts. Forecasts are from 1980Q2 to 1985Q1. Forecasts of real GNP, the GNP deflator, and nominal GNP are percentage changes from the previous quarter, and forecasts of the Treasury bill rate are cumulative changes in the quarterly average level. Data are from McNees (1986). Forecasts from WEFA were made in mid-quarter, and the others were made one month later.

to produce relatively accurate forecasts? One possibility is that there is only a limited amount of information in *all* macroeconomic time series that is relevant for forecasting broad aggregates like GDP or its price index and that a shrewdly chosen VAR model can capture much of that information.

At best, then, a VAR model is a satisfactory approximation to an underlying structure that would be better approximated by a larger, more complex model. That more complex model would include how government policymakers respond to economic events. The VAR approximation will be based on the average response over a particular sample period. A forecast from a VAR model will thus be an unconditional forecast in that it is not conditioned on any particular sequence of policy actions but rather on the average behavior of policymakers observed in the past. A forecast from a Keynesian model,

Table 2 Average Forecast Errors from Simulated Forecasts

Variable: Forecast Horizon (Quarters)	Chase	DRI	WEFA	VAR
Real GNP:				
1	4.1	4.0	4.2	5.3
2	3.1	3.1	2.9	4.1
4	2.5	2.5	2.2	2.8
6	2.3	2.3	1.9	2.4
GNP deflator:				
1	1.8	2.0	1.9	2.3
2	1.9	2.0	1.9	2.0
4	2.2	2.1	2.0	2.1
6	2.5	2.4	2.2	2.4
Nominal GNP:				
1	5.1	4.6	4.9	6.0
2	4.1	3.6	3.8	4.7
4	3.5	3.0	3.0	3.3
6	3.3	2.7	2.6	3.1
Treasury bill rate:				
1	1.5	1.4	. . .	1.3
2	2.2	2.1	. . .	2.1
4	2.9	2.6	. . .	2.8
6	3.5	3.2	. . .	3.5

Notes: Data are root mean squared errors (RMSEs) from postsample forecasts. Ranges for RMSEs are: one-quarter forecasts, 1970:4–1983:4; two-quarter forecasts, 1971:1–1983:4; four-quarter forecasts, 1971:3–1983:4; and six-quarter forecasts, 1972:1–1983:4. The VAR forecasts are simulated forecasts, as described in the text. Forecasts of real GNP, the GNP deflator, and nominal GNP are cumulative percentage changes, and forecasts of the Treasury bill rate are for its quarterly average level.

however, usually is based on a particular sequence of policy actions and is referred to as a conditional forecast—that is, conditional on that particular sequence. Despite the Lucas critique, many users of Keynesian models seek to determine the consequences of possible policy actions by simulating their model with different time paths of policy actions. But, although the Lucas critique was discussed above in reference to Keynesian models, it is equally valid for VAR models. To help emphasize this point, the next section reviews some details of using a VAR model for forecasting.

Forecasting with VAR Models

To forecast with the VAR model summarized in equations (8) and (9), one would estimate the a_{ij} coefficients, usually by ordinary least squares, and calculate period t values based on data for period $t-1$. One can then use the period t forecasts to calculate forecasts for period $t+1$; for example, inflation forecasts in the above model would be

$$\hat{p}_{t+1} = a_{21}\hat{R}_t + a_{22}\hat{p}_t + \hat{u}_{2,t+1}$$

$$= (a_{21}a_{11} + a_{22}a_{21})R_{t-1} + (a_{21}a_{12} + a_{22}^2)p_{t-1} + a_{21}\hat{u}_{1,t} + a_{22}\hat{u}_{2,t} + \hat{u}_{2,t+1}, \quad (10)$$

where the second line in (10) was obtained by taking the first line and substituting the right-hand sides of (8) and (9) for the estimated values of R_t and p_t, respectively. The above procedure can be repeated as many times as needed to produce as long a forecast as desired.

It is often assumed that the realizations of unknown error terms—$u_{1,t}, u_{2,t}$, and $u_{2,t+1}$—will all equal zero. One can discard that assumption to incorporate information that was not used to estimate the model. Suppose the above model uses monthly data, and at the beginning of a month one knows last month's average interest rate but not the inflation rate, which the Labor Department will release two weeks later. One could simply substitute the realized interest rate for the estimated rate in the calculations above; in equation (10) that would mean plugging in the realized value of $u_{1,t}$. Since the errors in a VAR are usually contemporaneously correlated, a realization of $u_{1,t}$ will also provide information about $u_{2,t}$. Specifically, the variances and covariances of the error terms are taken from the variance-covariance matrix that was estimated through period $t-1$ when the a_{ij} coefficients were estimated; the expected value of $u_{2,t}$ is then the ratio of the estimated covariance of u_1 and u_2 to the estimated variance of u_1 times the realization of $u_{1,t}$. This expected value of $u_{2,t}$ can then also be included in equations (8) and (9) in order to forecast inflation in periods t and $t+1$. One can easily apply this basic method for forecasting with a VAR, and the refinement for incorporating partial data for a period, to more complicated models with longer lags, more variables, and deterministic terms such as constants, time trends, and dummy variables.

With this background in mind, imagine that the true structure of the economy is given by the Keynesian model of equations (1) through (6) along with the monetary reaction function (7). Now suppose that the VAR model represented by equations (8) and (9) is estimated. Algebraic manipulation[7] yields the estimated coefficients of the VAR model as functions of the underlying structural coefficients and error terms in equations (8′) and (9′):

$$\pi_t = B_{1,t} + A_{11}\pi_{t-1} + A_{12}R_{t-1} + U_{1,t} \quad (8')$$

[7] In brief, substitute equations (1) and (2) into (6), solve for Y, then substitute the resulting expression for Y into equation (3), and rearrange terms so that π_t is on the left. Next, solve equations (4) and (7) for Y/Y^p, equate the resulting expressions, and rearrange terms so that R_t is on the left. The resulting two equations for π_t and R_t can be solved for each variable as an expression containing lagged values of π and R, exogenous variables, structural error terms, and underlying structural coefficients.

$$R_t = B_{2,t} + A_{21}\pi_{t-1} + A_{22}R_{t-1} + U_{2,t}, \tag{9$'$}$$

where

$$B_{1,t} = [(\alpha_1 + \alpha_2 - \beta_{11}T_t + G_t + (1-\beta_{11})(\alpha_3 - M_t))/\beta_{21}\theta_{51} + (\beta_{21} - \alpha_4\frac{\beta_{61}}{\beta_{41}})]/\delta$$

$$A_{11} = -\frac{\theta_{52}\pi_{t-1}}{\theta_{51}\delta}$$

$$A_{12} = \frac{\beta_{21} + (1-\beta_{11})\beta_{32}}{\beta_{21}\delta}$$

$$U_{1,t} = [e_{1,t} + e_{2,t} + (1-\beta_{11})e_{3,t}]/\beta_{21}\theta_{51}\delta$$

$$B_{2,t} = [\alpha_1 + \alpha_2 + \alpha_3(1-\beta_{11}) + \frac{\beta_{61}}{\beta_{41}}\alpha_4 + G_t - T_t - (1-\beta_{11})M_t]\frac{\beta_{41}\beta_{62} + \beta_{61}}{\beta_{21}\beta_{41}\theta_{51}\delta}$$

$$A_{21} = -\frac{\beta_{21}\theta_{51}(\beta_{41}\beta_{62} + \beta_{61})}{\beta_{21}\beta_{41}\theta_{51}\delta}$$

$$A_{22} = \frac{1}{\delta}$$

$$U_{2,t} = [\varepsilon_{1,t} + \varepsilon_{2,t} + \varepsilon_{3,t}(1-\beta_{11})]\frac{\beta_{41}\beta_{62} + \beta_{61}}{\beta_{21}\beta_{41}\theta_{51}\delta} + \varepsilon_{4,t}\frac{\beta_{61}}{\beta_{41}\delta} + \varepsilon_{5,t}\delta^{-1}$$

$$\delta = (1 - \frac{\beta_{21} + (1-\beta_{11})\beta_{32}}{\beta_{21}})(\beta_{62} + \frac{\beta_{61}}{\beta_{41}}).$$

Viewing the model as equations (8$'$) and (9$'$) reveals the problematic nature of conditional forecasting with the model. Suppose an analyst wishes to study the effect of a tighter monetary policy on the inflation rate by first obtaining a baseline forecast from the VAR model and then raising the interest rate prediction by a full percentage point for the next quarter. This step would be accomplished by feeding in a particular nonzero value for $u_{2,t+1}$ in equation (10). However, note that in terms of the underlying structure, the error term $U_{2,t}$ is a complicated composite of the five error terms from the equations of the underlying model. Yet for policy analysis it would be necessary to identify that composite error term as a monetary policy disturbance.[8]

An identification that ignores the distinction between VAR errors, the $u_{i,t}$s, and the underlying structural errors, such as the $\varepsilon_{j,t}$'s in the example above, can lead to absurd results. Suppose one simulates a tighter monetary policy in the model presented above by forcing the VAR model to predict higher interest rates; the outcome is a *higher* inflation prediction. The reason is that,

[8] This point is not new—see Cooley and LeRoy (1985).

in the quarterly macroeconomic time series of the last 50 years, the dominant shocks to interest rates and inflation have been aggregate demand shocks, and a positive aggregate demand shock raises interest rates, inflation, output, and employment. The VAR model captures these correlations. Asking the model to simulate a higher interest rate path will lead it to predict a higher inflation path as well. Now a clever user can tinker with the model—adding variables, changing the dates over which the model was estimated, and so forth—and eventually develop a VAR model that yields a lower inflation path in response to higher interest rates. At this point, though, the model would add little value beyond reflecting the user's prior beliefs.

To recap, VAR models are unsuited to conditional forecasting because a VAR residual tends to be such a hodgepodge. In addition, the models are vulnerable to the Lucas critique. Suppose that the monetary authority decided to put a higher weight on its inflation target and a lower weight on its output target and that its new reaction function could be represented by (7′):

$$R_t = R_{t-1} + (\beta_{61} - \phi)\frac{Y_t}{Y_t^p} + (\beta_{62} + \phi)\pi_t + \varepsilon_{6,t}. \tag{7'}$$

The interpretation of the VAR's coefficients in terms of the underlying structural coefficients would also change, with each instance of β_{61} changing to $\beta_{61} - \phi$ and each instance of β_{62} changing to $\beta_{62} + \phi$. Thus following a discrete change in the monetary strategy, the VAR's coefficients would be systematically biased and even the accuracy of its unconditional forecasts would be compromised.

Some authors, including Sims (1982), have questioned whether large policy changes in the United States have resulted in meaningful parameter instability in reduced forms such as VARs. One of the most dramatic changes in estimated coefficients in VAR equations for U.S. data occurred in an inflation equation. Table 3 is reproduced from Webb (1995) and shows significant changes in an inflation equation's coefficients estimated in different subperiods.[9] The subperiods, moreover, were determined by the author's review of minutes of the Federal Open Market Committee in order to find monetary policy actions that could indicate a discrete change in the monetary strategy. The results are thus consistent with the view that the monetary reaction function changed substantially in the mid-1960s and again in the early 1980s and that the changes in the economic structure played havoc with a VAR price equation's coefficients.

This section has thus presented two separate reasons for distrusting conditional forecasts from VAR models. First, their small size guarantees that residuals will be complicated amalgamations, and no single residual can be meaningfully interpreted as solely resulting from a policy action. Second,

[9] Consider, for example, the sum of coefficients on the nominal variables—inflation, the monetary base, and the nominal interest rate. In the early period the sum is 0.17, rising to 1.23 in the middle period, and then falling to 0.80 in the final period.

Table 3 Regression Results for Several Time Periods

1952Q2 to 1966Q4 $\bar{R}^2 = -0.08$

$$\hat{p}_t = \underset{(0.06)}{0.28} - \underset{(-0.54)}{0.08}p_{t-1} + \underset{(0.51)}{0.08}p_{t-2} + \underset{(0.72)}{0.11}p_{t-3} + \underset{(0.24)}{0.07}r_{t-1} + \underset{(0.28)}{0.02}c_{t-1} - \underset{(-0.01)}{0.01}m_{t-1} + \underset{(0.76)}{0.05}y_{t-1}$$

1967Q1 to 1981Q2 $\bar{R}^2 = 0.57$

$$\hat{p}_t = \underset{(-0.54)}{-2.78} + \underset{(2.30)}{0.30}p_{t-1} - \underset{(-0.28)}{0.04}p_{t-2} + \underset{(0.27)}{0.04}p_{t-3} + \underset{(2.56)}{0.33}r_{t-1} + \underset{(0.26)}{0.02}c_{t-1} + \underset{(4.87)}{0.60}m_{t-1} - \underset{(-1.52)}{0.08}y_{t-1}$$

1981Q3 to 1990Q4 $\bar{R}^2 = 0.51$

$$\hat{p}_t = \underset{(-1.54)}{-8.87} + \underset{(1.16)}{0.21}p_{t-1} + \underset{(0.53)}{0.09}p_{t-2} + \underset{(1.15)}{0.20}p_{t-3} + \underset{(1.07)}{0.20}r_{t-1} + \underset{(1.68)}{0.10}c_{t-1} + \underset{(1.02)}{0.10}m_{t-1} - \underset{(-0.21)}{0.15}y_{t-1}$$

1952Q2 to 1990Q4 $\bar{R}^2 = 0.59$

$$\hat{p}_t = \underset{(-1.42)}{-3.84} + \underset{(6.38)}{0.30}p_{t-1} + \underset{(2.89)}{0.23}p_{t-2} + \underset{(2.71)}{0.22}p_{t-3} + \underset{(0.07)}{0.005}r_{t-1} + \underset{(1.54)}{0.05}c_{t-1} + \underset{(2.95)}{0.17}m_{t-1} - \underset{(-0.59)}{0.22}y_{t-1}$$

Note: Coefficients were estimated by ordinary least squares; t-statistics are in parentheses.

applying the Lucas critique to VAR models implies that a VAR model's coefficients would be expected to change in response to a discrete policy change.

Several researchers who have recognized these deficiencies but were unwilling to give up the simplicity of the VAR approach have turned to structural VARs, or SVARs.[10] These models attempt to apply both economic theory that is often loosely specified and statistical assumptions to a VAR in order to interpret the residuals and conduct meaningful policy analysis. In many studies key statistical assumptions are that the economy is accurately described by a small number of equations containing stochastic error terms, and that these structural errors are uncorrelated across equations. The economic restrictions vary considerably from model to model; the common feature is that just enough restrictions are introduced so that the reduced-form errors, such as in equations (7′) and (8′) above, can be used to estimate the structural errors. For example, two of the restrictions used in a widely cited paper by Blanchard (1989) were (1) that reduced-form GDP errors were equal to structural aggregate demand errors, and (2) that reduced-form unemployment errors, given output, were equal to structural supply errors. After presenting those and other restrictions, the author noted "There is an obvious arbitrariness to any set of identification restrictions, and the discussion above is no exception" (p. 1150).

[10] A clear exposition of the SVAR approach is given by Sarte (1999).

It is often the case that a reader will find an identifying assumption of an SVAR somewhat questionable. A major difficulty of the SVAR approach is that there is no empirical method for testing a restriction. Moreover, if different models give different results, there are no accepted performance measures that can be used to identify superior performance. Since there are millions of possible SVARS that could be based on the last half century of U.S. macroeconomic data, their results will not be persuasive to a wide audience until a method is found to separate the best models from the rest.[11]

4. FINAL THOUGHTS ON CONDITIONAL FORECASTING

This article has discussed two approaches to macroeconomic forecasting. Both approaches have produced econometric models that fit observed data reasonably well, and both have produced fairly accurate unconditional forecasts. The VAR approach was found unsuitable for conditional forecasting and policy analysis. There is a wide division within the economics profession on the usefulness of large Keynesian models for policy analysis. At one extreme are those who accept the Lucas critique as a fatal blow and accordingly see little value in using Keynesian models for policy analysis. At the other extreme are analysts who are comfortable with traditional Keynesian models. In the middle are many economists with some degree of discomfort at using the existing Keynesian models, in part due to the features that allow the models to fit the historical data well but may not remain valid in the event of a significant policy change. But policy analysis will continue, formally or informally, regardless of economists' comfort with the models and with the strategies for using them. Decisions on the setting of policy instruments will continue to be made and will be based on some type of analysis.

