

Changes in Behavioral and Characteristic Determination of Female Labor Force Participation, 1975–2005

JULIE L. HOTCHKISS

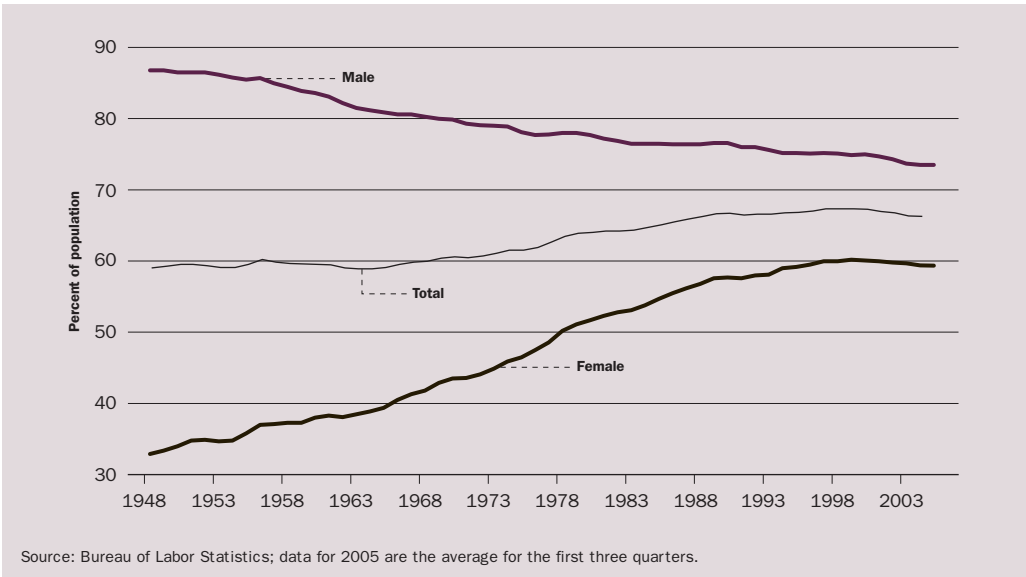
The author is a research economist and policy adviser in the regional group of the Atlanta Fed's research department and an adjunct professor of economics in the Andrew Young School of Policy Studies at Georgia State University. She thanks M. Melinda Pitts, Robert E. Moore, and John C. Robertson for helpful conversations and comments.

Since the late 1940s the percent of the male population participating in the labor force has steadily declined while female labor force participation has steadily increased (see Figure 1). A variety of factors have been found to have contributed significantly to the decline in male labor force participation: the institution of Social Security in 1935; its expansion to include disability insurance and Medicaid; and the Revenue Act of 1942, which granted tax incentives for firms to establish private pension plans (for example, see Burtless and Moffitt 1984; Cremer, Lozachmeur, and Pestieau 2004; Gruber 2000; and Lumsdaine, Stock, and Wise 1997). These policies provided men greater incentives both to claim a work-inhibiting disability and to retire earlier from the labor market. Another explanation for the decline in male labor force participation among all age groups is the increase in female labor force participation. With labor supply decisions often made in a household (husband-wife) setting, the increase in family income from more wives working provides an income effect incentive for husbands to decrease their labor supply.

The rise in female labor force participation has several explanations as well. A major determinant is the stream of biotechnological advancements that have provided women greater control over and timing of childbearing decisions since the 1940s (see Bailey 2004). This greater flexibility, along with advancements in household technologies (such as the introduction of the dishwasher and the microwave oven), has afforded women greater freedom and time to increase their educational attainment, providing yet another reason to devote more time to the labor market (see Goldin 1995). Further, changing social attitudes about the role of women and the appropriateness of women (and wives) working have increased job opportunities and, thus, incentives for women to enter the labor market (see Rindfuss, Brewster, and Kavee 1996).

While the ongoing decline in male labor force participation and the long-lived rise in female labor force participation have received much attention over the years,

Figure 1
Labor Force Participation Rates over Time, 1948–2005



more recent changes in the trend of labor force participation among women since the mid-1990s beg further scrutiny. The growth in female labor force participation began to flatten out in 1997 and has been declining since 2000. The purpose of this article is to dissect the changes in labor force participation decisions that have taken place among women aged twenty-five to fifty-four over the past thirty years.¹

Identifying the factors contributing to observed changes in labor force participation trends over time (particularly those affecting the recent decline) may help anticipate future changes in those trends. An important component of policymakers' expectations regarding productivity or output potential of the United States, and thus appropriate policy action, is the formation of expectations regarding available labor input, or the size of the workforce.² The results in this article suggest that the decline in female labor force participation rates between 2000 and 2005 was not entirely a response to a predictable change in macroeconomic conditions or to demographic changes. Consequently, a reversal is not obviously forthcoming or likely to be easily predictable.

Bradbury and Katz (2005) seem to have made the only investigation of the potential sources of the recent decline in female labor force participation; they identify the decline as being concentrated among more highly educated married women with young children.³ The analysis in this article delves deeper to disentangle changes in characteristics from changes in behavior, with given characteristics, of women over a long period of time. The results suggest that while changes in both observed characteristics and behavior have contributed to the decline in female labor force participation since 2000, unobserved—and thus unpredictable—changes are the largest contributors. The analysis also indicates that while the higher average unemployment rate in 2005 has put downward pressure on the labor supply of women, if the unemployment rate were to regain its 2000 level, women's labor force participation rate (keeping everything else at its 2005 level) would still be significantly lower than it was in 2000.

Theoretical and Empirical Construct

The labor-leisure choice model assumes that a person chooses a combination of hours of leisure and income (or an aggregate consumption bundle) in order to maximize utility. There is a trade-off between leisure and income in that consumption of more leisure (less work) results in less income. This utility maximization problem has a corner solution in which the person chooses to consume the maximum number of leisure hours possible (work zero hours).⁴ The decision to work (or participate in the labor market) boils down to the evaluation of what the market is willing to pay a person for his or her time relative to the value that person's time generates (in terms of additional utility) when consumed as leisure. This labor force participation decision can be expressed mathematically as

$$(1) \quad W_i - MRS_{i,H=0} \begin{cases} > 0 \Rightarrow LFP = 1 \\ \leq 0 \Rightarrow LFP = 0 \end{cases},$$

where W_i is the market wage that person i can earn in the market, $MRS_{i,H=0}$ is person i 's reservation wage (the utility gained at zero hours of work), and LFP is a binary choice variable that is equal to 1 if the person is a labor force participant and equal to 0 if the person is not in the labor force.

This theoretical construct translates into an operational estimation framework by assuming that the difference between a person's market wage and reservation wage can be represented by a linear function of observable characteristics about that person and an unobservable random component:

$$(2) \quad I_i^* = W_i - MRS_{i,H=0} = \beta_0 + \beta_1' X_{W,i} + \beta_2' X_{R,i} + \varepsilon_i = \begin{cases} > 0 \Rightarrow LFP = 1 \\ \leq 0 \Rightarrow LFP = 0 \end{cases}.$$

$X_{W,i}$ is a vector of observable characteristics that determine what wage person i could expect to earn in the market. One of the most important human capital characteristics determining labor market earnings is the woman's education level. Labor market experience is also important and will be proxied by age. Age squared is also included as a regressor to capture the concavity of the experience/age-labor force participation profile. Because living with a disability increases the cost (*ceteris paribus*) of participating in the labor market and may reduce the market wage available (see Hotchkiss 2003, chap. 3), a variable indicating the amount of disability income being received (if any) is also included as a regressor. The current labor market condition

-
1. While the behaviors of younger and older women deserve their own analyses, this article focuses on the change in labor force participation of the women who make up the bulk of the female labor force: those twenty-five to fifty-four years of age. The women in this age group made up roughly 68 percent of the female labor force in 2005. Kirkland (2002) has concluded that an increased emphasis on schoolwork (rather than working while attending school) has contributed significantly to the decline in teen labor force participation. DiNatale (2005a) attributes the rise in older women's labor force participation to the decline in retirement portfolios in 2001 and to better health care and longevity.
 2. Bradbury (2005) explores the implication of the lower labor force participation for assessing production slack in the economy.
 3. DiNatale (2005b) provides a much more cursory glance at the decline in labor force participation among working-aged women.
 4. There is another corner solution in which the person chooses to work the maximum number of hours possible, but this solution is considered practically infeasible.

is also important in determining the value of entering the labor market; as the probability of obtaining a job declines, the expected value of the market wage declines. The state unemployment rate will serve as a proxy for current labor market conditions.

$X_{R,i}$ is a vector of observable characteristics that determine the value of person i 's time out of the labor market. Factors that are expected to affect the value of a

This analysis suggests that while changes in observed characteristics and behavior have contributed to the decline in female labor force participation since 2000, unobserved—and thus unpredictable—changes are the largest contributors.

woman's time out of the labor market include whether she is married, how many children she has, and the amount of income she has access to in the absence of her earnings (nonlabor income, including any spousal earnings). In addition to the X variables already discussed, indicators for the woman's race are included to capture any differential labor market returns experienced across racial groups (for example,

as a result of discrimination) and cultural or social differences that might affect the marginal valuation of time spent out of the labor market.

ϵ_i is the random component, and assuming that $\epsilon_i \sim N(0, 1)$ means that the parameter coefficients in equation (2) are determined via maximum likelihood (ML) probit estimation.⁵

Data and Estimation Strategy

The previous theoretical construct indicates that changes in observed labor force participation rates can arise from three sources. One source is change in characteristics. For example, a woman's characteristics may change by her having children (which would be expected to raise her reservation wage, *ceteris paribus*) or by her attaining more education (which would raise her expected market wage). These changes in characteristics would be reflected in changes in the X s. While the unemployment rate is not a characteristic of the woman making the labor force participation decision *per se*, it is a characteristic of the environment in which the decision is made.

A second source of change is a change in behavior—a change in the way a woman's characteristics translate into her observed labor market participation decision. These changes will be reflected in changes in the estimated parameter coefficients, given a specific set of characteristics. Changes in parameter coefficients in a labor force participation equation can be thought of as reflecting changes in the marginal utility generated by the characteristics. For example, if the additional utility from participating in the labor market as a married woman increases (say, as the result of a decrease in relative market returns for men), then the parameter coefficient on the marriage indicator variable will increase. Or, if discriminatory behavior against women declines, the labor market return to a college degree might increase, raising the marginal utility from participating in the labor market for a woman with a college degree. This change would manifest itself in an observed change in behavior among college women (greater labor force participation) and a larger positive parameter coefficient on the college degree indicator variable. A change in the responsiveness to labor market conditions can also affect observed labor force participation. A change in responsiveness will be reflected through a change in the estimated parameter coefficient associated with the state unemployment rate.

The third source for change in labor force participation decisions is the force of unobservables. Innumerable factors enter into a woman's decision to participate in the labor market that are not observed and manifest themselves in the estimated intercept term. These factors might include changes in women's preferences not captured by

observables or changes in the labor market structure or institutions that affect the labor market valuation of human capital characteristics and thus the market wage. Unfortunately, this third source does not typically result in transparent policy implications.

The March Current Population Survey (CPS) from the Bureau of Labor Statistics (BLS) is used to evaluate changes in labor force participation behavior of women between the ages of twenty-five and fifty-four. The data cover the years 1976 through 2005, with the analysis focusing on the period 2000 through 2005. These data are used for two primary reasons. First, the BLS uses these data to estimate and report the labor force participation rate. Second, the data provide a consistent, long-running, and large sample on which to obtain parameter estimates. These data are cross-sectional, so separate labor force participation equations will be estimated for each year to “decompose” the changes in the labor force participation rate into changes in behavior (differences in estimated parameter coefficients across years) and changes in characteristics (differences in regressor values across years).⁶

Results

The 1994–2005 period. Table 1 presents sample means and estimated parameter coefficients from the ML probit labor force participation estimation for the years 2000 and 2005. This table provides the first clues about how changes in characteristics and behavior (*ceteris paribus*) have affected labor force participation decisions between 2000 and 2005. For example, women in 2005 were slightly more likely to have at least a college degree than they were in 2000. Given that more education increases the returns to supplying labor, this increase in education raises the probability of being in the labor force. However, the responsiveness of labor force participation to education declined slightly from 2000 to 2005. In other words, education (both college and high school) was providing less of a pull into the labor market in 2005 than in 2000, and this factor put downward pressure on labor force participation decisions.

Furthermore, the higher average state unemployment rate in 2005 put downward pressure on labor force participation; the negative parameter coefficient in 2000 on the state unemployment rate translates into a 1.9 percentage point decline in labor force participation for every 1 percent increase in the unemployment rate.⁷ But women were also apparently less sensitive to labor market conditions in 2005 (evidenced by the smaller negative parameter coefficient), meaning that a higher unemployment rate had less of an effect on labor force participation in 2005 relative to the effect it would have had in 2000; this smaller parameter coefficient translates

5. In a model of labor force participation of women, it might be prudent to model that decision jointly with that of her spouse (if married); for example, see Hotchkiss, Kassis, and Moore (1997). In the present article, it is assumed that labor supply decisions are made at an individual, rather than family, level, and the labor supply of other family members enters into a woman's labor force participation decision in the form of higher nonlabor income. A joint labor force participation analysis will be the subject of future research.

6. An alternative strategy might be to construct a synthetic panel of cohorts to determine whether the observed behavior is the result of collective changes within a certain group of women. A typical cohort definition is based on year of birth. The additional information that cohort identification might provide to the analysis was explored, but it was determined that, except for those older than fifty-five years, group behavior did not vary significantly across the sample time period. All analyses are performed using the March supplement weight because this is the only weight that is valid since 2002 and because some of the regressors come from the supplemental part of the survey. The results are essentially unchanged if the analyses are performed unweighted.

7. This marginal effect is calculated for every woman and then averaged across the sample.

Table 1
**Sample Means and Maximum Likelihood Probit Estimates for 2000 and 2005,
 Women Aged Twenty-five to Fifty-four**