One possibility is that policy analysis and economic forecasting will be seen as two different problems requiring two different types of models. Economists have constructed a large number of small structural models that can be quantified and used for policy analysis. A large number of statistical approaches to forecasting are available as well. It is not necessary that the same model be used for both.

Keynesian models, though, are still widely used for policy analysis, and there are actions that model builders could take to enhance the persuasiveness of their results. One would be to publish two forecasts on a routine basis—the usual forecast with add factors incorporating the modelers' judgment and a mechanical forecast with no add factors. In that way a user could easily distinguish the judgmental content from the pure model forecast. For example,

[11] Other authors have argued that SVAR results are not robust, including Cooley and Dwyer (1998) and Cecchetti and Rich (1999).

if one wanted to determine the possible impact of a tax cut on consumption, one would want to consider whether large add factors in a consumption equation such as equation (1) above were needed to achieve satisfactory results.

It would also be helpful for forecast consumers to know how much a model's specification has changed over time. Of course one hopes that new developments in economics are incorporated into models and that poorly performing specifications are discarded. As a result, some specification changes are to be expected. But if one saw that the consumption equation of a large model had been in a state of flux for several years, the numerous changes could signify that the model's analysis of a tax cut's effect on consumption was based on an unstable foundation.

In addition, it would be helpful to see more analysis of forecast errors. At a minimum, each forecast should be accompanied by confidence intervals for the most important variables stating the likely range of results. As the ex post errors indicate in Tables 1 and 2, these confidence intervals could be quite wide. For example, real GDP growth has averaged 2.8 percent over the last 30 years. In Table 2, the RMSE for four-quarter predictions of real growth from the best commercial forecasting service was 2.2 percent. Thus if a model predicted real growth to be the 2.8 percent average, and one used that RMSE as an approximate standard deviation of future forecast errors, then one would expect actual outcomes to be outside of a wide 0.6 to 5.0 percent range about 30 percent of the time. Now suppose that an exercise in policy analysis with that model revealed a difference of 1.0 percent for real GDP growth over the next year; a user might not consider that difference very meaningful, given the relatively large imprecision of the model's GDP forecast.

Finally, it would be especially helpful to have a detailed analysis of errors in a manner relevant for policy analysis. For example, continuing with the predicted effect of a tax cut, the model's predictions could be stated in the form of a multiplier that related the tax cut to a predicted change in real growth. That multiplier would be a random variable that could be statistically analyzed in the context of the whole model, and the user could be told the sampling distribution of that statistic. Also, one would want data on how well the model predicted the effects of tax cuts that had actually occurred in the past.

The unifying theme of these recommendations is for model builders to open the black box that generates forecasts. Until this supplementary information routinely accompanies the output of large forecasting models, many will see an exercise in policy evaluation as having unknowable properties and value it accordingly.

REFERENCES

Blanchard, Olivier Jean. "A Traditional Interpretation of Macroeconomic Fluctuations," *American Economic Review*, vol. 79 (December 1989), pp. 1146–64.

Brayton, Flint, Eileen Mauskopf, David Reifschneider, Peter Tinsley, and John Williams. "The Role of Expectations in the FRB/US Macroeconomic Model," *Federal Reserve Bulletin*, vol. 83 (April 1997), pp. 227–45.

Cecchetti, Stephen G., and Robert W. Rich. "Structural Estimates of the U.S. Sacrifice Ratio," Federal Reserve Bank of New York, 1999.

Christ, Carl. "Discussion," in *The Econometrics of Price Determination*. Washington: Board of Governors of the Federal Reserve System, 1972, pp. 330–31.

Cooley, Thomas F., and Mark Dwyer. "Business Cycle Analysis without Much Theory: A Look at Structural VARs," *Journal of Econometrics*, vol. 83 (March–April 1998), pp. 57–88.

__________, and Stephan F. LeRoy. "Atheoretical Macroeconomics: A Critique," *Journal of Monetary Economics*, vol. 16 (November 1985), pp. 283–308.

Fair, Ray. *A Short-Run Forecasting Model of the United States Economy.* Lexington, Mass.: D.C. Heath, 1971.

Friedman, Milton. "The Role of Monetary Policy," *American Economic Review*, vol. 58 (March 1968), pp. 1–17.

Klein, Lawrence. *Economic Fluctuations in the United States 1921–1941*. New York: Wiley, 1950.

Litterman, Robert. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience," *Journal of Business and Economic Statistics*, vol. 4 (January 1986), pp. 25–38.

Lucas, Robert. "Econometric Policy Evaluation: A Critique," in Karl Brunner and Allan H. Meltzer, eds., *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, vol. 1 (1976), pp. 19–46.

__________. "Econometric Testing of the Natural Rate Hypothesis," in *The Econometrics of Price Determination*. Washington: Board of Governors of the Federal Reserve System, 1972, pp. 50–59.

__________, and Thomas Sargent. "After Keynesian Macroeconomics," Federal Reserve Bank of Minneapolis *Quarterly Review*, vol. 3 (Spring 1979), pp. 1–16.

Lupoletti, William, and Roy Webb. "Defining and Improving the Accuracy of Macroeconomic Forecasts: Contributions from a VAR Model," *Journal of Business*, vol. 59 (April 1986), pp. 263–85.

McNees, Stephen. "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts," *Journal of Business and Economic Statistics*, vol. 4 (January 1986), pp. 5–16.

Sarte, Pierre-Daniel G. "An Empirical Investigation of Fluctuations in Manufacturing Sales and Inventory within a Sticky-Price Framework," Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 85 (Summer 1999), pp. 61–84.

Sims, Christopher A. "Policy Analysis with Econometric Models," *Brookings Papers on Economic Activity*, 1:1982, pp. 107–165.

__________. "Macroeconomics and Reality," *Econometrica*, vol. 48 (January 1980), pp. 1–48.

Tinbergen, J. *Business Cycles in the United States: 1912–1932*. League of Nations, 1939.

Tobin, James. "Comments," in J. Kmenta and J. B. Ramsey, eds., *Large-Scale Macro-Econometric Models*. New York: North-Holland, 1981, pp. 391–92.

Webb, Roy. "Forecasts of Inflation from VAR Models," *Journal of Forecasting*, vol. 14 (May 1995), pp. 267–85.

The Importance of Systematic Monetary Policy for Economic Activity

Michael Dotsey

How the Federal Reserve reacts to economic activity has significant implications for the way the economy responds to various shocks. Yet the importance of these responses has received limited attention in the economic literature. Much of the literature devoted to the economic effects of monetary policy concentrates on the impact of random monetary policy shocks. By contrast, this article analyzes the effects of the systematic, or predictable, portion of policy. Specifically, I compare how different specifications of an interest rate rule affect a model economy's response to a technology shock and a monetary policy shock. In the case of a technology shock, the central bank's adjustment of the interest rate is totally an endogenous response to economic events. The experiments show that, when there are significant linkages between real and nominal variables, the economy's response to changes in technology depends on the behavior of the monetary authority. With a monetary policy shock—for example, an unexpected change in the interest rate—the effects of that shock will depend on how the central bank subsequently reacts to changes in inflation and output. In general, the way shocks propagate through the economic system is intimately linked to the systematic behavior of the monetary authority.

The results of the experiments have a number of significant implications. Most importantly, the specification of the interest rate rule, which dictates how the monetary authority moves the interest rate in response to inflation and real activity, fundamentally affects economic behavior. The economy's behavior

Michael.Dotsey@rich.frb.org. I wish to thank Bob Hetzel, Andreas Hornstein, Pierre Sarte, and Mark Watson for a number of useful suggestions. I have also greatly benefited from many helpful discussions with Alex Wolman. The views expressed herein are the author's and do not necessarily represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

may be very different depending on the parameters that govern how the central bank reacts to inflation and the state of the economy, as well as the degree of concern it has for interest rate smoothing. For example, the central bank's systematic behavior can alter the correlations between variables in the model.[1] This type of policy effect calls into question whether changes in a policy instrument that are the result of a changing policy emphasis can be adequately approximated as shocks to an unchanging policy rule.

The article's emphasis on the effects of systematic monetary policy places it in a long tradition dating back to Poole (1970), who discussed the implications of different types of policy rules. In that paper, and in subsequent extensions to a flexible-price rational expectations environment, the primary purpose was to compare a policy that used a monetary instrument with one that employed an interest rate instrument.[2] An outcome of that literature was that the systematic component of monetary policy was important. Of significance was the way that informational frictions interacted with monetary policy, which allowed certain types of feedback rules to improve the information content of the nominal interest rate. This sharpening of information occurred only when the nominal interest rate was determined endogenously, implying that the systematic portion mattered only when money was the instrument of policy. Futhermore, the systematic portion mattered solely in the way that it affected expectations of future policy and not because it affected the current money stock.

In other types of models, such as those of Fischer (1977) and Taylor (1980), which included nominal frictions such as sticky prices and wages, an important element was the effect that systematic monetary policy had on the economy. In short, anticipated money mattered. But in these papers monetary policy was largely depicted through changes in money, rather than interest rates.

Recently, there has been renewed interest in the effects of monetary policy when the policy instrument is more accurately depicted as the interest rate. These investigations share some of the same features of the earlier models of Fischer and Taylor in that some nominal variables, usually prices, are assumed to be sticky. That is, the price level only adjusts gradually to its long-run equilibrium value. Some notable examples of this research can be found in Batini and Haldane (1997), McCallum and Nelson (1997), Rotemberg and Woodford (1997), and Rudebusch and Svensson (1997). The concern of these papers is, however, somewhat different than the one emphasized here. They concentrate on both the welfare effects and the variability of output and inflation that are induced by different forms of interest rate rules. In this article, I instead

[1] Rotemberg and Woodford (1997) perform a detailed analysis of the effects that different parameter values have on the second moments of various variables and on whether or not their economy has a unique solution.

[2] Prominent examples of this literature are Dotsey and King (1983, 1986) and Canzoneri, Henderson, and Rogoff (1983).

emphasize the qualitatively different ways that a model economy behaves for a variety of specifications of monetary policy.

Both types of investigations are important and complementary. Economic welfare analysis is important because it is the primary concern of policy analysis. But welfare measures and variances cannot in themselves inform us whether various rules yield similar forms of behavior that are just more or less volatile, or if behavior is changed in more fundamental ways. On these key matters, the article is more in the spirit of the work of Christiano and Gust (1999) and McCallum (1999), who also investigate the differences in impulse response functions when the feedback parameters of a given policy rule are varied. Even so, they use models that are different from the one used here.

The article proceeds as follows. Section 1 sketches the underlying model that is common to the analysis. The key feature of the model is the presence of price rigidity. I also indicate how an economy with sticky or sluggishly adjusting prices behaves when the money stock is held constant, and when a policy rule that results in real business cycle behavior of real quantities is implicitly followed. The latter policy rule essentially negates the real effects of price stickiness by keeping the price level and the markup constant. This exercise provides some intuition on how the model works. Section 2 describes the form of the interest rate rules investigated. These rules derive from the work of John Taylor (1993). Under the first rule, the monetary authority responds both to deviations in lagged inflation from target and to lagged output from its steady-state value. The second rule adds a concern for smoothing the behavior of the nominal interest rate. In Section 3, I analyze the response of the model economy to a permanent increase in the level of technology. Section 4 investigates the effect of an unanticipated increase in the nominal interest rate on the economy. Section 5 concludes.

1. THE MODEL

For the purpose of this investigation, I use a framework that embeds sticky prices into a dynamic stochastic model of the economy. Under flexible prices and zero inflation, the underlying economy behaves as a classic real business cycle model. The model is, therefore, of the new neoclassical synthesis variety and displays features that are common to much of the current literature using sticky-price models.[3] Agents have preferences over consumption and leisure, and rent productive factors to firms. For convenience, money is introduced via a demand function rather than entering directly in utility or through a shopping time technology. Firms are monopolistically competitive and face a fixed

[3] Examples of this literature are Goodfriend and King (1997), and Chari, Kehoe, and McGrattan (1998).

schedule for changing prices. Specifically, one-quarter of the firms change their price each period, and each firm can change its price only once a year. This type of staggered time-dependent pricing behavior, referred to as a Taylor contract, is a common methodology for introducing price stickiness into an otherwise neoclassical model.

Consumers

Consumers maximize the following utility function:

$$U = E_0 \sum_{t=0}^{\infty} \beta^t [\ln(C_t) - \chi n_t^{\zeta}],$$

where $C = [\int_0^1 c(i)^{(\varepsilon-1)/\varepsilon} di]^{\varepsilon/(\varepsilon-1)}$ is an index of consumption and n is the fraction of time spent in employment.

Consumers also face the following intertemporal budget constraint:

$$P_t C_t + P_t K_{t+1} \leq W_t n_t + [r_t + (1-\delta)] P_t K_t + Div_t,$$

where $P = [\int_0^1 p(i)^{1-\varepsilon} di]^{1/(1-\varepsilon)}$ is the price index associated with the aggregator C; W is the nominal wage rate; r is the rental rate on capital; δ is the rate that capital depreciates; and Div are nominal dividend payments received from firms.

The relevant first-order conditions for the representative consumer's problem are given by

$$(1/C_t)(W_t/P_t) = \chi \varsigma n_t^{\varsigma - 1} \tag{1a}$$

and

$$(1/C_t) = \beta E_t (1/C_{t+1})[r_{t+1} + (1-\delta)]. \tag{1b}$$

Equation (1a) equates the marginal disutility of work with the value of additional earnings. An increase in wages implies that individuals will work harder. Equation (1b) describes the optimal savings behavior of individuals. If the return to saving (r) rises, then households will consume less today, saving more and consuming more in the future.

The demand for money, which is just assumed rather than derived from optimizing behavior, is given by

$$\ln(M_t/P_t) = \ln Y_t - \eta_R R_t, \tag{2}$$

where Y is the aggregator of goods produced in the economy and is the sum of the consumption aggregator C and an analogous investment aggregator I. The nominal interest rate is denoted R, and η_R is the interest semi-elasticity of money demand. One could derive the money demand curve from a shopping time technology without qualitatively affecting the results in the article.

Firms

There is a continuum of firms indexed by j that produce goods, $y(j)$, using a Cobb-Douglas technology that combines labor and capital according to

$$y(j) = a_t k(j)^{\alpha} n(j)^{1-\alpha}, \tag{3}$$

where a is a technology shock that is the same for all firms. Each firm rents capital and hires labor in economywide competitive factor markets. The cost-minimizing demands for each factor are given by

$$\psi_t a_t (1-\alpha)[k_t(j)/n_t(j)]^{\alpha} = W_t/P_t \tag{4a}$$

and

$$\psi_t a_t \alpha [k_t(j)/n_t(j)]^{\alpha-1} = r_t, \tag{4b}$$

where ψ is real marginal cost. The above conditions imply that capital-labor ratios are equal across firms.

Although firms are competitors in factor markets, they have some monopoly power over their own product and face downward-sloping demand curves of $y(j) = (p(j)/P)^{-\varepsilon} Y$, where $p(j)$ is the price that firm j charges for its product. This demand curve results from individuals minimizing the cost of purchasing the consumption index C and an analogous investment index. Firms are allowed to adjust their price once every four periods and choose a price that will maximize the expected value of the discounted stream of profits over that period. Specifically, a firm sets its price in period t to

$$\max_{p_t(j)} E_t \sum_{h=0}^{3} \Delta_{t+h} \phi_{t+h}(j),$$

where real profits at time $t+h$, $\phi_{t+h}(j)$, are given by $[p_t^*(j) y_{t+h}(j) - \psi_{t+h} P_{t+h} y_{t+h}(j)]/P_{t+h}$, and Δ_{t+h} is an appropriate discount factor that is related to the way in which individuals value future as opposed to current consumption.[4]

The result of this maximization is that an adjusting firm's relative price is given by

$$\frac{p_t^*(j)}{P_t} = \frac{\varepsilon}{\varepsilon - 1} \frac{\sum_{h=0}^{3} \beta^h E_t\{\Delta_{t+h}\psi_{t+h}(P_{t+h}/P_t)^{1+\varepsilon} Y_{t+h}\}}{\sum_{h=0}^{3} \beta^h E_t\{\Delta_{t+h}(P_{t+h}/P_t)^{\varepsilon} Y_{t+h}\}}. \tag{5}$$

Furthermore, the symmetric nature of the economic environment implies that all adjusting firms will choose the same price. One can see from equation (5) that in a regime of zero inflation and constant marginal costs, firms would set their relative price $p^*(j)/P$ as a constant markup over marginal cost of $\frac{\varepsilon}{\varepsilon-1}$. In general, a firm's pricing decision depends on future marginal costs, the future

[4] Specifically, the discount factor is the ratio of the marginal utility of consumption in period $t+h$ to the marginal utility of consumption in period t.

aggregate price level, future aggregate demand, and future discount rates. For example, if a firm expects marginal costs to rise in the future, or if it expects higher rates of inflation, it will choose a relatively higher current price for its product.

The aggregate price level for the economy will depend on the prices the various firms charge. Since all adjusting firms choose the same price, there will be four different prices charged for the various individual goods. Each different price is merely a function of when that price was last adjusted. The aggregate price level is, therefore, given by

$$P_t = \left[\sum_{h=0}^{3}(1/4)(p_{t-h}^{*})^{1-\varepsilon}\right]^{(1/(1-\varepsilon))}. \tag{6}$$

Steady State and Calibration

An equilibrium in this economy is a vector of prices p_{t-h}^{*}, wages, rental rates, and quantities that solves the firm's maximization problem, the consumers' optimization problem, and one in which the goods, capital, and labor markets clear. Furthermore, the pricing decisions of firms must be consistent with the aggregate pricing relationship (6) and with the behavior of the monetary authority described in the next section. Although I will look at the economy's behavior when the Fed changes its policy rule, the above description of the private sector will remain invariant across policy experiments.

The baseline steady state is solved for the following parameterization. Labor's share, $1-\alpha$, is set at 2/3, $\zeta = 9/5, \beta = 0.984, \varepsilon = 10, \delta = 0.025$, $\eta_R = 0$, and agents spend 20 percent of their time working. These parameter values imply a steady-state ratio of I/Y of 18 percent, and a value of $\chi = 18.47$. The choice of $\zeta = 9/5$ implies a labor supply elasticity of 1.25, which agrees with recent work by Mulligan (1999). A value of $\varepsilon = 10$ implies a steady-state markup of 11 percent, which is consistent with the empirical work in Basu and Fernald (1997) and Basu and Kimball (1997). The interest sensitivity of money demand is set at zero. The demand for money is generally acknowledged to be fairly interest insensitive and zero is simply the extreme case. Since the ensuing analysis concentrates on interest rate rules, the value of this parameter is unimportant.

The economy is buffeted by a technology shock modeled as a random walk and assumed to have a standard deviation of 1 percent. Thus, increases in technology have a permanent effect on the economy. This specification is consistent with the assumptions of much of the empirical work in this area.

The Model under Constant Money Growth

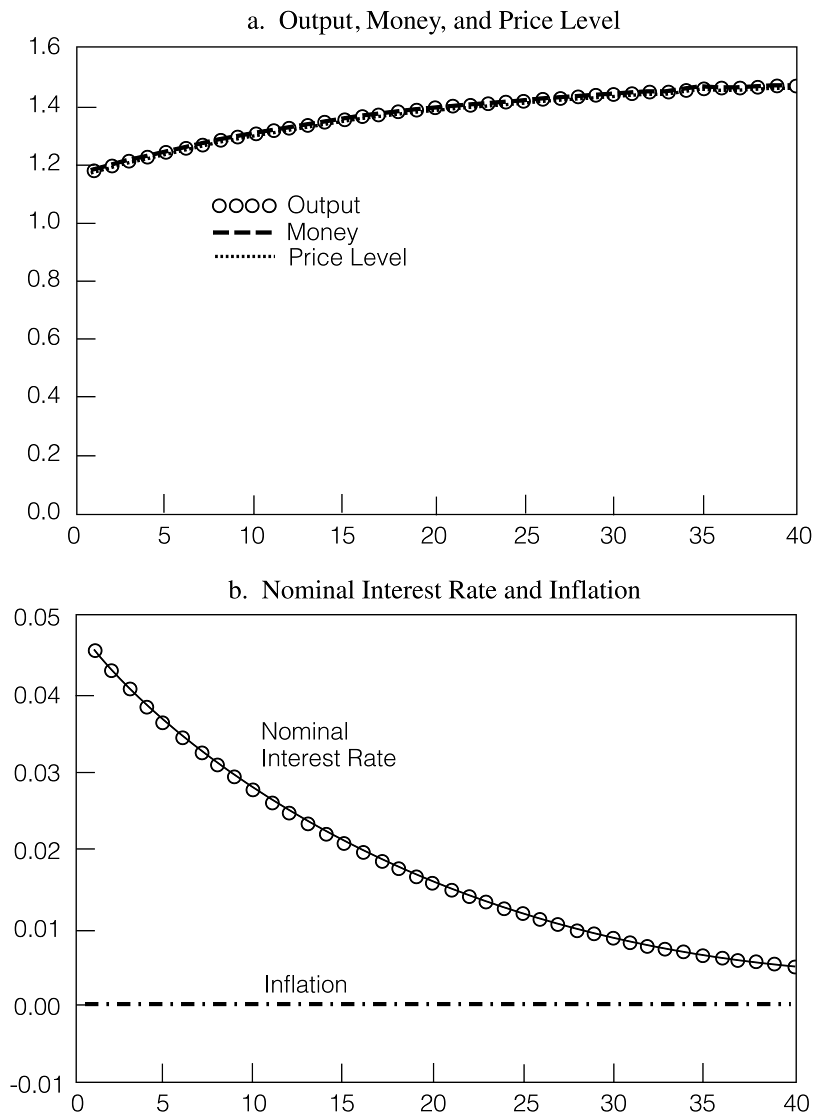
In this section, I analyze the response of the model economy to a technology shock under a constant money growth rule. As a preliminary matter, it is worthwhile to recall how a standard real business cycle (RBC) model would behave when subjected to such shocks. The behavior of real variables in the baseline RBC model is closely mimicked in this model with a rule that keeps the price markup and the inflation rate close to their steady-state values. The behavior of the economy under such a rule is of independent interest as well, because some recent work indicates that a constant markup would be a feature of optimal monetary policy (e.g., King and Wolman [1999]).

The reason this policy produces a response in real variables very much like that obtained in a model with flexible prices can be seen by examining equation (5). If prices were flexible, then each firm would choose the same price, and relative prices would equal unity. Real marginal cost would then be $\frac{\epsilon-1}{\epsilon}$, which is exactly the steady-state value of marginal cost under staggered price setting and zero inflation. If the steady-state inflation rate were zero, then stabilizing real marginal cost at its steady-state value would imply that firms would have no desire to deviate from steady-state behavior and would keep their relative price constant at one. Thus, in an environment of zero average inflation, stabilizing marginal cost or the markup leads to firm behavior that replicates what firms would do in a world of flexible prices. In short, when inflation rates are close to zero, one can find a policy that virtually replicates the behavior found in a flexible price model.

Figures 1a and 1b show the deviations of output, money stock, price level, nominal interest rate, and inflation from their steady-state values in response to a permanent increase in the level of technology under a rule that keeps the markup approximately constant (it varies by less than 0.0002 percent from its steady-state value of 0.11). Output initially jumps by 1.2 percent and then gradually increases to its new steady-state value. The money supply grows one-for-one with output, and given an income elasticity of one and an interest elasticity of zero, its behavior is consistent with prices growing at their steady-state rate of 2 percent. Consequently, inflation remains at its steady-state rate. The slight uptick in the nominal interest rate is, therefore, entirely due to a small increase in the real rate of interest.

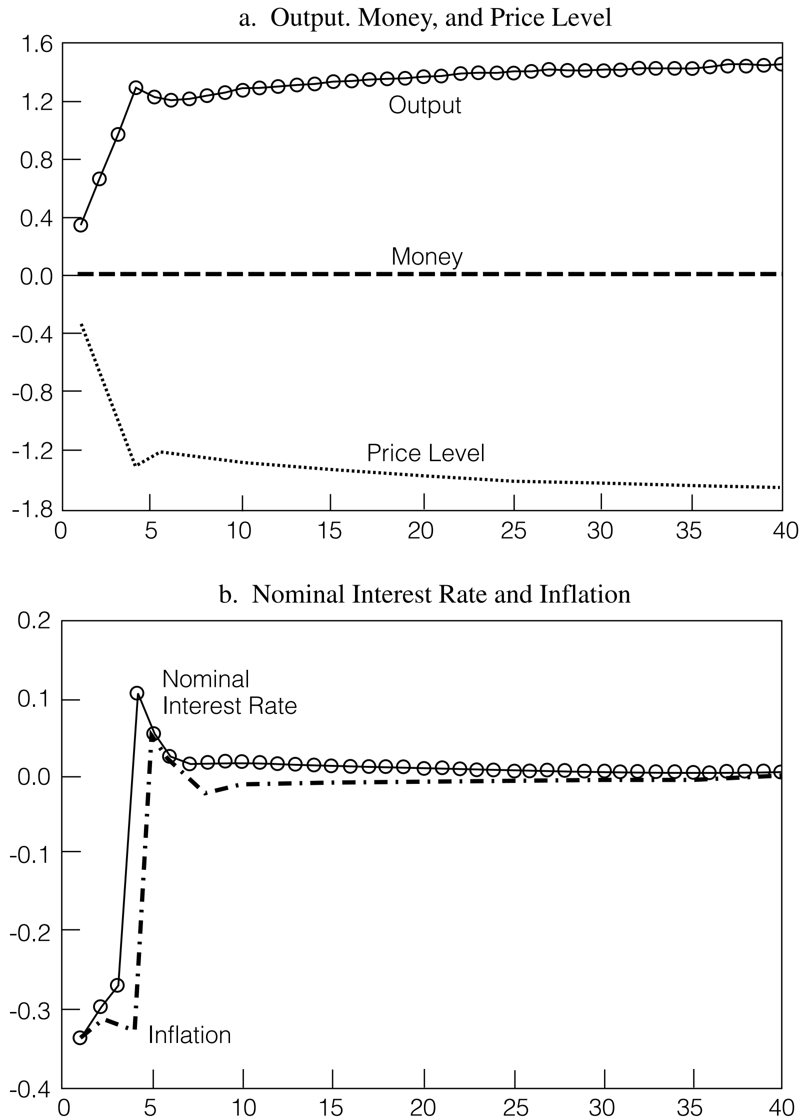
In contrast, Figures 2a and 2b depict the behavior of the economy in response to the same shock but with money supply growth kept at its steady-state rate of 2 percent. From the money demand curve, equation (2), it is clear that nominal income growth cannot deviate from steady state. If prices were flexible, they would fall by enough so that output would behave as shown in Figure 1a. But, because 75 percent of the firms are unable to adjust their prices, the price level declines by much less, and as a result the response of real output is damped. As additional firms adjust their price over time, the price level falls and output eventually reaches its new steady state. Falling prices imply disinflation and a decline in the nominal interest rate. This behavior is shown in Figure 2b.

Figure 1 Constant Markup and Technology Shock



By analyzing the various interest rate rules, it will become evident that they differ in their ability to produce the type of output behavior associated with flexible prices. The above discussion should help in clarifying why that is the case.

Figure 2 Constant Money Growth and Technology Shock



2. MONETARY POLICY

For studying the effects that the systematic part of monetary policy has on the transmission of the various shocks to the economy, I shall be fundamentally concerned with two basic types of policy rules. These rules employ an interest rate instrument and fall into the category broadly labeled as Taylor (1993) type rules. Both rules are backward-looking and allow the central bank to respond to deviations of past inflation from its target and past output levels from the steady-state level of output. However, one rule implies interest smoothing on the part of the monetary authority. Specifically, the first rule is given by

$$R_t = \bar{r} + \pi^* + 0.75(\pi_{t-1} - \pi^*) + 0.6(Y_{t-1} - \overline{Y_{t-1}}), \tag{7}$$

where π^* is the inflation target of 2 percent, and $\overline{Y}$ is the steady-state level of output. Under this rule, the central bank responds only to readily available information when adjusting the nominal rate of interest. When inflation is running above target or output is above trend, the central bank tightens monetary policy by raising the nominal interest rate. This type of rule restores the inflation rate to 2 percent after the shock's effects dissipate. The rule differs from the original one proposed by Taylor in that it includes a response to last quarter's lagged inflation rather than current yearly inflation, and the coefficient on inflation is somewhat smaller than that initially specified by Taylor (1993). This specification is adopted for two reasons. First, as emphasized by McCallum (1997), the elements of a feedback rule should involve only variables that are readily observable. Although contemporaneous output and inflation are observed in the stylistic setting of the model, in practice these variables may be observed only with a lag.[5] Second, the rule specified in (7) is explosive for the parameters chosen by Taylor. In the above lagged specification, explosive behavior results if the monetary authority responds too aggressively to both inflation and output.[6]

The second rule is similar to the first but adds a degree of interest rate smoothing. The actual interest rate can be thought of as a weighted average of some target that depends on the state of the economy and last period's nominal interest rate. The greater the weight on the nominal interest rate, the more concerned the monetary authority is for smoothing the interest rate. This rule is given by[7]

$$R_t = \bar{r} + \pi^* + 0.75R_{t-1} + 0.75(\pi_{t-1} - \pi^*) + 0.15(Y_{t-1} - \overline{Y_{t-1}}). \tag{8}$$

[5] In actuality the Fed does not observe potential or steady-state output either. It must respond to estimates.

[6] If the interest rate rule was specified in terms of the deviation of a four-quarter average of inflation from target, then a coefficient on inflation's deviation from target of 1.5 and a coefficient on output's deviation from a potential of 0.5 would produce well-behaved economic responses to shocks. For a detailed discussion of issues regarding determinacy and instability, see Rotemberg and Woodford (1997) and Christiano and Gust (1999).

[7] I initially tried to use the same coefficient on output as in the first rule, but the behavior of the economy was erratic. Scaling down the coefficient on lagged real activity produced more reasonable behavior.