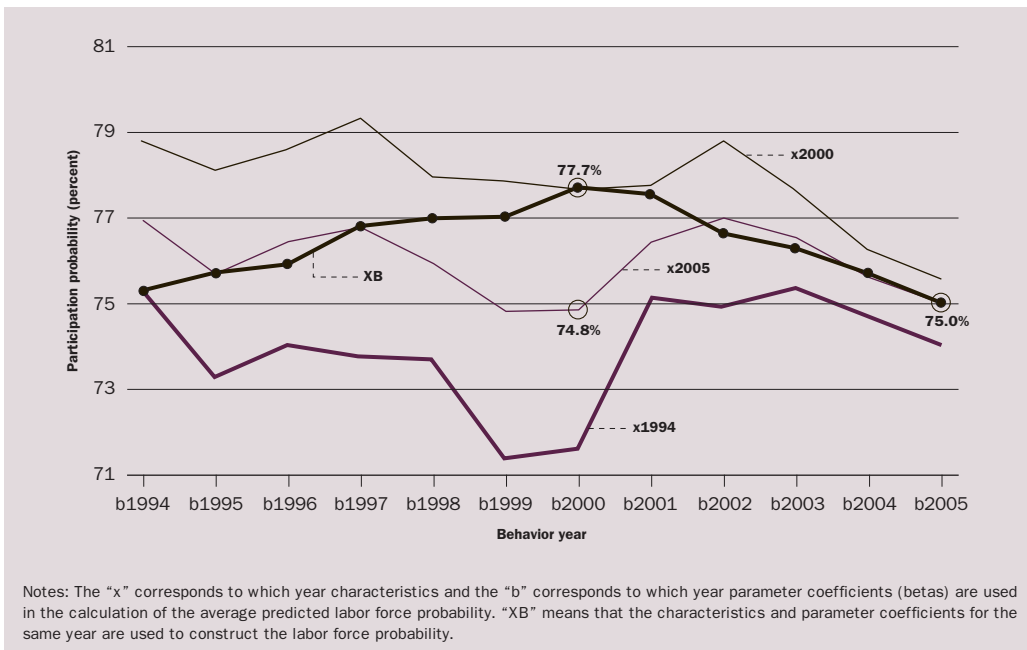
Characteristic	Weighted sample means		Dependent variable = probability of labor force participation	
	2000	2005	2000	2005
Age	39.4	39.7	0.0501 (0.0125)	0.4723 (0.0101)
Age squared	1,618.43	1,649.56	-0.0007 (0.0002)	-0.0006 (0.0001)
# children YT 6	0.285	0.293	-0.3661 (0.0163)	-0.3168 (0.0128)
# children aged 6–18	0.729	0.718	-0.0654 (0.0102)	-0.0385 (0.0081)
Married, spouse present	63.4%	62.6%	-0.1188 (0.0231)	-0.1122 (0.0190)
High school graduate	60.9%	58.9%	0.6763 (0.0275)	0.6236 (0.0230)
College degree or more	28.2%	30.4%	1.0051 (0.0327)	0.9695 (0.0264)
Nonlabor income (per yr.)	\$43,496	\$43,047	-0.0028 (0.0002)	-0.0022 (0.0001)
Hispanic	5.4%	7.3%	-0.1795 (0.0430)	-0.2115 (0.0287)
Black	13.4%	13.3%	0.0166 (0.0309)	0.0165 (0.0244)
Disability income (per yr.)	\$68.20	\$67.57	-0.0001 (0.00002)	-0.0001 (0.00002)
State unempl. rate	4.45%	5.79%	-0.0698 (0.0110)	-0.0109 (0.0077)
Intercept	—	—	-0.0091 (0.2418)	-0.3897 (0.1988)
N	29,718	46,862		

Notes: Means in bold are significantly different from one another at least at the 95 percent confidence level across years. All parameter coefficients are significantly different from one another across years except for the coefficients on black. All coefficients are significantly different from zero at least at the 95 percent confidence level except black and intercept (2000) and black and the state unemployment rate (2005). All dollar values are inflated to 2004 values using the consumer price index.

into a 0.3 percentage point decline in labor force participation for every 1 percent increase in the unemployment rate.

To determine how much of the observed decline in labor force participation among women is due to changes in characteristics and behavior (parameter coefficients), labor force participation in each year is simulated using a common set of parameter coefficients. For example, the average probability of women in 2005 participating in the labor market is calculated assuming the women in that year behaved as women did in, say, the year 2000. The deviation in the simulated labor force participation (using 2005 women's characteristics and 2000 parameter coefficients) and

Figure 2
**Actual and Simulated Labor Force Participation of Women
 Aged Twenty-five to Fifty-four, 1994, 2000, and 2005**



the actual labor force participation in 2005 indicates how much of the observed difference in labor force participation between 2000 and 2005 was due to changes in behavior and how much was due to changes in characteristics. This decomposition technique is subsequently described in more detail.

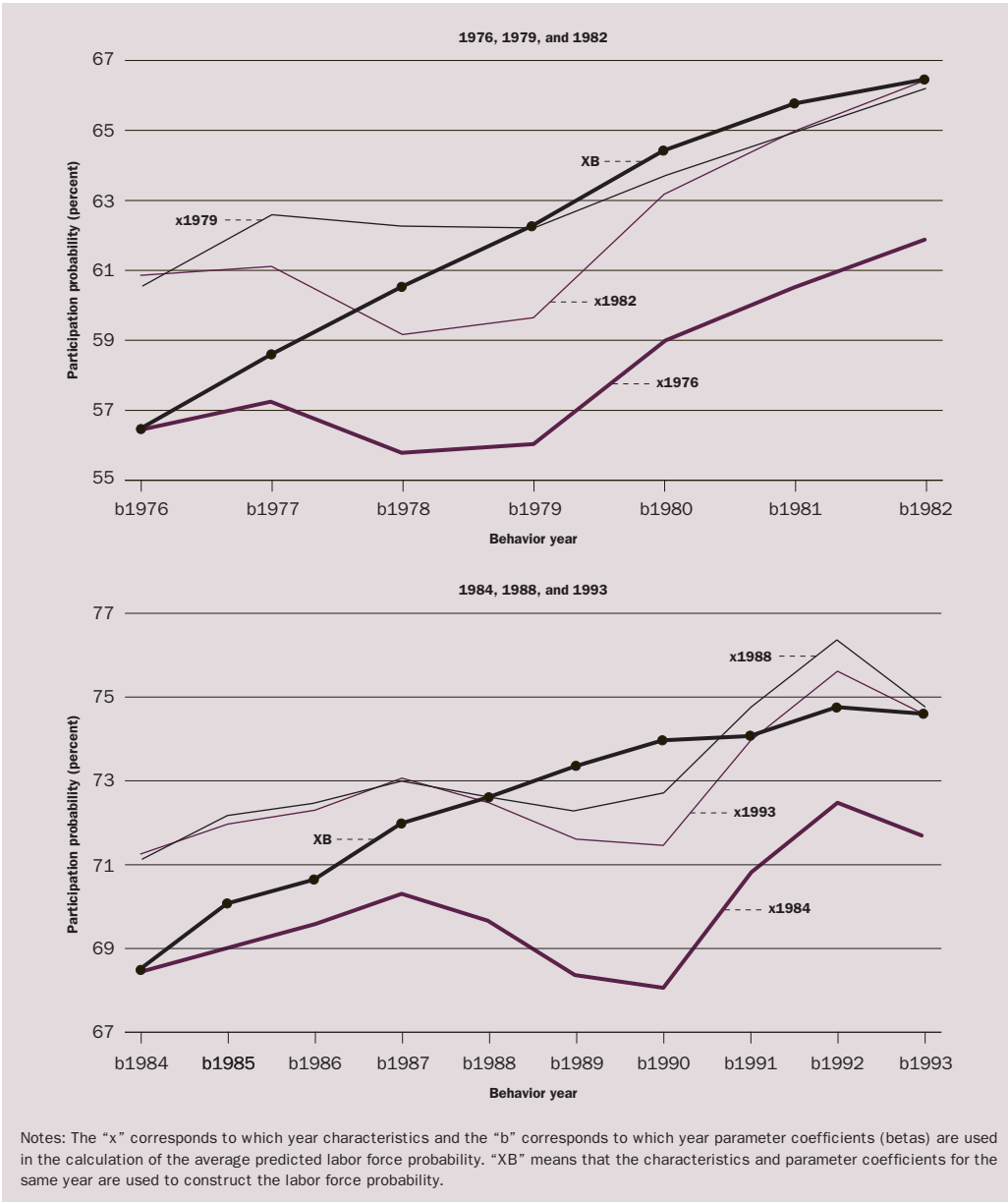
Figure 2 plots selected results from this simulation exercise. First, separate probit models were estimated for each year between 1994 and 2005 in order to generate year-specific parameter coefficients ($\hat{\beta}^t$).⁸ These year-specific parameter coefficients were then combined with each year's sample characteristics (X_i^t) to simulate the expected labor force participation decision of women in each year, given the behavior across different years.