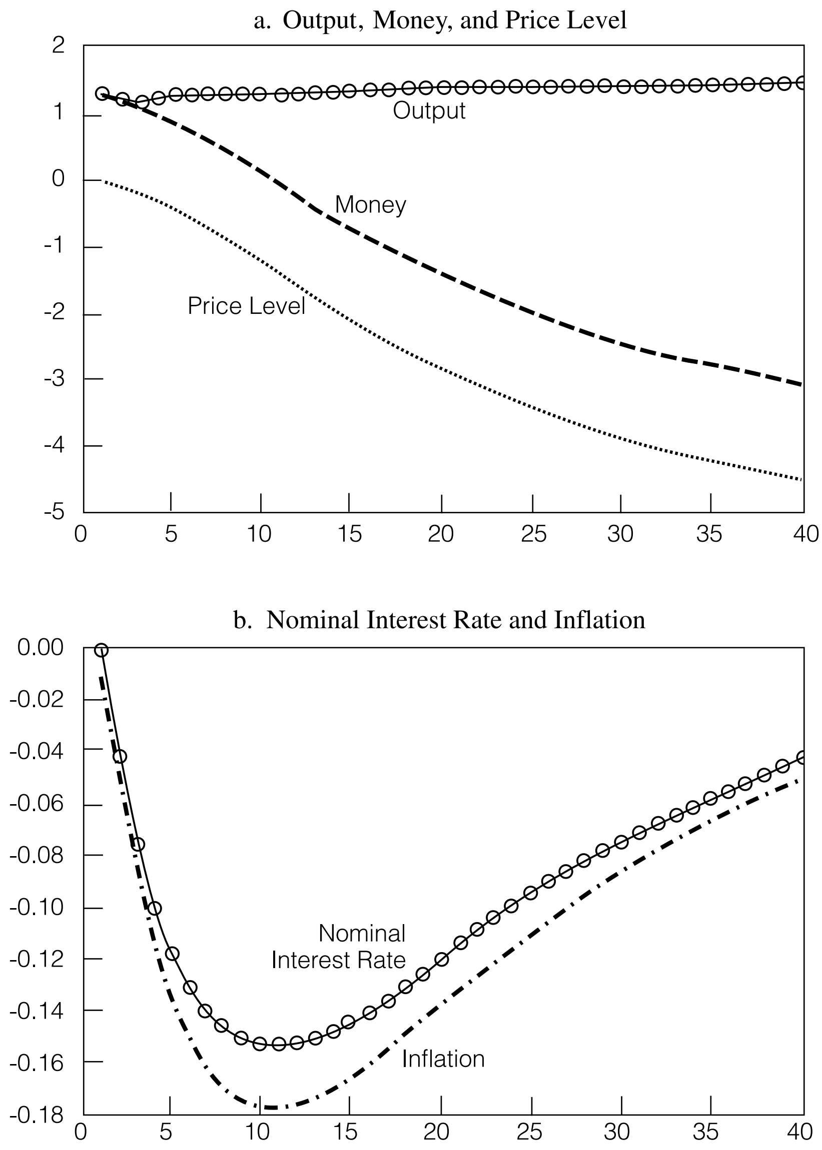
Contrary to many theoretical and empirical studies, the model experiments I run in the ensuing section take a far-from-typical perspective concerning the effects of monetary policy than is usually taken in theoretical and empirical studies. Standard investigations attempt to determine how the economy reacts to policy shocks represented as unexpected disturbances to either money growth rates or the interest rate set by the Fed. While those analyses tackle an interesting problem, only recently have economists begun to analyze the economic effects of the systematic component of policy. By concentrating on the sensitivity of the economy's responses to various shocks under different policies, the article has a different emphasis from much of the recent work on systematic policy. The analysis is, therefore, similar in emphasis to recent papers by McCallum (1999) and Christiano and Gust (1999).

3. A COMPARISON OF THE POLICY RULES

This section analyzes the way the model economy reacts to a technology shock under the two different interest rate rules. These responses are depicted in Figures 3 and 4, where as before all changes represent deviations from steady-state values. Figures 3a and 3b and Figures 4a and 4b refer to rules 1 and 2, respectively. The differences across the policy rules are striking, especially when one also considers the behavior depicted in Figures 1 and 2. Although output increases on impact under each rule, the magnitude of the increase varies greatly across policy regimes, ranging from less than 0.4 percent for a money growth rule to approximately 2.5 percent under the Taylor rule that uses interest rate smoothing. The impulse response for output under the Taylor rule that does not employ interest rate smoothing is closest to the response shown in Figures 1 and 2 for a standard RBC model.

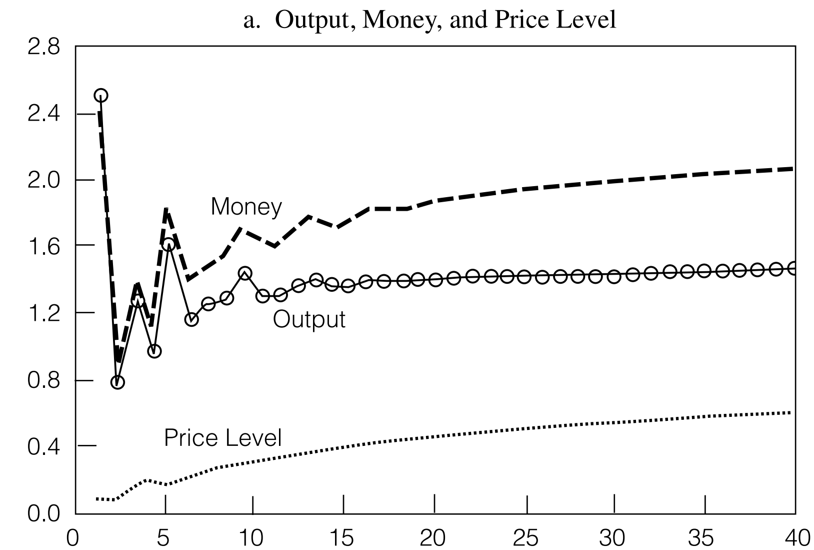
The nominal behavior of the economy is also very different under the various rules. Under the first rule the price level barely moves on impact but then falls as the effect of the technology shock works its way through the economy. This behavior is in sharp contrast to that associated with the constant money growth rule in which the price level declines on impact. It is, therefore, not staggered price setting that is responsible for the initial stickiness of the price level but the specification of the interest rate rule. Because the nominal interest rate responds only to lagged variables, it doesn't react initially. Consequently, all the money demanded at the initial interest rate is supplied, and there is no need for price adjustment to equilibrate the money market. With output slightly below its new steady-state value, the nominal interest begins to decline, and it continues to decline in response to falling prices. It is important to emphasize that the decline in the interest rate does not represent an easing of policy but rather an endogenous response to an economic shock. That is, the central bank is not attempting to independently stimulate the economy.

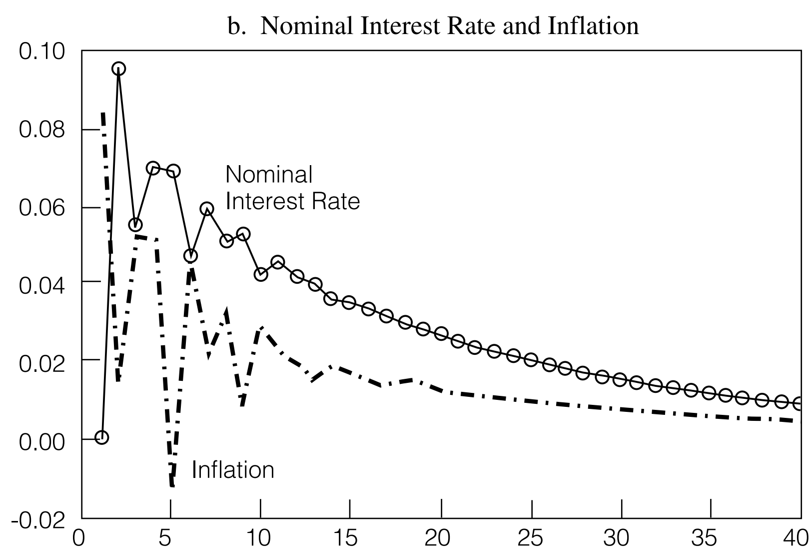
Figure 3 Taylor Rule and Technology Shock



Under the second rule the economy booms. Output rises by an extraordinary 2.5 percent, and with it is an accompanying increase in marginal cost as firms must bid up the wage to induce additional labor supply. The increase in marginal cost implies that adjusting firms will raise their prices. In contrast

Figure 4 Smooth Taylor Rule and Technology Shock





to the previous example, the economy experiences inflation in response to the increase in technology. Because prices will be rising over time, the current period is a relatively good time to consume, and output demand is high as well. The increase in inflation as well as the increase in output above its new steady-state level causes the central bank to raise interest rates. As in the previous case, the subsequent rise in the interest rate should not be interpreted as an attempt to shock an overheated economy but simply as the central bank's usual response to strong economic growth. The endogenous rise in the interest rates, as we shall see in the next section, is responsible for the dramatic fall in economic activity. The initial overshooting is subsequently corrected, and output then gradually approaches its new steady-state level. It is important to note that under the two rules the marked difference in the impulse response functions is not due to the somewhat smaller coefficient on lagged output in the second rule. If that coefficient were increased to 0.6, then the response of the economy would be similar but the volatility, or saw-toothed behavior, of the variables would be more pronounced.

The difference in the functions is due to the interest rate smoothing present in the second policy rule. Under the first policy rule, any increase in inflation is aggressively reacted to because the monetary authority does not have to take into account the past level of the interest rate. Knowing the relatively aggressive nature of policy, individuals and firms expect less inflation, creating less pressure to raise prices. The subsequent downward path of prices makes postponing purchases optimal. As a result, there is less demand pressure in response to the shock. Output does not rise to its new steady-state level on impact, and there is no upward pressure on marginal cost. Under the second policy rule, the monetary authority will be less aggressive, so prices are expected to rise. Such expectations spur consumers to purchase goods today, resulting in relatively strong aggregate demand. The economy booms and the combined effect of expected inflation and upward pressure on marginal cost causes firms to raise prices.

There are a number of points to take away from the analysis presented in this section. First and foremost is that the systematic component of monetary policy is key in determining the economy's reaction to shocks. For example, with no interest rate smoothing, inflation and prices are negatively correlated with output, while they are positively correlated when the monetary authority smooths the interest rate. As a consequence of sticky prices, both positive and negative correlations between real and nominal variables are possible. The type of correlation observed may be entirely due to the systematic behavior of monetary policy and have nothing to do with the structure of the economy.

4. A FURTHER INVESTIGATION OF TAYLOR RULES

In this section I illustrate the sensitivity of the model economy's responses under the two policy rules to a transitory tightening of monetary policy as reflected in a 100 basis point increase in the nominal interest rate. As with the case of the technology shock, the responses are very different. These responses are displayed in Figures 5 and 6, with 5a and 5b depicting the response under the first rule and 6a and 6b reflecting behavior under the second rule.

Under rule 1 output actually rises on impact, while under the interest rate smoothing rule output falls. This difference in behavior occurs because the unexpected rise in the nominal rate under the first rule will accommodate modest inflation. As long as inflation doesn't accelerate—behavior the rule is designed to prevent—the nominal rate will gradually return to steady state, and there will be upward movement in prices as well as strong economic growth. The economy only suffers a mild recession four quarters into the future.

The presence of interest rate smoothing in this case means that any upward movement in real economic activity or inflation will drive the interest rate even higher. Rather than acting as an anchor as in the previous section, the interest smoothing term implies a much more aggressive response to nominal growth. Because today's interest rate is high, all things being equal, the next period's interest rate will be high as well. Individuals and firms understand the nature of the rule, and, therefore, an increase in nominal activity is inconsistent with interest rate behavior under this rule. Output, prices, and inflation decline immediately in response to the rise in the nominal interest rate. It is noteworthy that the nominal interest is more volatile under the policy rule that reflects a concern for interest rate smoothing.

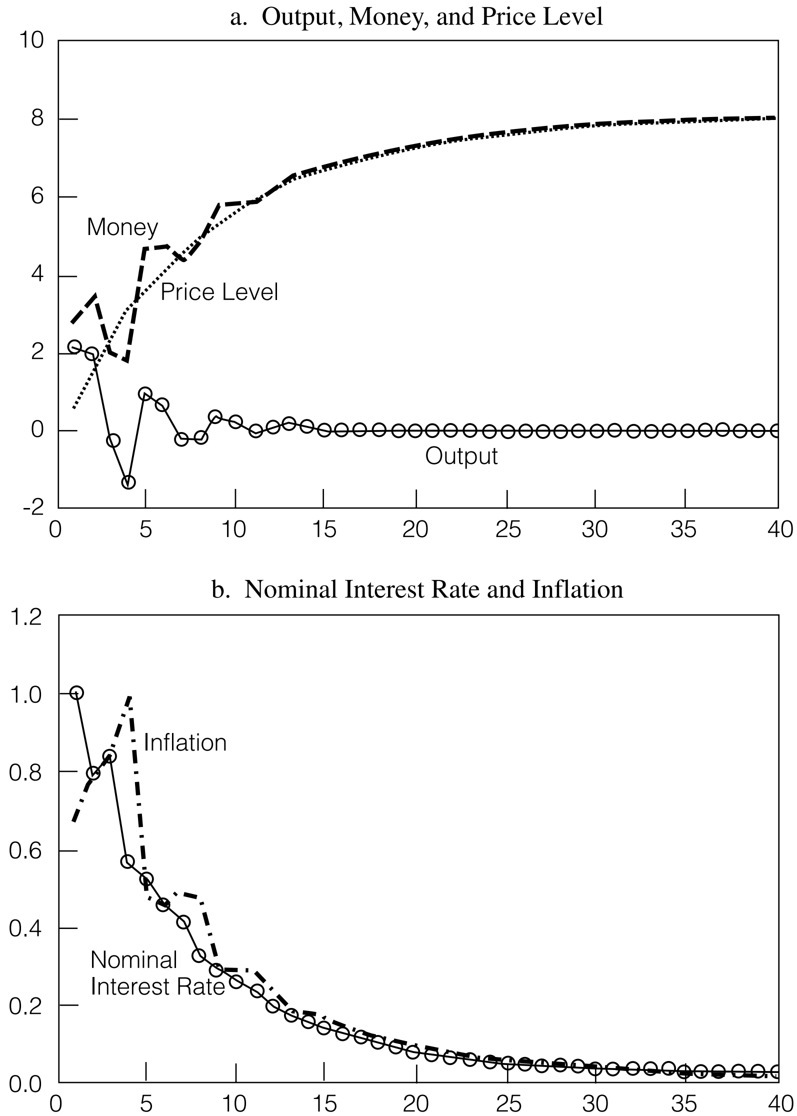
Thus, if the Fed were to significantly and periodically alter its reaction to the past behavior of interest rates, policy would appear to operate with variable impact effects and variable lagged effects. It would do so not because the changes in the policy rule are reflected in small quantitative differences in the economy's response to policy shocks but because these changes in policy may actually lead to an economy that qualitatively responds in a different way altogether.

Admittedly, the behavioral changes analyzed may be severe and the model economy may not reflect important elements of actual behavior, but the experiments in this and the preceding section send a strong message that the form of the policy rule is far from innocuous.

5. CONCLUSION

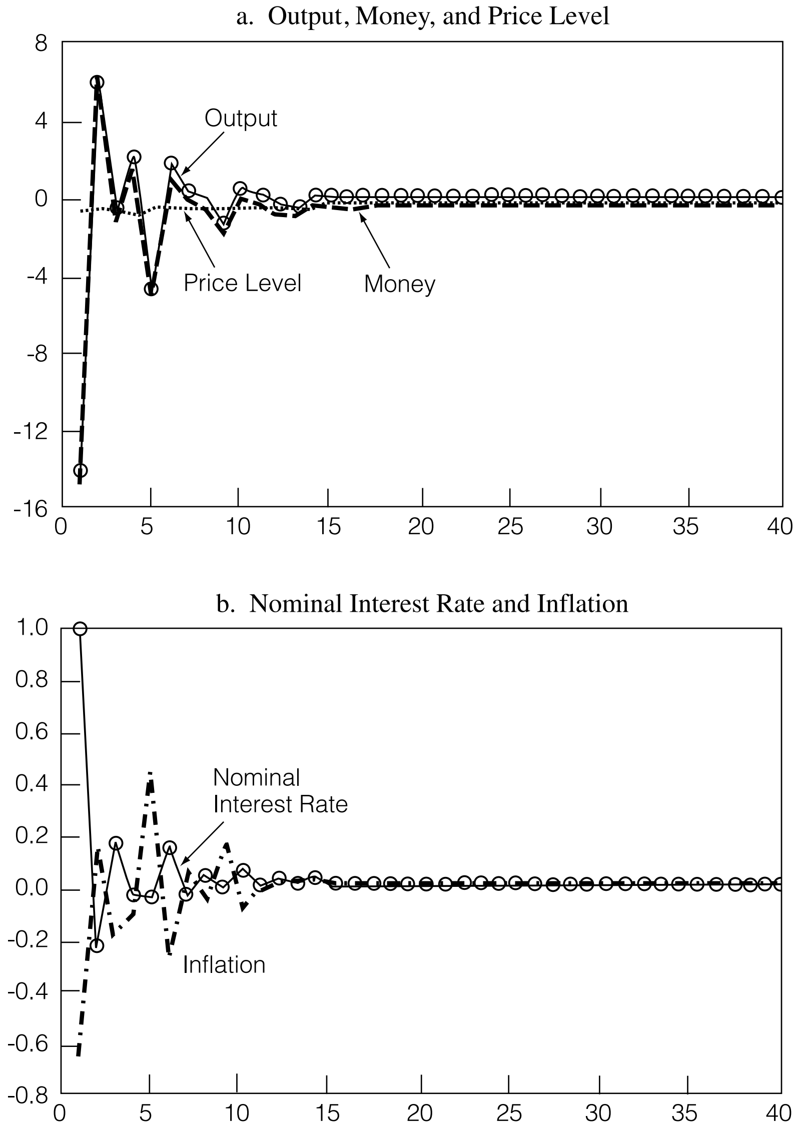
The basic conclusion of this article is that money matters. More to the point, monetary policy matters, and specifically the systematic part of monetary policy matters. While most studies have devoted a great deal of effort to understanding and quantifying the economic effects of monetary policy shocks, my results indicate that it may be equally if not more important to determine the appropriate design of a policy rule. From my own perspective, which is influenced by numerous (or perhaps endless) policy debates, what is typically discussed is not what monetary disturbance should impact the economy but what response

Figure 5 Taylor Rule and Policy Shock



should policy have to the economy. Significant tightenings of policy are generally not an attempt to shock the economy but the Fed's realization that inflation and expected inflation have risen and that tightening is appropriate. The degree of the response may, and probably does, vary in different periods. And it

Figure 6 Smooth Taylor Rule and Policy Shock



may be inappropriate to model these changes as shocks to an unvarying rule. As I have shown, the sign of correlations among economic variables can differ across rules. That type of behavior would not be captured by appending a shock to a given policy rule. The message from the above exercises is that it may be more appropriate to model the coefficients in the response function as random, rather than attaching some randomness to an invariant rule.

REFERENCES

Basu, Susanto, and John Fernald. “Returns to Scale in U.S. Production: Estimates and Implications,” *Journal of Political Economy,* vol. 105 (April 1997), pp. 249–83.

Basu, Susanto, and Miles S. Kimball. “Cyclical Productivity with Unobserved Input Variation,” *NBER* Working Paper 5915. February 1997.

Batini, Nicoletta, and Andrew G. Haldane. “Forward-Looking Rules for Monetary Policy,” in John B. Taylor, ed., *Monetary Policy Rules.* Chicago: University of Chicago Press, 1999.

Canzoneri, Matthew, Dale Henderson, and Kenneth Rogoff. “The Information Content of Interest Rates and the Effectiveness of Monetary Policy Rules,” *Quarterly Journal of Economics,* vol. 98 (November 1983), pp. 545–66.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. “Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem?” Staff Report 217. Minneapolis: Federal Reserve Bank of Minneapolis, May 1998.

Christiano, Lawrence J., and Christopher J. Gust. “Taylor Rules in a Limited Participation Model,” *NBER* Working Paper 7017. March 1999.

Dotsey, Michael, and Robert G. King. “Informational Implications of Interest Rate Rules,” *American Economic Review,* vol. 76 (March 1986), pp. 33–42.

__________. “Monetary Instruments and Policy Rules in a Rational Expectations Environment,” *Journal of Monetary Economics,* vol. 12 (September 1983), pp. 357–82.

Fischer, Stanley. “Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule,” *Journal of Political Economy,* vol. 85 (February 1977), pp. 191–205.

Goodfriend, Marvin S., and Robert G. King. “The New Neoclassical Synthesis and the Role of Monetary Policy,” *NBER Macroeconomics Annual 1997.* Cambridge, Mass.: MIT Press, 1997.

King, Robert G., and Alexander L. Wolman. “What Should the Monetary Authority Do When Prices are Sticky?” in John B. Taylor, ed., *Monetary Policy Rules.* Chicago: University of Chicago Press, 1999.

McCallum, Bennett T. “Analysis of the Monetary Transmission Mechanism: Methodological Issues.” Manuscript. February 1999.

———. "Comment" [on "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy"]. *NBER Macroeconomics Annual 1997.* Cambridge, Mass.: MIT Press, 1997, pp. 355–59.

———, and Edward Nelson. "Performance of Operational Policy Rules in an Estimated Semiclassical Structural Model," in John B. Taylor, ed., *Monetary Policy Rules.* Chicago: University of Chicago Press, 1999.

Mulligan, Casey B. "Substitution over Time: Another Look at Life Cycle Labor Supply," *NBER Macroeconomics Annual 1998.* Cambridge, Mass.: MIT Press, 1999.

Poole, William W. "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Model," *Quarterly Journal of Economics,* vol. 84 (May 1970), pp. 197–216.

Rotemberg, Julio J., and Michael Woodford. "Interest Rate Rules in an Estimated Sticky Price Model," in John B. Taylor, ed., *Monetary Policy Rules.* Chicago: University of Chicago Press, 1999.

Rudebusch, Glenn D., and Lars E. O. Svensson. "Policy Rules for Inflation Targeting," in John B. Taylor, ed., *Monetary Policy Rules.* Chicago: University of Chicago Press, 1999.

Taylor, John B. "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy,* vol. 39 (December 1993), pp. 195–214.

———. "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy,* vol. 88 (February 1980), pp. 1–24.

An Empirical Investigation of Fluctuations in Manufacturing Sales and Inventory within a Sticky-Price Framework

Pierre-Daniel G. Sarte

The macroeconomics literature has recently witnessed a resurgence of interest in issues related to nominal price rigidities. In particular, advances in computational methods have allowed for the analysis of fully articulated quantitative general equilibrium models with inflexible prices.[1] Because nominal price rigidities create predictable variations in sales, these models provide a natural setting for the study of inventory behavior. Specifically, firms that face increasing marginal costs wish to smooth production and, given predictable variations in sales, can naturally use inventories to accommodate any difference between a smooth production volume and sales.

Hornstein and Sarte (1998) study the implications of sticky prices for inventory behavior under different assumptions about the nature of the driving process. Regardless of whether the economy is driven by nominal demand or real supply shocks, the authors find that an equilibrium model with inflexible prices can replicate the main stylized facts of inventory behavior. Namely, production is more volatile than sales while inventory investment is positively correlated with sales at business cycle frequencies. More importantly, their study also makes specific predictions about the dynamic adjustment of inventories and sales to these shocks. In response to a permanent positive money growth

Pierre.Sarte@rich.frb.org. I wish to thank Andreas Hornstein and Mark Watson for helpful discussions. I also wish to thank Yash Mehra, Wenli Li, and Roy Webb for thoughtful comments. The opinions expressed herein are the author's and do not necessarily represent those of the Federal Reserve Bank of Richmond or the Federal Reserve System. Any errors are, of course, my own.

[1] See Goodfriend and King (1997) for a survey of recent work.

innovation, both sales and inventories contemporaneously rise before gradually returning to the steady state. In contrast, a permanent positive technology shock leads to a rise in sales and a fall in inventories on impact. As time passes by, sales increase monotonically and eventually reach a new higher steady-state level.

In this article, we estimate a structural vector autoregression (SVAR), where money is constrained to be neutral in the long run, in order to gauge the degree to which these theoretical dynamic adjustment paths hold in the data. Using manufacturing data, we find that the impulse response of sales and inventories to nominal shocks is generally consistent with the predictions of a sticky-price model. Furthermore, both sales and inventories also behave as predicted in the long run in response to a technology shock. Contrary to theory, however, we find that inventories contemporaneously rise in response to a positive innovation in technology. In all cases, the data indicate significantly more sluggishness in the dynamic adjustment of sales and inventories to shocks than implied by current models with sticky prices. The latter finding is consistent with earlier work by Feldstein and Auerbach (1976), as well as Blinder and Maccini (1991), using stock-adjustment equations. More recently, Ramey and West (1997) also find that the inventory:sales relationship is unusually sluggish. They are able to explain this result by appealing either to persistent shocks to the cost of production or to a strong accelerator motive within a linear quadratic framework.

Although the earlier analysis in Hornstein and Sarte (1998) makes specific predictions regarding the dynamic response of sales and inventories to various shocks, it does not assess the relative importance of these shocks as sources of fluctuations. Here we use our estimated VAR to acquire some insight into the significance of both real and nominal shocks in generating fluctuations in sales and the inventory:sales ratio. We find that nominal shocks generally contribute little to the forecast error variance in the latter variables at both short and long horizons. Instead, consistent with earlier work such as King, Plosser, Stock, and Watson (1991), fluctuations in real variables tend to be dominated by real disturbances. Moreover, these empirical findings tend to hold consistently throughout different historical episodes at the business cycle frequency. One exception concerns monetary disturbances that play a noticeably more important role in generating inventory:sales ratio fluctuations in the early 1990s.

This article is organized as follows. We first set up and motivate an empirical model that is consistent with generic restrictions implied by an equilibrium model of inventory behavior. In particular, we assume that money is neutral in the long run and that the inventory:sales ratio is a stationary process without trend. Note that we do not impose any a priori restrictions that are directly tied to the assumption of sticky prices. The next section examines various integration and cointegration properties of the data under consideration. We then analyze the impulse responses of sales and the inventory:sales ratio to various shocks. We also try to gauge the relative importance of these shocks as sources of fluctuations in the latter variables. After that, we offer some cautionary remarks regarding the specific empirical implementation in this article. The final section concludes the analysis.

1. INVENTORY FLUCTUATIONS: THEORETICAL MOTIVATION

To set the stage and notation for the econometric specification, we will provide some theoretical background on the behavior of inventories. The basic framework we have in mind is one in which firms use inventories to smooth production in a setting with staggered nominal prices.[2] The assumption of inflexible price adjustment provides a natural role for production smoothing as the underlying factor driving inventory behavior. In particular, nominal price rigidity creates predictable variations in sales. Suppose, for instance, that the nominal price set by a given firm is fixed over some time interval. If the general price level increases over that time interval, then the firm's relative price correspondingly falls and its sales rise, all else being equal. Given this rising sales path, the firm also attempts to minimize total production costs by keeping production relatively smooth. Inventories can then be used to make up for the differences between production and sales. In addition to identifying this sticky-price motive, we, like Khan (1987), assume that firms may also hold inventories to avoid costly stock-outs.

Within the context of this framework, the dynamic adjustment of inventories and sales to various shocks will generally depend on how preferences and technology are specified. In the long run, however, the model exhibits basic neoclassical properties that can be used for the purposes of identification. One of these properties suggests that money is neutral and, moreover, that changes in the steady-state level of sales ultimately arise from innovations in technology. With this in mind, we let the long-run component of the sales process evolve according to

$$s_t^* = \delta_s + s_{t-1}^* + \Phi_s(L)a_t, \tag{1}$$

where s_t^* denotes the log level of sales and a_t captures shocks to technology. The lag polynomial $\Phi_s(L)$, as well as all other polynomials described below, is assumed to have absolutely summable coefficients with roots lying outside the unit circle. Observe that equation (1) implicitly assumes that the sales process possesses a unit root. We formally test this assumption later in this article.

In principle, the steady-state level of inventories can be thought of as being determined by the two forces we described previously. Note that in a

[2] See Hornstein and Sarte (1998) for details of the model.

model with rigid prices, firms naturally wish to hold inventories to accommodate any difference between predictable variations in sales and a smooth production volume. Moreover, by using inventories to avoid costly stock-outs, firms generally target some appropriate inventory:sales ratio in the long run. Although the short- and medium-run dynamics of inventories typically depend on both these forces, Hornstein and Sarte (1998) note that in the steady state, the level of inventories reflects almost exclusively the stock-out avoidance motive. Accordingly, we may express long-run inventories as

$$n_t^* = s_t^* + \xi, \tag{2}$$

where n_t^* denotes the log level of inventories and ξ is some target inventory:sales ratio. It immediately follows from (1) and (2) that inventories and sales share a common stochastic trend whose growth rate is $\delta_s + \Phi_s(L)a_t$. Furthermore, as we make clear below, the inventory:sales ratio becomes a stationary stochastic process.

Since in this article we are partly interested in how monetary shocks affect the dynamics of inventories and sales, we must specify our beliefs about the behavior of money. To this end, we let the long-run component of money evolve according to

$$m_t^* = \delta_m + \Phi_m(L)[a_t, \eta_t]', \tag{3}$$

where m_t^* is the log level of money and η_t denotes money innovations. Note that, as in Gali (1999), we allow monetary policy to respond permanently to long-run changes in technology, a_t. This assumption captures the idea that the Federal Reserve reacts to permanent real changes in the economic environment in its effort to keep prices stable. We further assume that a_t and η_t are serially and mutually uncorrelated shocks.

While we have assumed that long-run changes in sales are ultimately determined by technological considerations, sales may actually respond to a variety of economic shocks in the short run. More specifically, the level of sales, s_t, may deviate temporarily from its long-run value because of money shocks or transitory real demand shocks. Such real shocks may include temporary changes in tastes, for instance. Therefore, a complete process for sales can be described as

$$s_t = s_t^* + \psi_s(L)[a_t, \eta_t, e_t]', \tag{4}$$

where e_t captures a mixture of temporary real demand shocks. These are assumed to be serially uncorrelated as well as uncorrelated with a_t and η_t. In principle, the fact that s_t depends on all shocks in the model allows for flexible short-run dynamics. The aim of our empirical exercise is, in part, to gauge whether these short-run dynamics are consistent with the predictions of a model with nominal price rigidities. Taking the first difference in equation (4) and substituting equation (1) into it yields

$$\Delta s_t = \delta_s + \Phi_s(L)a_t + (1-L)\psi_s(L)[a_t, \eta_t, e_t]', \tag{5}$$

which represents one of the structural equations to be estimated.

As in (4), one generally expects the level of inventories to be sensitive to all shocks in the short run. Consequently, we may write the following stochastic process for inventories:

$$n_t = n_t^* + \psi_n(L)[a_t, \eta_t, e_t]'. \tag{6}$$

Note that the theoretical framework we have been using predicts a testable cointegrating restriction. In particular, while (1) and (2) suggest that both inventories and sales are integrated of order one (often denoted I(1)), these equations combined with (4) also suggest that the difference between inventories and sales is stationary (or I(0)). Formally, we can use (6) along with equations (1), (2), and (4) to show that

$$n_t - s_t = \xi + \{\psi_n(L) - \psi_s(L)\}[a_t, \eta_t, e_t]'. \tag{7}$$

The above equation clearly indicates that the inventory:sales ratio will deviate from its long-run value at high and medium frequency. By construction, these deviations are never permanent.