The line labeled "XB" reflects the labor force participation predicted for each year's sample of women, given their own parameter coefficients (the average across i of $\Phi(X_i^t \hat{\beta}^t)$ for the sample of women in year t , where Φ is the standard normal cumulative distribution function). This sample average probability of participating in the labor market is analogous to the population labor force participation rate. The line labeled "x1994" reflects the predicted labor force participation for the 1994 sample of women in each year (t), assuming that they behaved as the women in year t behaved (this is the average across i in year t of $\Phi(X_i^{1994} \hat{\beta}^t)$). The line labeled "x2000" is the average of $\Phi(X_i^{2000} \hat{\beta}^t)$, and the line labeled "x2005" is the average of $\Phi(X_i^{2005} \hat{\beta}^t)$.

The distance between two lines reflects differences in the characteristics of women holding the responsiveness to those characteristics (or behavior) fixed. For

8. The comparison goes back only to 1994 in this figure since a major CPS questionnaire change in 1994 has been shown to have changed the classification of female labor force participants. Analyses of earlier years are subsequently explored.

Figure 3
Actual and Simulated Labor Force Participation of Women
Aged Twenty-five to Fifty-four, 1976–93



Notes: The “x” corresponds to which year characteristics and the “b” corresponds to which year parameter coefficients (betas) are used in the calculation of the average predicted labor force probability. “XB” means that the characteristics and parameter coefficients for the same year are used to construct the labor force probability.

example, the vertical distance between the x1994 line and the x2000 line measures the difference in predicted labor force participation between 1994 and 2000 that is accounted for by differences in women’s characteristics in those two years. Moving along any of the lines shows how the labor force participation decisions of any one sample of women would have changed given the estimated parameter coefficients across the years; this pattern indicates the importance of changes in behavior across the years in determining labor force participation.

Comparing the predicted and simulated labor force participation probabilities shows that the bulk of the decline in predicted labor force participation between 2000 and 2005 derived from a change in characteristics rather than a change in behavior. In other words, from 2000 to 2005 characteristics changed in such a way as to reduce labor force participation from 77.7 percent to 74.8 percent (a 2.9 percentage point decline, holding behavior constant at 2000 values). Furthermore, behavior changed in such a way as to put upward pressure on labor force participation from 74.8 percent (2005 women behaving like 2000 women) to 75.0 percent (a 0.2 percentage point increase). These influences resulted in a net decline in labor force participation between 2000 and 2005 of 2.7 percentage points.⁹ Labor force participation decreased slightly over the entire 1994–2005 time period by 0.2 of a percentage point. Over this entire time period, characteristic changes alone would have increased predicted labor force participation by 1.7 percentage points, while behavioral changes alone would have reduced labor force participation by 1.9 percentage points. In other words, between 1994 and 2005 the characteristics of women (for example, fewer children, less likely to be married, more likely to be a college graduate) put upward pressure on labor force participation, but behavioral changes added downward pressure.

The movements of labor force participation rates of men and women since the late 1990s have taken more parallel, rather than converging, paths.

The 1970s and 1980s. The results for the 1994–2005 time period are in sharp contrast to the changes observed for women during the 1970s and 1980s. Figure 3 depicts the same simulation described above for different year groups: 1976 to 1982 and 1984 to 1993.¹⁰ Two main observations are worth highlighting from these graphs. First, both sets of years saw a much stronger impact on labor force participation from both characteristic changes (movement between the lines) and behavioral changes (movement along each line). Second, the slowing in the impact of both characteristic and behavioral changes is evident in the early 1990s.¹¹

Table 2 summarizes the changes depicted in Figures 2 and 3. The first column of numbers shows the net change in labor force participation rates, the second column shows how labor force participation would have altered from changes in characteristics only (holding coefficients constant), and the third column shows how labor force participation would have altered from changes in behavior only (measured by differences in parameter coefficients). For the 1970s and 1980s, both behavior and characteristics changed in such a way as to put upward pressure on labor force participation, with changes in behavior contributing the most to the net change over the

-
9. The algebraic decomposition of the net change in the expected labor force participation between year t and year $t - k$ is expressed as $\Phi(X_t^i \hat{\beta}^t) - \Phi(X_{t-k}^i \hat{\beta}^{t-k}) = [\Phi(X_t^i \hat{\beta}^t) - \Phi(X_{t-k}^i \hat{\beta}^{t-k})] + [\Phi(X_{t-k}^i \hat{\beta}^{t-k}) - \Phi(X_{t-k}^i \hat{\beta}^t)]$. The decomposition can be performed using different endpoints as the base case. Here the latter year (t) is used as the point of reference. Generally, the conclusions are essentially the same, even if some of the details of the decomposition may vary, if the earlier year is used as the point of reference.
10. The analyses for these earlier years differ from the 1994–2005 analysis only in that there is no regressor for disability income; this variable is not available in the CPS until 1989. The analysis does not start earlier than 1976 because a person's nonlabor income is not calculable (without matching respondent records) prior to this survey year. Also, 1983 is excluded because of erratic estimation results.
11. Sample means and estimated parameter coefficients for select years are contained in the appendix (Table A1).

Table 2

Net Changes in Predicted Labor Force Participation and the Contribution of Changes in Behavior and Characteristics, Women Aged Twenty-five to Fifty-four

Year span	Net change in predicted labor force participation (percentage points)	Portion of net change accounted for by changes in characteristics (percentage points)	Portion of net change accounted for by changes in behavior (percentage points)
1976–82	+10.0	+4.4	+5.6
1984–93	+6.1	+2.8	+3.3
1994–2005	-0.2	+1.7	-1.9
2000–05	-2.7	-2.8	+0.1

periods.¹² In the 1990s, however, while characteristic changes (for example, higher educational attainment, less marriage, fewer children) continued to put upward pressure (albeit by a smaller amount) on labor force participation, behavioral changes more than offset those characteristic changes by contributing downward pressure.

Comparison with men. One interpretation of the observed recent decline in female labor force participation is that women are losing ground in their efforts to compete with men and to make comparable contributions to the labor market (for example, see Bradbury and Katz 2005). An alternative interpretation, given that the labor force participation rate of men has been declining steadily for decades (see Figure 1), is that women have achieved as much parity regarding labor force participation decisions as they and their partners want and that the same forces driving labor force participation of men downward are now acting upon those decisions of women. For example, the income effect of rising real wages may now dominate the substitution effect for women. One way to explore how similar the recent experience of women is to that of men is to perform the same analysis described previously for samples of men over the same time periods and to compare those results to those obtained for women. Table 3 decomposes the net changes in male labor force participation across different time periods into the contributions made by changes in characteristics and changes in behavior.

The first thing to notice in this table is that behavioral changes have consistently contributed downward pressure and have been largely responsible for the decline in the labor force participation of men over each period. While women’s behavior changed through the 1970s and 1980s to provide upward pressure on labor force participation, their behavior changed in the 1990s to resemble the changes occurring in men’s behavior across all periods, putting downward pressure on participation decisions. By contrast, however, characteristic changes of men continued to push participation downward, while the characteristics of women continued to contribute positively to their labor force participation (except in just the most recent years). The net result is that the movements of labor force participation rates of men and women since the late 1990s have taken more parallel, rather than converging, paths.