To complete the econometric specification, we allow monetary policy to respond to various shocks not only in the long run but also in the short run. The latter assumption along with equation (3) yields

$$\Delta m_t = \delta_m + \Phi_m(L)[a_t, \eta_t]' + (1-L)\psi_m(L)[a_t, \eta_t, e_t]'. \tag{8}$$

At this point, we wish to stress that the identifying restrictions made in this section are, in fact, quite generic and unrelated to the notion of sticky prices per se. Therefore, if the results below turn out to be consistent with the notion of nominal rigidities, this outcome will not be as a direct consequence of the identifying strategy used. It remains that different identification strategies may yield different results. Because our restrictions are relatively general, however, they encompass a broad range of models.[3]

2. ECONOMETRIC METHOD AND DATA ANALYSIS

Using Long-Run Restrictions for the Purpose of Identification

We can summarize our model thus far in the form of a vector moving average,

$$\mathbf{Y}_t = T(L)\varepsilon_t, \tag{9}$$

[3] See Cooley and Dwyer (1998) for a thorough discussion of the pitfalls associated with the identification of SVARs.

where $\mathbf{Y}_t = (\Delta s_t, \Delta m_t, n_t - s_t)'$ and $\varepsilon_t = (a_t, \eta_t, e_t)$. The matrix polynomial $T(L)$ consists of the polynomials $\Phi_a(L)$, $\Phi_m(L)$, $\psi_s(L)$, $\psi_n(L)$, and $\psi_m(L)$ in equations (1) through (8). In addition, embedded in $T(L)$ are long-run restrictions implied by our model that can be used to identify each of the three structural shocks. Specifically, the matrix of long-run multipliers, $T(1)$, may be written as

$$T(1) = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{12} & a_{21} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}. \tag{10}$$

Thus, the first row of $T(1)$ reflects our restriction that only technology shocks alter the level of sales in the long run (in the steady state, sales should equal production). Money, therefore, is neutral, and we can appropriately constrain the estimation of the sales growth equation to identify technology shocks. To see how to impose the restrictions contained in $T(1)$, note first that under the assumption that $T(L)$ is invertible, $T(1)^{-1}$ is also lower block triangular. In estimating the sales growth equation, therefore, it suffices to set the long-run elasticity of Δs_t with respect to both Δm_t and $n_t - s_t$ to zero.

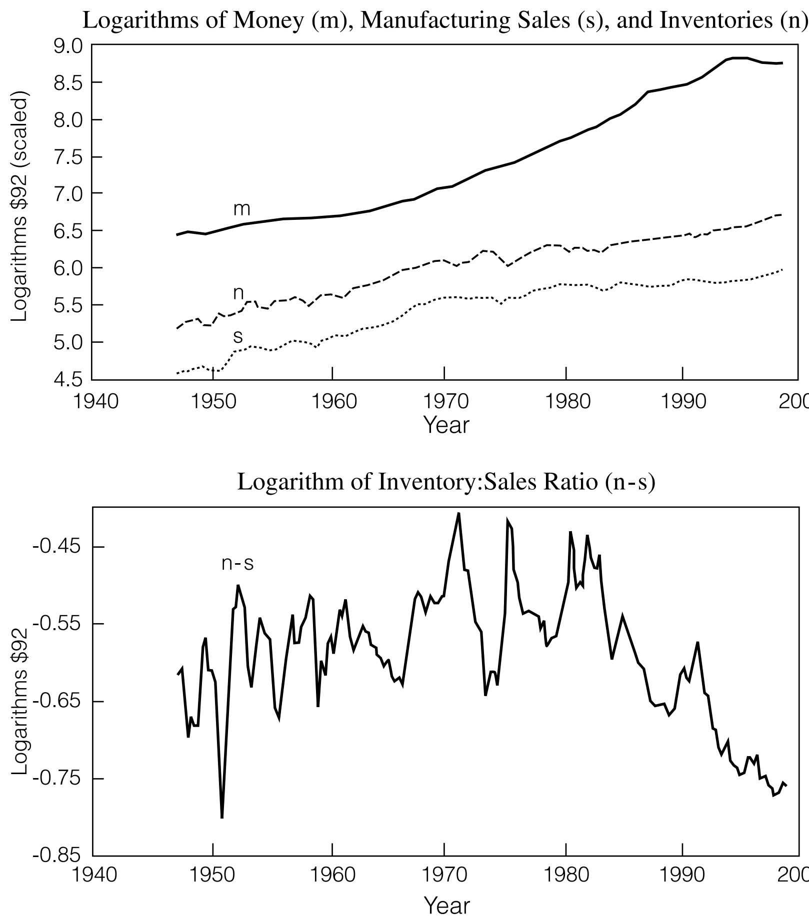
Real transitory demand shocks cannot, by definition, have long-run effects on any of the variables in the model. As the second row of $T(1)$ suggests, this restriction, already imposed in estimating the sales growth regression, can be used to identify money shocks. In other words, except for permanent changes in technology, long-run changes in money are associated only with their own innovations, as equation (3) illustrates. To uncover money innovations, therefore, we estimate the money growth equation subject to the restriction that the long-run elasticity of Δm_t with respect to $n_t - s_t$ be set to zero. It remains that the long-run elasticity of Δm_t with respect to Δs_t will generally not be zero. To account for the presence of this contemporaneous endogenous variable in the money growth regression, we use the fact that the structural disturbances are assumed to be mutually uncorrelated and use the residual from the sales growth regression as an instrument. The econometric methodology used here, therefore, follows that of Shapiro and Watson (1988), Blanchard and Quah (1989), King, Plosser, Stock, and Watson (1991), as well as many others.

It follows that the last remaining innovation captures real temporary demand shocks. In particular, the third row of $T(1)$ suggests that the latter shocks can simply be uncovered by estimating the inventory:sales ratio equation without any restrictions. We use the residuals from both the sales growth and money growth regressions to instrument for Δs_t and Δm_t in this last regression.

Cointegration Properties of the Data

Before proceeding with the estimation, we first investigate the cointegrating restriction implied by (7). As with the majority of the empirical literature on inventory behavior, this article focuses mainly on manufacturing inventories.

Figure 1



More specifically, the notion of production smoothing applies best to manufactured goods, as pointed out in Hornstein (1998). In Section 4, we shall take the research one step further by showing that the econometric specification above may be ill-suited to both the retail and service sectors. We add one cautionary note, however, regarding our assumption that money may respond to long-run innovations in technology (recall equation [3]). In all likelihood, this assumption is most relevant for aggregate shocks rather than sectoral shocks. Our model does not allow us to disentangle these shocks. Consequently, shocks captured by a_t should be interpreted as a linear combination of both aggregate and sectoral innovations.

Table 1 Cointegration Statistics—1947:1–1998:3

a. Results from Unrestricted Levels Vector Autoregression: Largest Eigenvalues of Estimated Companion Matrix

VAR(6) with constant and trend

Real	**Imaginary**	**Modulus**
0.98	0.04	0.98
0.98	−0.04	0.98
0.85	0.15	0.86
0.85	−0.15	0.86
0.63	0.47	0.78
0.63	−0.47	0.78

b. Multivariate Unit-Root Statistics: Stock and Watson q^f Statistic

H0: 3 unit roots vs. H1: at most 2 unit roots

Number of Lags	$q^f_T(3,2)$ statistic	P value
1	−42.51	2.75
2	−45.10	1.75
3	−42.27	2.75
4	−34.68	9.75
5	−31.89	14.75

H0: 2 unit roots vs. H1: at most 1 unit root

Number of Lags	$q^f_T(2,1)$ statistic	P value
1	−8.99	83.25
2	−11.94	67.50
3	−9.39	81.25
4	−8.69	85.00
5	−7.87	88.75

Figure 1 shows the logarithms of money, as defined by M1 (i.e., currency and demand deposits), manufacturing inventories, and sales of finished goods. The data are quarterly U.S. observations spanning the period 1947:1 to 1998:3. Early figures for M1 were obtained from the Monetary Statistics of the United States since they were unavailable from the Board of Governors dataset. The inventory and sales data were downloaded from the National Income and Products Accounts on February 19, 1999. Regressions were run over the period 1948:3 to 1998:3 to allow for six lags. The plots of the variables display familiar, clear upward trends with inventories being the most volatile component. Note that inventories and sales indeed seem to share the same trend over the period considered. Figure 1 also plots the logarithm of the inventory:sales ratio, $(n - s)$, which appears relatively stable. One possible exception concerns the period

Table 1 Cointegration Statistics—1947:1–1998:3 (cont.)

Johansen's Likelihood Ratio Statistics

	$2(\mathcal{L}_1 - \mathcal{L}_0)$	5% critical value
H0: $h = 0$ vs. H1: No restrictions:	$-T\sum_{i=1}^{3}\log(1-\lambda_i) = 41.97$	29.51
H0: $h = 0$ vs. H1: $h = 1$:	$-T\log(1-\lambda_1) = 25.99$	20.77
H0: $h = 1$ vs. H1: No restrictions:	$-T\sum_{i=2}^{3}\log(1-\lambda_i) = 15.98$	15.20
H0: $h = 1$ vs. H1: $h = 2$:	$-T\log(1-\lambda_2) = 12.95$	14.03

Notes: $T = 201$, where T is the sample size, $\lambda_1 = 0.1213$, $\lambda_2 = 0.0624$, and $\lambda_3 = 0.0148$, where the λ_i's refer to the square of the canonical correlations.

Saikkonen's Estimator for Cointegrated Regressions

Variable	Null Hypothesis	Estimates
s	-1	-1
m	0	0
n	1	0.95 (0.02)

Wald Test for the Cointegrating Vector $(-1, 0, 1)$: $\chi^2_{[1]} = 12.08$.

beginning in the early 1990s in which this ratio seems to have started to fall.[4] On the whole, however, it would be difficult to argue that the inventory:sales ratio does not fluctuate around a constant mean. Alternatively, Figure 1 loosely suggests that inventories and sales are cointegrated.

A univariate analysis of the three variables plotted in Figure 1 suggests that they can each be characterized as an I(1) process with positive drift. Our concern, however, is mostly with a multivariate analysis of the relationship described by (9). Accordingly, Tables 1a and 1b present a number of statistics that relate to the three-variable system, $\widetilde{\mathbf{Y}}_t = (s_t, m_t, n_t)'$.

Panel a of Table 1 shows the largest eigenvalues of the companion matrix associated with a VAR(6) estimated with a constant and a linear trend. Under the assumption that only one cointegrating restriction links the variables in $\widetilde{\mathbf{Y}}_t$, the companion matrix should have two unit eigenvalues corresponding to two common stochastic trends. All other eigenvalues should be less than one in modulus. These results follow directly from Stock and Watson's (1988) common trends representation. The point estimates displayed in Table 1a indeed

[4] Regressions were also run over the period 1947:1 to 1990:1 to check for robustness with respect to this feature of the data. Our empirical results, however, were largely unaffected.

support the hypothesis of two common trends or, alternatively, that there exists a single cointegrating restriction in our three-variable system.

Panel b presents more formal tests of cointegration developed by both Stock and Watson (1988) and Johansen (1988). Stock and Watson's $q_{\tau}^{f}(k, m)$ statistic tests the null of k unit roots against the alternative of m, ($m < k$), unit roots using Stock and Watson's (1989) dynamic Ordinary Least Squares (OLS) procedure. Specifically, if there are n variables and h cointegrating vectors, the procedure estimates h regression equations containing a constant, $n - h$ regressors in levels, as well as leads and lags of the first differences in these regressors as right-hand-side variables. The τ subscript indicates that a linear trend is included in the regressions. In panel b of Table 1, we note that the $q_{\tau}^{f}(3, 2)$ statistic is consistent with rejecting the null of no cointegrating restrictions against the alternative of at least one cointegrating restriction. In particular, the P values are generally small regardless of the number of lags used in the dynamic OLS equations. In addition, the $q_{\tau}^{f}(2, 1)$ statistic suggests rejecting the alternative of two cointegrating restrictions against the null of one cointegrating vector. Put together, these results provide evidence of only one cointegrating vector in our three-variable system.

Panel b also presents results obtained from Johansen's Likelihood Ratio Trace and Maximum Eigenvalue statistics. For these statistics, we can think of the number of unit roots as the number of variables less the number of cointegrating relations. Consider first the likelihood ratio test for the null of zero cointegrating relation against the alternative of three cointegrating relations. For this test, the likelihood ratio statistic, $2(\mathcal{L}_1 - \mathcal{L}_0)$, is 41.97, which is greater than 29.51. Therefore, the null hypothesis is rejected at the 5 percent significance level. Similarly, the test statistic for the null of zero cointegrating restriction against the alternative of one restriction is $25.99 > 20.77$. It follows that the null hypothesis of no cointegration is rejected by this second test as well.

To see whether a second cointegrating relation potentially exists, consider the likelihood ratio test for the null of $h = 1$ against the alternative of $h = 3$. In this case, the test statistic is $15.98 > 15.20$ so that the null hypothesis is, in fact, rejected at the 5 percent significance level. However, the likelihood ratio test statistic for the null of one cointegrating relation against the alternative of two relations is $12.95 < 14.03$. Therefore, although the Johansen tests generally suggest one cointegrating relation, they also offer conflicting evidence as to the presence of a second cointegrating restriction.

Finally, panel b gives an estimate of the cointegrating relation associated with the vector of variables (s_t, m_t, n_t) using Saikkonen's (1991) procedure.[5]

[5] This procedure is essentially that of dynamic OLS. In this case, the regression involves the level of sales as the dependent variable; as right-hand-side variables it involves the level of inventories, a constant but no deterministic trend, as well as leads and lags of the differences in inventories.

Although the Wald Statistic suggests rejecting the null hypothesis that the cointegrating vector is proportional to $(-1, 0, 1)$, the point estimates are broadly consistent with the notion that the inventory:sales ratio is stationary.

3. ESTIMATION OF A THREE-VARIABLE SYSTEM

The results presented in this section are based on the estimation of the Vector Error Correction Model (VECM) implied by equation (9). Each regression equation is estimated using six lags of Δs_t, Δm_t, the error-correction term $n_t - s_t$, as well as a constant. As we indicated earlier, the triangular nature of the long-run multiplier matrix and the assumption that the structural error terms are mutually uncorrelated allows us to recursively estimate each equation in the system. In estimating the money growth equation, the residual from the sales growth regression was used to instrument for contemporaneous endogenous variables. Similarly, in estimating the inventory:sales ratio equation, the residual from the money growth regression was added to the list of instruments.[6]

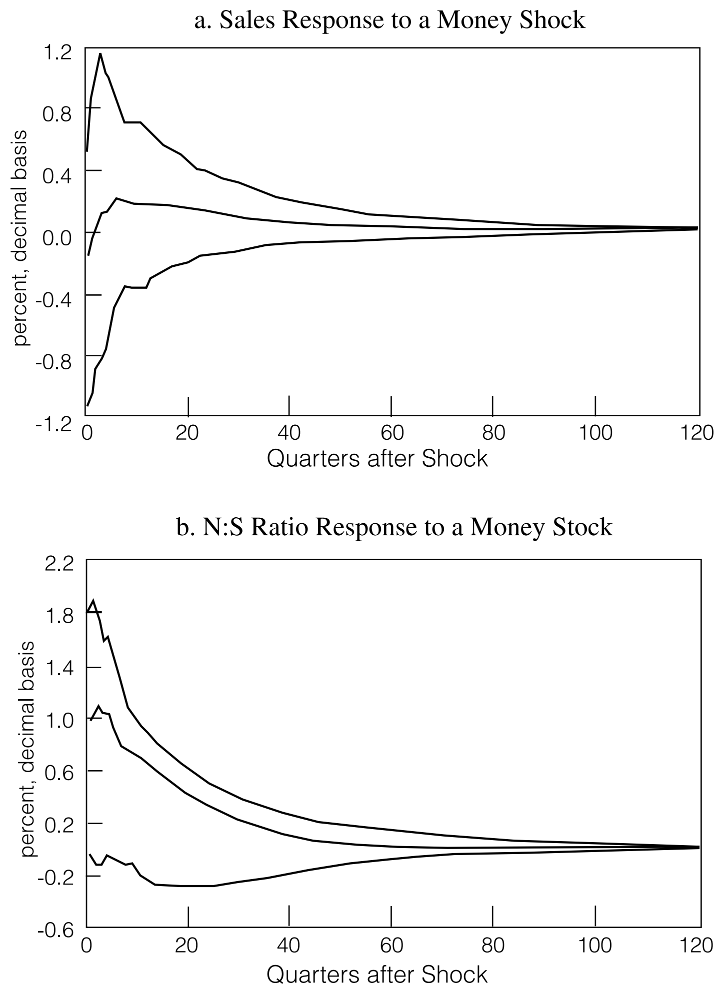
Estimated Structural Impulse Responses

Figure 2 displays the estimated impulse response function obtained from the system summarized by (9). The 95 percent confidence bands also displayed in Figure 2 were computed using Monte Carlo simulations. These simulations were carried out by using draws from the normal distribution for the technology, money growth, and temporary real demand innovations. One thousand Monte Carlo draws were completed.