Which characteristics? The bulk of the decline in labor force participation rates among both men and women between 2000 and 2005 is accounted for by changes in characteristics. The difference in average characteristics between 2000 and 2005 in Table 1 suggests which characteristic changes might have been most influential in lowering women’s labor force participation between these two years.¹³ Specifically, declines in characteristics positively influencing labor force participation or increases

Table 3

Net Changes in Predicted Labor Force Participation and the Contribution of Changes in Behavior and Characteristics, Men Aged Twenty-five to Fifty-four

Year span	Net change in predicted labor force participation (percentage points)	Portion of net change accounted for by changes in characteristics (percentage points)	Portion of net change accounted for by changes in behavior (percentage points)
1976–82	-0.14	-0.04	-0.10
1984–93	-1.26	-0.12	-1.14
1994–2005	-1.48	-0.11	-1.37
2000–05	-1.78	-1.54	-0.24

in characteristics negatively influencing labor force participation are candidates. These would include the declines in the percent of high school graduates and of blacks, the increases in number of children younger than six years and the percent of Hispanics, and the rise in the unemployment rate.

Figure 4 simulates the hypothetical question, What would the average probability of labor force participation have been in 2005 if women's characteristics indicated on the horizontal axis equaled the average for women in 2000 (holding all other characteristics at their 2005 levels)? This simulation shows that the most important characteristics contributing to the decline in the labor force participation rate between 2000 and 2005 are the rise in the number of children under age six and the unemployment rate.¹⁴ However, even if the unemployment rate were to regain its 2000 level of 4.5 percent, women's labor force participation would rise (keeping all other characteristics at their 2005 levels) only to 75.4 percent, which is still 2.3 percentage points below the labor force participation rate for women in 2000.

Which behaviors? The results in Table 2 indicate that the combined behavioral changes between 2000 and 2005 among women put a slight upward pressure on their labor force participation decisions. But some behavioral changes individually contributed to the lower observed labor force participation in 2005. The estimation results for 2000 and 2005 in Table 1 show that those variables on which the estimated coefficients became less positive or more negative are the candidates for having lowered labor force participation. For example, the smaller positive coefficient on both the high school and college education dummy variables means that education had less of a pull into the labor market for women in 2005 than it did in 2000. Other factors that lowered the predicted labor force participation for women include being Hispanic, being black (very slightly), the impact of disability income (very slightly), and the intercept term. In contrast, between 2000 and 2005 predicted labor force participation increased (*ceteris paribus*) among women who are married, have children, have more nonlabor income, and face stronger local labor market conditions (the state unemployment rate).

12. Blau and Kahn (2005) document and investigate the source of changes in hours of work among women between 1980 and 2000. They also identify behavioral changes as the major contributor over that time to changes in hours of work.

13. The sample means and estimated parameter coefficients for men for selected years are presented in the appendix, Table A2.

14. However, the sample mean numbers of children under age six are not significantly different from one another in 2000 and 2005 (see Table 1).

Figure 4
Simulated Labor Force Participation Probabilities for Women with 2005 Characteristics

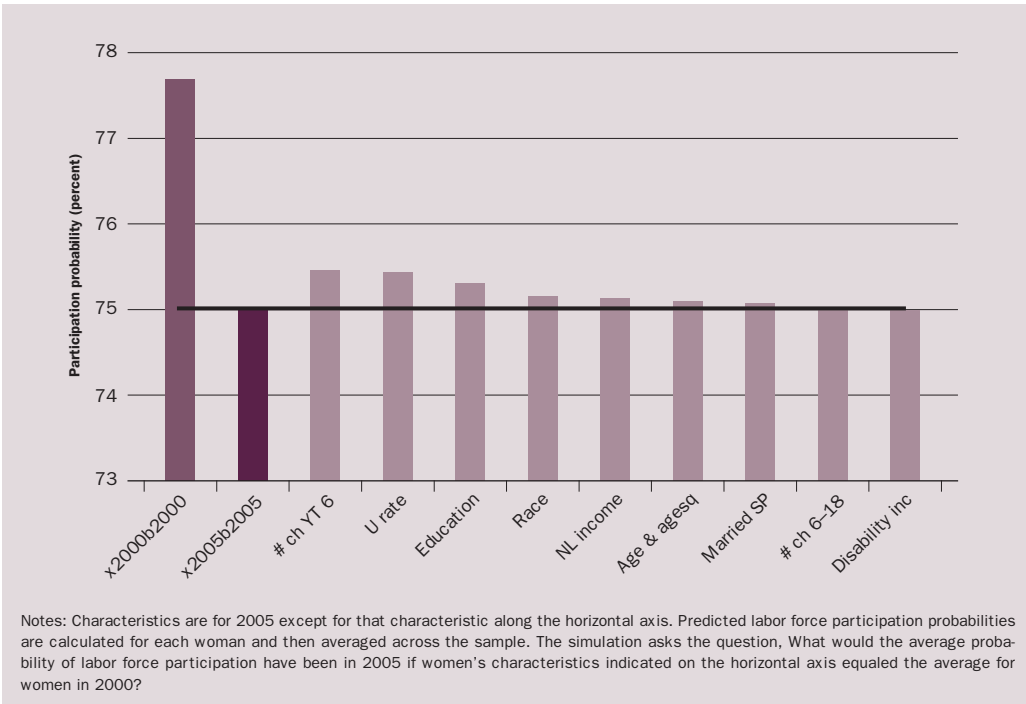
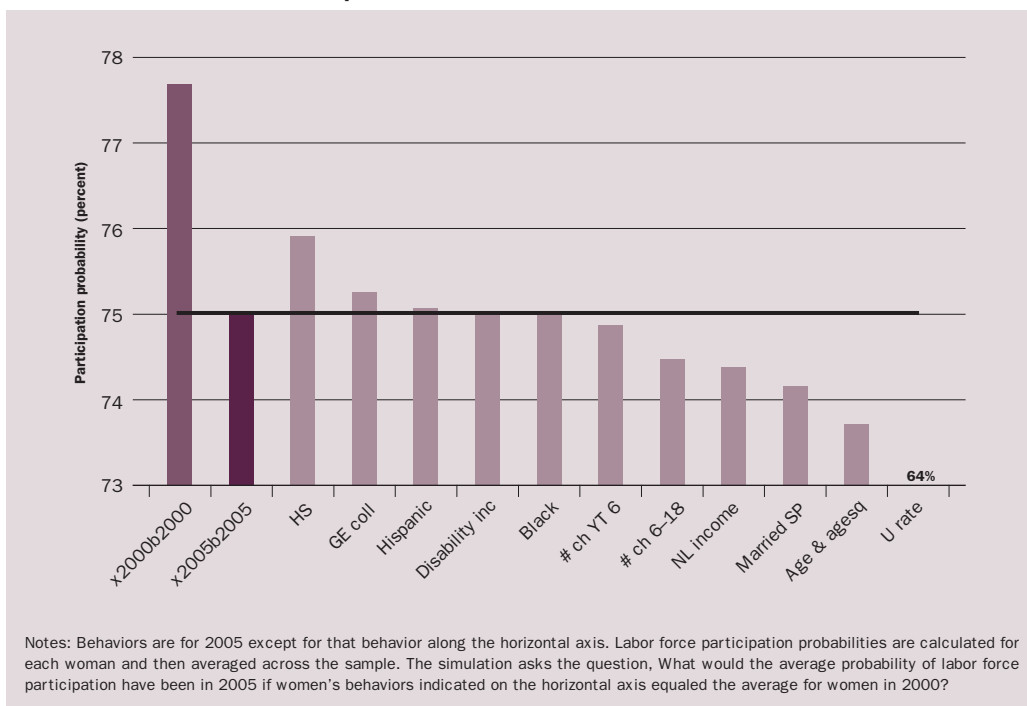


Figure 5 plots the predicted labor force participation rate in 2000 and in 2005 for women along with simulated 2005 labor force participation that would result from changing one coefficient at a time. The hypothetical is, What would the average probability of labor force participation have been in 2005 if women's behavior matched that in 2000 with regard to the characteristics indicated along the horizontal axis (keeping all other behavior at its 2005 level)? For example, if college graduates responded in 2005 as they did in 2000, leaving everything else about 2005 women and their behavior unchanged, the labor force participation rate would have been 75.3 percent instead of the actual 75 percent. (The combined effect of 2000 behavior among both high school and college graduates would have resulted in a labor force participation rate of 76.2 percent.) Regarding women's response to labor market conditions, if women in 2005 were as sensitive to changes in the unemployment rate as they were in 2000, the labor force participation rate would have actually been only 64 percent in 2005.

Like changes in characteristics, no single observed behavioral change or combination of behavioral changes can account for the full 2.7 percentage point drop in the participation rate from 2000 to 2005. The reason is that behavioral changes unexplained by observed factors are largely responsible for the observed decline in labor force participation rates. These unexplained behavioral changes manifest themselves in the estimate of the intercept term. Replacing the estimated 2005 intercept term with that estimated for 2000 (leaving all other parameter coefficients and characteristics at their 2005 values) results in a predicted labor force participation rate of 85 percent. Unfortunately, there is no way to know exactly what the intercept term is capturing.¹⁵

Figure 5

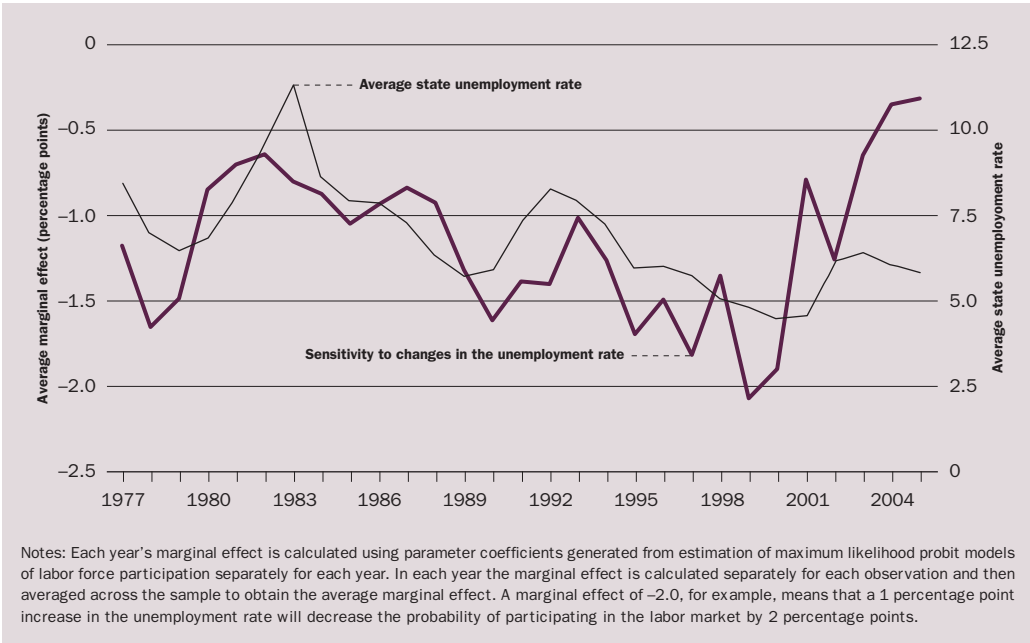
Simulated Labor Force Participation Probabilities for Women with 2005 Behavior

Some studies have demonstrated that labor force participation rates have fallen particularly dramatically among college-educated, married women (for example, Bradbury and Katz 2005). Certainly, highly educated women are more likely to marry highly educated men (Hotchkiss and Pitts 2005; Neal 2004; Herrnstein and Murray 1994) and would therefore likely feel more secure in leaving the labor market (with their husband's higher earning power and lower probability of unemployment). This observation suggests that the interaction of some of the regressors (for example, education and marriage) may help reduce some of the intercept's explanatory power. Several alternative specifications and interactions were explored, resulting in practically no change in the results. Also, recall that the nonlabor income measure includes a woman's spouse's earnings (if she is married). Therefore, it seems that any change in preferences between 2000 and 2005 that may be reflected in the estimates of the intercept term of the regression is not particularly correlated with changes in behavior related to marriage, children, or educational attainment.

A closer look at labor market conditions. The conclusion from the preceding two sections is that no single observable characteristic or behavioral change would

- With unobservable factors making such a large contribution to observed behavior, a natural question is how well the model fits the data. Comparing actual (reported by the BLS) labor force participation rates for women aged twenty-five to fifty-four, deviations of the average predicted labor force participation in each year range from 0.02 to 0.4 of a percentage point with the median deviation at 0.1 of a percentage point. The large contribution of the intercept term does not imply a poor fit of the model; it just means there are deviations in labor force participation probabilities across time that cannot be identified through changes in observed characteristics.

Figure 6
Average Marginal Effect of a 1 Percentage Point Change in the Unemployment Rate on the Probability of Participating in the Labor Market, Women Aged Twenty-five to Fifty-four, 1977–2005



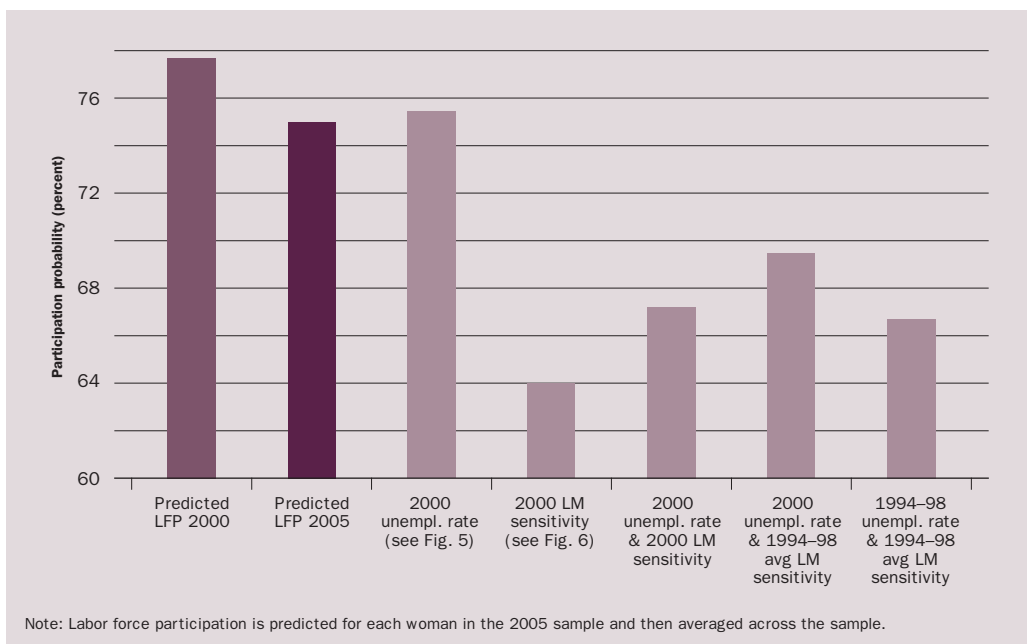
overwhelm the influence of the change in unobservables between 2000 and 2005 in order to return the labor force participation rate of women to the level seen in 2000. Two of the most dramatic differences between these years, however, include the higher unemployment rate and the lower sensitivity to labor market conditions in 2005 relative to 2000. Looking more closely at the sensitivity to labor market conditions, Figure 6 plots the average marginal effect of a 1 percentage point change in the state unemployment rate on the probability of participating in the labor force in each year from 1977 to 2005.¹⁶ On the secondary axis the figure also plots the average of the state unemployment rate in each year.