We now interpret these impulse response functions in terms of a production-smoothing model with nominal rigidities. Let us first focus our attention on the effect of a money growth innovation. In a framework with staggered prices, Hornstein and Sarte (1998) suggest that in response to a money growth shock, sales should contemporaneously rise before gradually reverting back to the steady state. To see why this is true, note that a firm that does not adjust its price following an increase in nominal demand naturally sees its sales rise on impact. Moreover, its relative price continues to decline as long as its nominal price remains fixed. These results occur because other firms eventually increase their price so that the price level rises. Firms that do adjust their price immediately following the money growth innovation set their price high enough so that their sales initially fall. In the aggregate, however, the latter firms typically represent a small fraction of the total number of firms and aggregate sales initially rise. Looking at the point estimates of the sales response to a money

[6] See Shapiro and Watson (1988) for details of how to estimate just-identified SVARs using an instrumental variables approach.

Figure 2



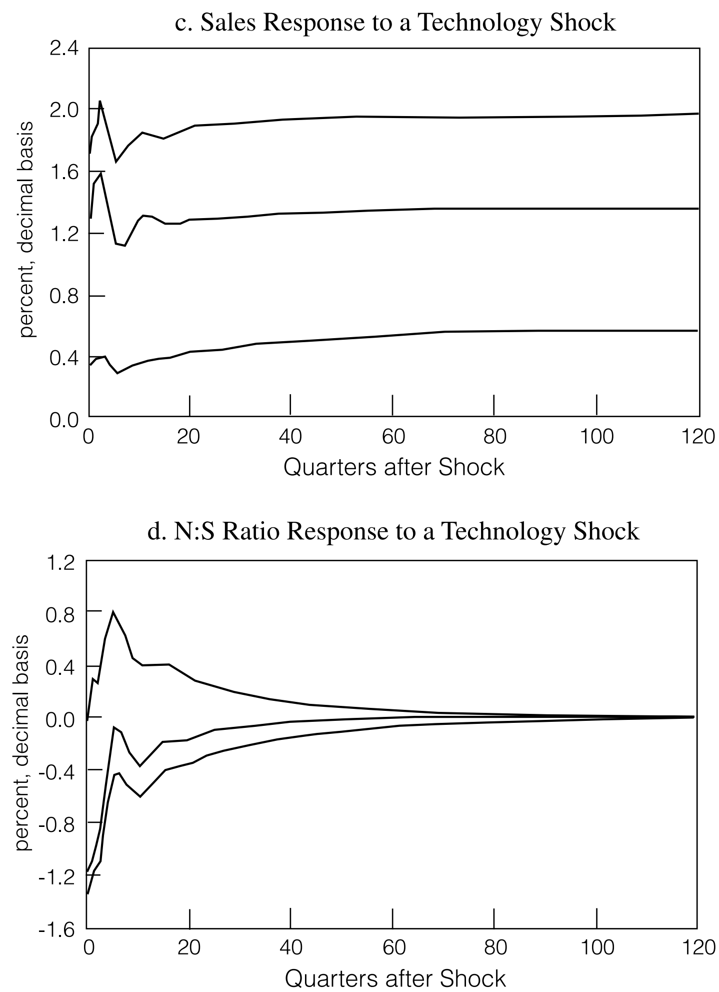
innovation in Figure 2a, we see that sales actually fall when the shock occurs. However, immediately following this initial response, sales increase before reverting back to the steady state. This dynamic adjustment in sales, therefore, is almost compatible with the predicted response in Hornstein and Sarte (1998). The main difference lies in the contemporaneous response that appears negative in the data. On the one hand, this difference may be evidence that a relatively nontrivial fraction of firms actually do adjust their price at the time of the shock. On the other hand, the upper bound of the confidence interval suggests a positive initial response of sales as expected. Furthermore, the subsequent dynamic adjustment in sales is consistent with that of a sticky-price model.

We now turn to the dynamic response of the inventory:sales ratio to a money growth innovation. In theory, the combination of production smoothing and sticky-price forces predicts that inventories should rise on impact in response to a positive nominal shock. Because the inventory:sales ratio is constant in the steady state, and nominal shocks have no long-run effect on sales, inventories then gradually fall back so as to meet some target inventory:sales ratio. Alternatively, changes in the inventory:sales ratio fall back to zero. To understand the nature of this dynamic adjustment, recall that a firm that does not adjust its price in response to a nominal shock initially experiences a rise in sales. Afterwards, sales continue to rise as long as its price remains unchanged. Given that this firm also smooths production over its pricing cycle, it must initially increase production by *more* than sales. This large initial increase in production effectively allows output to grow relatively slowly over the remainder of the firm's pricing cycle. Therefore, firms that keep their price fixed following a money growth shock increase their inventory holdings at the outset. Now, what about firms that do change their price at the time the shock occurs? Since these firms also smooth production over their pricing cycle, they initially reduce output by *less* than the fall in sales they experience. Consequently, inventory holdings increase for the latter firms as well. In the aggregate, therefore, inventory holdings should *unambiguously rise* on impact in response to a positive nominal demand shock. Looking at the response of the inventory:sales ratio to a money shock in Figure 2b, we see that it rises on impact by approximately 1 percent. Since sales contemporaneously fall by 0.4 percent in response to the same shock, the level of inventories does indeed rise at the outset by about 0.6 percent as suggested by our sticky-price framework.[7]

When examining the dynamic adjustment of sales and inventories to a technology shock, Hornstein and Sarte (1998) suggest that total sales should contemporaneously rise in response to a technology shock. This result is mainly driven by the firms that respond to the innovation. In particular, a productivity increase implies a fall in the marginal cost of production. Firms that immediately respond to the shock, therefore, lower their price and see their sales increase. Furthermore, during the transition, aggregate sales continue to rise monotonically to a higher steady state as more firms also reduce their price. The sales response in Figure 2c indeed broadly suggests that sales first increase in response to a technology shock and eventually reach a new higher steady-state level. The dynamic adjustment, however, is not monotonic. Specifically, sales appear to overshoot the new steady state twice during the early portion

[7] Letting $n{:}s$ denote the inventory:sales ratio, observe that the change in inventories is then given by $\Delta n = \Delta n{:}s + \Delta s = 0.01 - 0.004 = 0.06$.

Figure 2



of the transition phase. This oscillatory impulse response in sales is somewhat difficult to reconcile with a standard sticky-price model. It may suggest that some firms find it difficult to know exactly where to set a new price following the shock. In particular, the overshooting suggests that these firms may initially set their price too low. The subsequent corrective rise in price that would then occur causes a temporary decline in sales. It remains that, as expected, sales ultimately rise in the long run relative to their initial level.

As with the dynamic adjustment to a monetary innovation, the response of inventories to a technology shock hinges on the production-smoothing behavior of firms. Consider first the behavior of firms that adjust their price immediately following the shock. As we have just seen, these firms initially lower their price so that their sales at first increase. However, these firms then face declining sales over the remainder of their pricing cycle. This result stems from the fact that, following the initial adjustment, their price remains fixed while the price level continues to fall. Therefore, firms that adjust their price on impact also raise production but by a lesser amount than the initial sales increase. These firms consequently experience a fall in inventory holdings.

For the firms that do not adjust their price at the time the shock occurs, sales initially decrease as adjusting firms cause the aggregate price level to fall. Since these firms anticipate further declines in sales while their price remains fixed, they reduce production on impact by more than the initial decline in sales. Therefore, inventory holdings contemporaneously fall for the latter firms as well. It follows that aggregate inventory holdings should *unambiguously decline* immediately following the technology shock. As sales eventually rise to a higher steady state, inventories should then rise by the same amount in the long run to keep the inventory:sales ratio constant.

When we examine the inventory:sales ratio response to a technology shock in Figure 2d, we see that it falls by approximately 0.8 percent on impact in response to the innovation. Given the 1.2 percent rise in sales that contemporaneously follows the same technology shock, inventories then rise by about 0.4 percent at the time the shock hits. Since, on the contrary, a framework with sticky prices predicts an unambiguous initial decline in inventory holdings, the implied initial reaction of inventories in the data represents evidence against such a framework. However, we note that the lower bound of the 95 percent confidence interval for the impact response of sales is relatively small at about 0.35 percent. Because of the contemporaneous fall in the inventory:sales ratio by 0.8 percent, the level of inventories would also fall if we were to use the lower bound on the contemporaneous sales response. The latter observation mitigates the evidence against a sticky-price framework implied by the point estimates.

Thus far, the dynamic adjustment of sales and the inventory:sales ratio to both nominal demand and technology shocks are roughly consistent with what might have been predicted from a rigid price framework. The two main exceptions are (1) the implied initial response of inventories to technology shocks, and (2) the extremely sluggish dynamic adjustment of both sales and the inventory:sales ratio to shocks, as shown in Figure 2. In the case of the inventory:sales ratio's response to a money innovation, for instance, the half-life of the impulse is approximately 25 quarters or more than six years. While typical sticky-price models deliver nowhere near this kind of sluggishness in real variables, Ramey and West (1997) also note that the inventory:sales relationship exhibits a very high degree of persistence. In fact, these findings turn out to be a reflection of a well-known problem in the empirical literature on inventory behavior. Specifically, Feldstein and Auerbach (1976) point out early on the incongruity inherent to the notion that firms may take years to adjust to a sales shock, while the widest swings in inventory levels seldom amount to more than a few days' production. More recently, Blinder and Maccini (1991, p. 81) write that "one major difficulty with stock-adjustment models is that adjustment speeds turn out to be extremely low," a comment referring to the estimation of stock-adjustment equations generally.[8] They further note that "a natural reaction is that the slow estimated adjustment speeds must be an artifact of econometric biases. One potential source of such bias is omitted variables." As with the estimation of stock-adjustment equations, we should be conscious that the structural equations we estimate may also be subject to the latter source of bias.

Forecast Error Variance Decompositions

Having investigated the way in which the variables in (9) empirically respond to various structural shocks, we now wish to gauge the importance of each of these shocks in determining short-run variations in the data. We have seen that the dynamic adjustment of sales and the inventory:sales ratio, and hence inventories, to a money shock is generally consistent with the predictions of a sticky-price framework in which firms also smooth production. In some sense, however, this concept may be of secondary importance to a monetary policymaker if money shocks only play a small role in determining real variables. King, Plosser, Stock, and Watson (1991), for example, present compelling evidence to that effect in the case of aggregate variables. Of primary importance is the role played by each structural shock in determining short- and medium-run fluctuations in the data, as reflected when decomposing the variance of the k-step-ahead forecast errors.

Consider the moving-average process given by (9) and let $T(L) = T_0 + T_1 L + T_2 L^2 + \ \ldots \ + T_k L^k + \ \ldots \ ,\ L^j x_t = x_{t-j}$, while $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$. Then, we may write the k-step-ahead forecast error in $\mathbf{Y}$ as

$$\mathbf{Y}_{t+k} - E_{t-1}\mathbf{Y}_{t+k} = \sum_{j=0}^{k} T_k \varepsilon_{t+k-j}. \tag{11}$$

For our purposes, what we wish to assess is the fraction of variance in the left-hand side of equation (11) that is attributable to each of the structural shocks. In other words, we ask the question: To the degree that the actual data differ from the optimal forecast, which of the structural shocks is most

[8] See Lovell (1961) for instance.

Table 2 Decompositions of Forecast Error Variance

a. Fraction of Sales Forecast Error Variance Attributed to Shocks

Horizon	Technology Shock	Money Shock	Real Demand Shock
1	0.70 (0.21)	0.22 (0.16)	0.08 (0.13)
4	0.92 (0.17)	0.06 (0.12)	0.02 (0.12)
8	0.97 (0.13)	0.02 (0.09)	0.01 (0.07)
12	0.99 (0.11)	0.01 (0.08)	0.00 (0.05)
16	0.99 (0.09)	0.01 (0.07)	0.00 (0.05)
20	0.99 (0.07)	0.01 (0.05)	0.00 (0.02)
∞	1.00 (0.00)	0.00 (0.00)	0.00 (0.00)

b. Fraction of Inventory:Sales Ratio Forecast Error Variance Attributed to Shocks

Horizon	Technology Shock	Money Shock	Real Demand Shock
1	0.90 (0.27)	0.08 (0.26)	0.02 (0.21)
4	0.85 (0.26)	0.01 (0.25)	0.14 (0.24)
8	0.79 (0.22)	0.01 (0.25)	0.20 (0.24)
12	0.75 (0.22)	0.05 (0.24)	0.20 (0.24)
16	0.73 (0.21)	0.06 (0.25)	0.21 (0.24)
20	0.72 (0.21)	0.09 (0.24)	0.18 (0.24)
∞	0.69 (0.20)	0.14 (0.24)	0.17 (0.25)

responsible for this difference? Note that our identifying restrictions imply that 100 percent of the sales forecast error variance is explained by the technology shock at the infinite horizon. At shorter horizons, however, both nominal and real demand disturbances are allowed to contribute to fluctuations in sales. Table 2, panel a, shows that in fact, this contribution is relatively minor.[9] At the one-quarter horizon, technology shocks already explain 70 percent of the forecast error variance in sales. The bulk of the remaining variance is attributable to money growth shocks while real demand disturbances play a very small role. At the four-quarter horizon, technology shocks account for 92 percent of the variation in sales. At the three-year horizon, virtually all of the forecast error fluctuations in sales can be explained by technology shocks.

Focusing on fluctuations in the inventory:sales ratio, we again find that they tend to be dominated by real disturbances. In this case, however, it is interesting that as the forecast horizon lengthens, the important role played by technology

[9] Standard errors are in parentheses.

innovations diminishes somewhat at the expense of real demand disturbances. Also, by contrast to sales above, forecast errors in the inventory:sales ratio are not restricted to be uniquely driven by real disturbances in the long run. As a result, we find that at the infinite horizon, monetary disturbances explain approximately 14 percent of the forecast error in the inventory:sales ratio. While this number may not be too consequential, it is slightly larger than most other findings concerning the role of nominal shocks in determining the behavior of real variables. Gali (1992), for example, finds that after 20 quarters, money supply shocks only explain 9 percent of the variation in aggregate output.

Historical Decompositions

The variance decompositions in Table 2 show the relative importance of each structural shock in explaining variations in both sales and the inventory:sales ratio *on average*. It is also interesting to note that these shocks may matter more or less during various historical episodes. Figures 3 and 4 plot the historical forecast error decompositions in sales and the inventory:sales ratio at the 12-quarter horizon. This 12-quarter horizon concept of the business cycle is adopted from King, Plosser, Stock, and Watson (1991).

Figure 3 confirms that while money shocks have historically played a small role in explaining fluctuations in sales, technology shocks have played a more substantial role. Interestingly, this finding appears to remain consistent throughout the entire sample period considered. Temporary real demand disturbances take on relatively more importance in explaining sales fluctuations in the 1990s. On the whole, the largest forecast errors occur in the mid-1970s and, as might have been expected, coincide with the oil price shock of 1973.