A few striking observations can be made from Figure 6. First, as expected, the marginal effects are always negative; an increase in the unemployment rate reduces a woman's likelihood of participating in the labor market.¹⁷ Second, the years 1999 through 2005 were unique for women relative to earlier years. In no other years was women's sensitivity to labor market conditions as strong as in 1999 and 2000. And in no other years was women's sensitivity weaker than in 2004 and 2005. These years were also notable for their relatively low (1999 and 2000) and relatively high (2004 and 2005) average state unemployment rates. These observations lead to the third observation that labor market sensitivity tends to be countercyclical, demonstrating less sensitivity (smaller negative number) during years of higher unemployment rates.

What would be the impact on 2005 labor force participation if both the unemployment rate and labor market sensitivity returned to their 2000 levels? Figure 7 plots the actual labor force participation rates in 2000 (77.7 percent) and in 2005 (75 percent), plus what the labor force participation rate would be assuming 2005 characteristics and behavior, except for changing the unemployment rate or women's sensitivity to labor market conditions. The first simulation is the same as that in Figure 5: a 75.4 percent

Figure 7

Average Predicted Labor Force Participation (LFP) in 2000 and 2005 and Simulated LFP Assuming 2005 Characteristics, Except as Noted, Women Aged Twenty-five to Fifty-four



labor force participation rate if only the unemployment rate returned to its 2000 level. The second simulation is what is seen in Figure 6: a 64 percent labor force participation rate if only the responsiveness of women to changes in the unemployment rate returned to its 2000 level. The third simulation indicates that if both the unemployment rate and women's labor market sensitivity returned to their 2000 levels, the labor force participation rate would be 67.2 percent.

Because both the low unemployment rate and strong labor market sensitivity in 2000 are unique, assuming a return to the environment and behavior that existed that year may not be realistic. To illustrate what might be a more reasonable future path of labor force participation rates for women, Figure 7 presents two additional scenarios. The first assumes the low 2000 unemployment rate but applies a 1994-98 average labor market sensitivity. Under this scenario, the labor force participation of women (assuming other factors remain at their 2005 values) would be 69.4 percent, still below the 75 percent labor force participation rate of 2005. The second, even more conservative, scenario assumes a 1994-98 average unemployment rate and average labor market sensitivity. In this case, labor force participation would be only 66.7 percent.

While both the influence of current labor market conditions and the responsiveness of women to those conditions are relatively strong, the simulations in Figure 7

16. States were not all individually identified in the CPS until 1977. The earlier analyses average the unemployment rates for those states that were combined in 1976. Also, throughout this article the state unemployment rate for March was used. The results are nearly identical if a lagged unemployment rate (for example, February) was used.

17. The model structure assumes a symmetric labor force participation response to increases and decreases in the unemployment rate.

are consistent with the earlier conclusion that even significant changes among observable factors will not overcome the gap that emerged between labor force participation rates in 2000 and 2005.

Conclusions

After decades of consistent increases, the labor force participation of women began to flatten out in the late 1990s and decline after 2000. This article investigates the changes in the labor force participation rate among women aged twenty-five to fifty-four that have occurred over the past thirty years. For each decade between 1975 and 2005, the changes in labor force participation rates are decomposed into the portions

Since 2000, the labor force participation behavior of women appears to be moving in parallel to that of men albeit with a significant gap in labor force participation rates that may not ever close.

explained by changes in either women's characteristics or women's behavior across each decade. Characteristics that have traditionally pushed women out of the labor market (number of children, being married, low education) have declined each decade, but by smaller and smaller amounts. More influential than changing characteristics, however, has been changing behavior, pulling

women into the labor market for any given set of characteristics. The rate of change in behavior has also declined over the past thirty years. Indeed, behavioral change between 1994 and 2005 had a direct negative influence on the observed decline in the labor force participation rate during these decades.

Special attention is given to the unprecedented 2.7 percent decline in the labor force participation rate between 2000 and 2005, which can be explained by changes in both behavior and characteristics, with weaker labor market conditions in 2005 being one of the characteristics providing the greatest downward pressure. However, if the unemployment rate had been at the 2000 level of 4.5 percent in 2005 (holding everything else at 2005 levels), the labor force participation rate of women would still have been 2.3 percentage points lower than in 2000. Other characteristic changes that contributed to the decline in female labor force participation included a lower percent of high school graduates, a greater percent of Hispanic women, and women having more children under the age of six, on average.

Among observable behavioral changes, the largest contributor to the labor force participation rate decline between 2000 and 2005 was the weaker pull of education into the labor market. The lower probability of Hispanic and black women participating in the labor market in 2005 and the stronger push of disability income on women out of the labor market also contributed to the decline. One behavior that changed quite dramatically between 2000 and 2005 was women's sensitivity to labor market conditions. In 2000 a 1 percentage point drop in the unemployment rate (*ceteris paribus*) would have led to an increase in labor force participation of 1.9 percentage points, whereas the same drop in the unemployment rate in 2005 would have led to only a 0.3 percentage point increase in the labor force participation rate. The combined impact of the labor market returning to its average 1994–98 conditions and the average sensitivity to those conditions would result in a labor force participation rate of 66.7 percent, lower than in 2005.

While changes in specific, observable behavior and characteristics can be identified as contributing to the decline in female labor force participation since 2000, it is important to realize that there remains one key determinant of labor force participation decisions about which the analysis in this article has little to say—that is, unobservable,

or unexplained, behavior. Even though clearly observable factors have worked to reduce female labor force participation, other observable factors have been working to increase it. For example, women's responses to changes in nonlabor income, marriage, children, and labor market conditions have provided less of a push out of the labor market in 2005 than in 2000. Taking into account the impact of all observable characteristics and behavioral changes between 2000 and 2005, it is the change in unobservable factors that has had the strongest impact on the decline in labor force participation rates over this period. The large role that unobservables play in the determination of labor force participation is not unique to the 2000–05 period, nor is it unique to women.

The presence of unobservables is not very satisfying or informative from a policy perspective. Nonetheless, their large role in the determination of labor force participation rates suggests that a rebound of the labor market to the environment that existed in 2000 is not likely to cause female labor force participation to rebound to 2000 levels without changes in unobservable factors that cannot be predicted. The pull of college education into the labor force apparently began to weaken for women in the late 1990s but has slowed in its decline. Also, the push for women out of the labor force caused by marriage and children continues to weaken, as does the push resulting from higher nonlabor income. Indeed, it is striking how, since 2000, the labor force participation behavior of women appears to be moving in parallel to that of men albeit with a significant gap in labor force participation rates that may not ever close. Further investigation of how labor force participation decisions are made in a family context and how these joint (spousal) decisions have changed over time is the next obvious step in the ongoing scrutiny of the declining labor force participation rates of women.