The latter observation also applies to the forecast errors in the inventory:sales ratio as suggested by Figure 4. Again we note that money generally plays a small role in driving inventory:sales ratio fluctuations, as implied by the variance decompositions in Table 2. In contrast, we also find that the importance of the monetary component, even if small on average, noticeably increases in the early 1990s. Put another way, Figure 4 suggests that even if monetary fluctuations have traditionally represented a small portion of fluctuations in the inventory:sales ratio, this does not imply that monetary disturbances are always unimportant.

4. CAUTIONARY REMARKS

An important part of the empirical analysis above has been the assumption that the inventory:sales ratio is stationary around a constant mean. As we have seen, various cointegration tests have generally confirmed this hypothesis for manufacturing inventories. Moreover, the notion of a stationary ratio is typically explained on the grounds that stock-outs are costly and, therefore, that

Figure 3 Historical Forecast-Error Decomposition: Sales

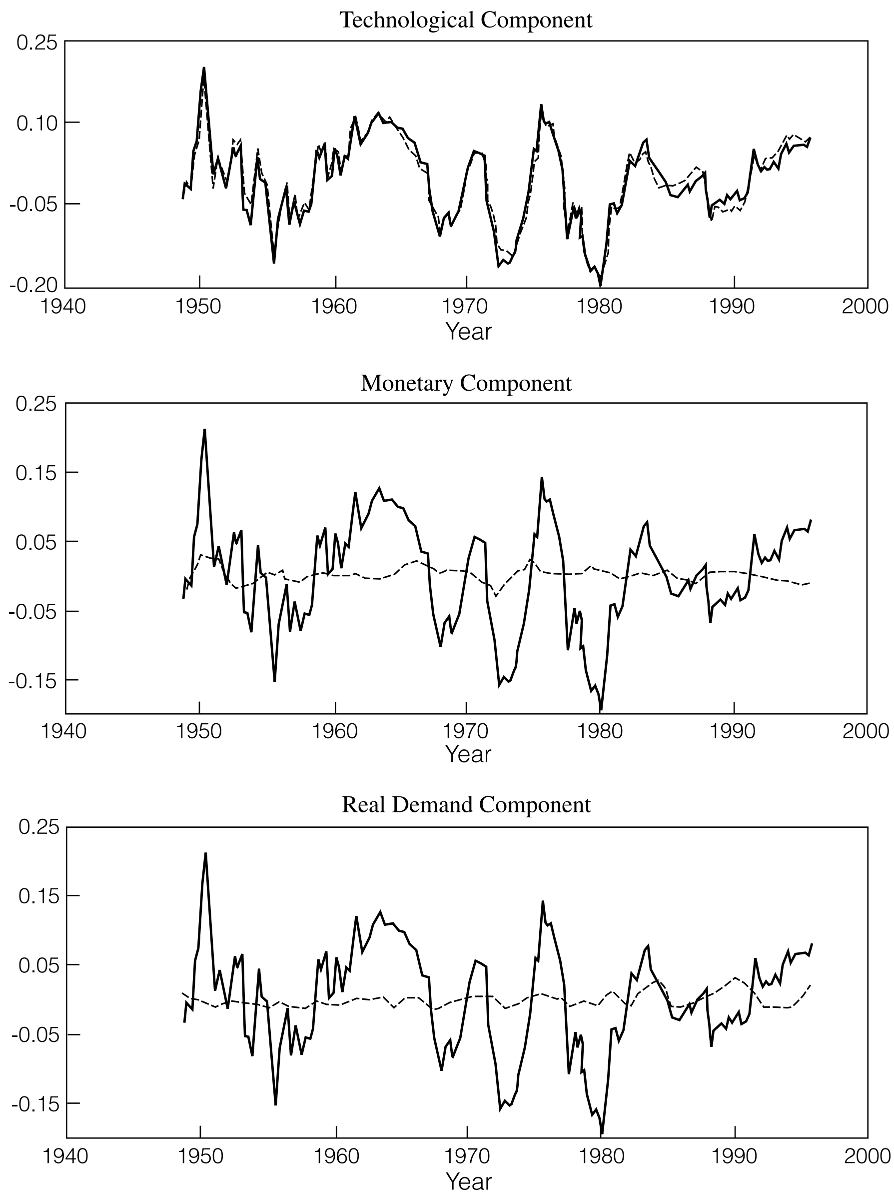


Figure 4 Historical Forecast-Error Decomposition: N:S Ratio

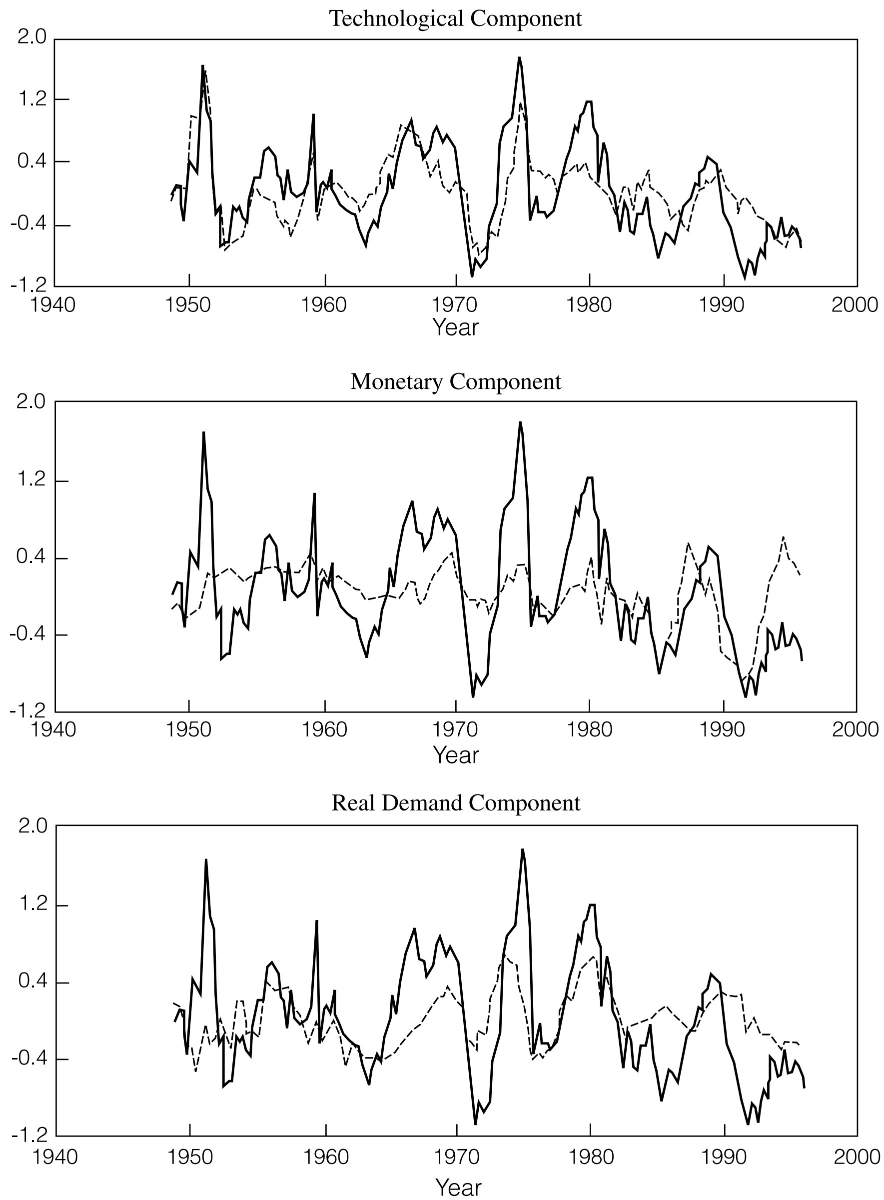
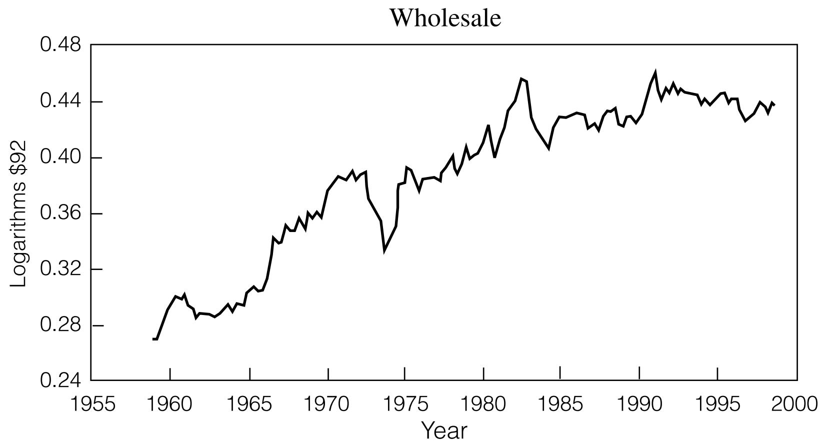
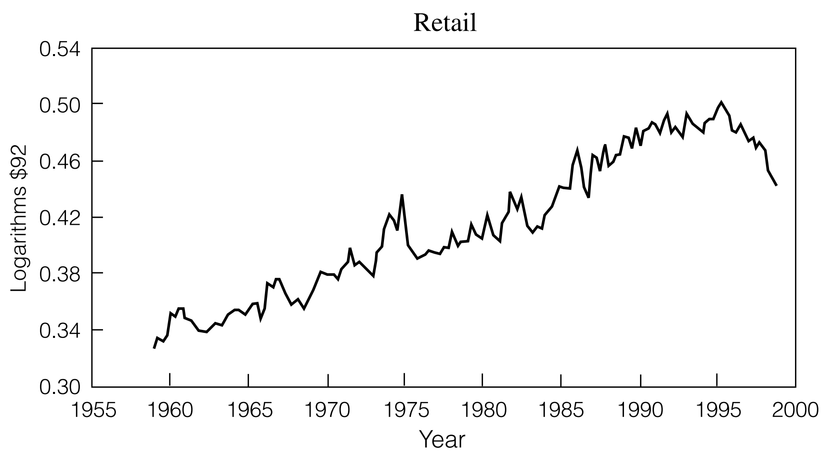


Figure 5 Logarithm of the Inventory:Sales Ratio (*n*−*s*)





firms generally try to meet some target inventory:sales ratio in the long run. The fall in the inventory:sales ratio that begins in the early 1990s (Figure 1) is sometimes taken as evidence of the just-in-time inventory method taking hold in the United States. The inventory:sales ratio in both the wholesale and retail sectors, however, reveals a much different story. Figure 5 suggests that for much of the period under consideration, the inventory:sales ratio in both these sectors has actually trended *upwards.* Both the ratios seem to stabilize in the early 1990s, perhaps again because of the widespread emergence of the just-in-time method. Nevertheless, it remains that much of the increase in the inventory:sales ratio up to the early 1990s, in both the wholesale and retail sectors, represents somewhat of a puzzle.

To explain this puzzle, one can speculate that, over time, consumers have gained easier access to a wide variety of goods through improved means of communication and transportation. As a result, back orders for any one business are less likely to arise since consumers can simply acquire the same goods elsewhere. So not having goods on hand more readily results in lost sales, which effectively drives up the cost of stock-outs and, consequently, inventory:sales ratios. Alternatively, consistent improvements in technology may simply have reduced storing costs over time. This would have made it easier for wholesalers and retailers to avoid stock-outs and is directly consistent with increasing inventory:sales ratios. Food products, for instance, have become increasingly storable because of consistent innovations in preservatives technology. Whatever the case may be, Figure 5 makes it clear that traditional theories of inventory behavior need to be amended to account for the data in the wholesale and retail sectors. Perhaps a focus away from production smoothing is even necessary.

5. CONCLUSIONS

We have used an SVAR to acquire some insight into the dynamic responses of manufacturing sales and inventories to both nominal demand and real supply shocks. We assumed that money is neutral in the long run and, moreover, that the inventory:sales ratio can be properly characterized as a stationary process without trend. We then found that the estimated dynamic adjustments to nominal demand and real supply shocks are generally consistent with those of an equilibrium model of inventory behavior with inflexible prices. However, the degree of sluggishness exhibited by both sales and the inventory:sales ratio, and hence inventories, in response to these shocks is much greater than that suggested by current sticky-price models. The latter findings confirm earlier observations by Blinder and Maccini (1991) and, more recently, Ramey and West (1997).

We also used our empirical framework to gauge the relative importance of both nominal and real disturbances as sources of fluctuations in the manufacturing sector. The results indicate that nominal shocks generally contribute little to the forecast error variance in both sales and the inventory:sales ratio at all horizons. Instead, fluctuations in real variables are mainly driven by real disturbances. In addition, the latter results appeared to hold consistently at the business cycle frequency throughout the sample period under consideration.

REFERENCES

Blanchard, Olivier J., and Danny Quah. “The Dynamic Effects of Aggregate Demand and Supply Disturbances,” *American Economic Review*, vol. 79 (September 1989), pp. 655–73.

Blinder, Alan S., and Louis J. Maccini. “Taking Stock: A Critical Assessment of Recent Research on Inventories,” *Journal of Economic Perspectives*, vol. 5 (Winter 1991), pp. 73–96.

Cooley, Thomas F., and Mark Dwyer. “Business Cycle Analysis without Much Theory: A Look at Structural VARs,” *Journal of Econometrics*, vol. 83 (March–April 1998), pp. 57–88.

Feldstein, Martin, and Alan Auerbach. “Inventory Behavior in Durable Goods Manufacturing: The Target-Adjustment Model,” *Brookings Papers on Economic Activity*, 2:1976, pp. 351–96.

Gali, Jordi. “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?” *American Economic Review*, vol. 89 (March 1999), pp. 249–71.

__________. “How Well Does the IS-LM Model Fit Postwar U.S. Data?” *Quarterly Journal of Economics*, vol. 107 (May 1992), pp. 709–38.

Goodfriend, Marvin, and Robert G. King. “The New Neoclassical Synthesis and the Role of Monetary Policy,” in Ben Bernanke and Julio Rotemberg, eds., *NBER Macroeconomic Annual 1997*. Cambridge, Mass.: MIT Press, 1997.

Hornstein, Andreas. “Inventory Investment and the Business Cycle,” Federal Reserve Bank of Richmond *Economic Quarterly*, vol. 84 (Spring 1998), pp. 49–72.

__________, and Pierre-Daniel G. Sarte. “Staggered Prices and Inventories: Production Smoothing Reconsidered,” Working Paper 98–8. Richmond: Federal Reserve Bank of Richmond, December 1998.

Johansen, Soren. “Statistical Analysis of Cointegration Vectors,” *Journal of Economic Dynamics and Control*, vol. 12 (April 1988), pp. 231–54.

Khan, James A. “Inventories and the Volatility of Production,” *American Economic Review*, vol. 77 (September 1987), pp. 667–79.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson. “Stochastic Trends and Economic Fluctuations,” *American Economic Review*, vol. 81 (September 1991), pp. 819–40.

Lovell, Michael C. “Manufacturers’ Inventories, Sales Expectations, and the Acceleration Principle,” *Econometrica*, vol. 29 (July 1961), pp. 293–314.

Ramey, Valerie A., and Kenneth D. West. "Inventories," *NBER* Working Paper 6315. December 1997.

Saikkonen, P. "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory*, vol. 7 (March 1991), pp. 1–21.

Shapiro, Mathew D., and Mark W. Watson. "Sources of Business Cycle Fluctuations," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1988.* Cambridge, Mass.: MIT Press, 1988.

Stock, James H., and Mark W. Watson. "A Simple MLE of Cointegrating Vectors in Higher Order Integrated Systems," *NBER* Technical Working Paper 83. December 1989.

__________. "Testing for Common Trends," *Journal of the American Statistical Association,* vol. 83 (December 1988), pp. 1097–1107.