Appendix

Additional Tables

Table A1
Sample Means and Maximum Likelihood Probit Parameter Estimates
(and Standard Errors) for 1977, 1984, and 1994, Women Aged Twenty-five to Fifty-four

Characteristic	Weighted sample means			Dependent variable = probability of labor force participation		
	1977	1984	1994	1977	1984	1994
Age	38.3	37.4	38.3	0.0503 (0.0110)	0.0687 (0.0110)	0.0818 (0.0117)
Age squared	1,545.65	1,474.06	1,535.84	-0.0008 (0.0001)	-0.0011 (0.0001)	-0.0012 (0.0002)
# children YT 6	0.323	0.338	0.329	-0.5585 (0.0160)	-0.4924 (0.0149)	-0.4911 (0.0148)
# children aged 6-18	1.090	0.816	0.739	-0.0925 (0.0075)	-0.1318 (0.0089)	-0.1230 (0.0096)
Married, spouse present	75.8%	69.5%	64.5%	-0.2669 (0.0234)	-0.1567 (0.0228)	0.0333 (0.0232)
High school graduate	59.9%	63.0%	63.7%	0.4497 (0.0203)	0.6114 (0.0225)	0.7597 (0.0254)
College degree or more	14.8%	19.5%	23.7%	0.8041 (0.0297)	1.0166 (0.0301)	1.1540 (0.0318)
Nonlabor income (per yr.)	\$43,384	\$38,103	\$37,566	-6.6x10 ⁻⁶ (3.0x10 ⁻⁷)	-5.5 x10 ⁻⁶ (3.0x10 ⁻⁷)	-4.5x10 ⁻⁶ (3.1x10 ⁻⁷)
Hispanic	2.0%	3.1%	4.7%	0.0721 (0.0630)	-0.0914 (0.0501)	-0.1867 (0.0393)
Black	11.1%	11.9%	12.9%	0.1189 (0.0295)	0.0615 (0.0301)	-0.06181 (0.0292)
Disability income (per yr.)	—	—	\$60.82	—	—	-5.1x10 ⁻⁵ (0.0000)
State unempl. rate	8.46%	8.51%	7.18%	-0.0344 (0.0048)	-0.0281 (0.0040)	-0.0449 (0.0063)
Intercept	—	—	—	0.2711 (0.2110)	-0.1732 (0.2097)	-0.6303 (0.2269)
N	30,187	32,984	33,411			

Notes: All means across years are significantly different from one another at least at the 95 percent confidence level except age, age squared, and number of children younger than six (1977 versus 1994); and number of children younger than six, high school graduate, and nonlabor income (1984 versus 1994). The parameter coefficients are all significantly different from one another at the 95 percent confidence level. All parameter coefficients are significantly different from zero at least at the 95 percent confidence level except for the coefficient on Hispanic and the intercept (1977 and 1984) and married, spouse present (1994). All dollar values are inflated to 2004 values using the consumer price index.

Table A2
Sample Means and Maximum Likelihood Probit Parameter Estimates
(and Standard Errors) for 1977, 1984, 2000, and 2005, Men Aged Twenty-five to Fifty-four

Characteristic	Weighted sample means				Dependent variable = probability of labor force participation			
	1977	1984	2000	2005	1977	1984	2000	2005
Age	38.2	37.4	39.4	39.6	0.1144 (0.0174)	0.1026 (0.0163)	0.0263 (0.0172)	0.0557 (0.0140)
Age squared	1,541.75	1,467.60	1,617.29	1,639.64	-0.0016 (0.0002)	-0.0015 (0.0002)	-0.0006 (0.0002)	-0.0009 (0.0002)
# children YT 6	0.345	0.338	0.273	0.272	0.0503 (0.0288)	-0.0131 (0.0264)	0.1224 (0.0333)	0.0911 (0.0280)
# children aged 6–18	0.939	0.673	0.593	0.570	0.0068 (0.014)	-0.0090 (0.0161)	0.0686 (0.0187)	0.0899 (0.0161)
Married, spouse present	78.4%	70.1%	62.2%	59.9%	0.6289 (0.0375)	0.6372 (0.0346)	0.5041 (0.0328)	0.5586 (0.0289)
High school graduate	51.7%	55.2%	58.7%	57.4%	0.4369 (0.0324)	0.5328 (0.0324)	0.5251 (0.0343)	0.36153 (0.0300)
College degree or more	23.1%	26.5%	29.2%	29.6%	0.6520 (0.0436)	0.7957 (0.0418)	0.9313 (0.0430)	0.7012 (0.0370)
Nonlabor income (per yr.)	\$17,104	\$18,895	\$25,646	\$25,448	-7.9x10 ⁻⁶ (5.5x10 ⁻⁷)	-6.4x10 ⁻⁶ (5.3x10 ⁻⁷)	-3.6x10 ⁻⁶ (3.6x10 ⁻⁷)	-2.9x10 ⁻⁶ (2.7x10 ⁻⁷)
Hispanic	1.8%	2.8%	5.0%	7.1%	-0.4292 (0.0911)	-0.5172 (0.0675)	-0.1937 (0.0603)	-0.2512 (0.0390)
Black	9.5%	10.2%	11.6%	11.1%	-0.2230 (0.0441)	-0.2693 (0.0432)	-0.3309 (0.0398)	-0.3297 (0.0323)
Disability income (per yr.)	—	—	\$100.63	\$89.79	—	—	-0.0001 (-1.5x10 ⁻⁵)	-9.0x10 ⁻⁵ (-1.1x10 ⁻⁵)
State unempl. rate	8.4%	8.5%	4.4%	5.8%	-0.0135 (0.0084)	-0.04514 (-0.3912)	-0.0485 (0.0157)	-0.0552 (0.0107)
Intercept	—	—	—	—	-0.7469 (0.3340)	0.0064 (-0.3912)	0.8426 (0.3413)	0.2407 (0.2748)
N	27,698	30,216	27,874	42,138				

Notes: All means across years are significantly different from one another at least at the 95 percent confidence level except number of children younger than six (1977 versus 1984) and number of children younger than six, high school graduate, college degree or more, nonlabor income, and black (2000 versus 2005). The parameter coefficients are all significantly different from one another at the 95 percent confidence level. All estimated parameter coefficients are significantly different from zero at least at the 95 percent confidence level except for the coefficients on children younger than six, children aged six to eighteen, and the state unemployment rate (1977); children younger than six, children aged six to eighteen, and the intercept (1984); age (2000); and the intercept (2005). All dollar values are inflated to 2004 values using the consumer price index.

REFERENCES

- Bailey, Martha J. 2004. More power to the pill: The impact of contraceptive freedom on women's labor supply. Vanderbilt University, photocopy, April.
- Blau, Francine D., and Lawrence M. Kahn. 2005. Changes in the labor supply behavior of married women: 1980–2000. NBER Working Paper 11230, March.
- Bradbury, Katharine. 2005. Additional slack in the economy: The poor recovery in labor force participation during the business cycle. Federal Reserve Bank of Boston Public Policy Briefs 05-2.
- Bradbury, Katharine, and Jane Katz. 2005. Women's rise: A work in progress. Federal Reserve Bank of Boston *Regional Review* 14, no. 3:58–67.
- Burtless, Gary, and Robert A. Moffitt. 1984. The effect of Social Security benefits on the labor supply of the aged. In *Retirement and economic behavior*, edited by Henry J. Aaron and Gary Burtless. Washington, D.C.: Brookings Institution.
- Cremer, Helmuth, Jean-Marie Lozachmeur, and Pierre Pestieau. 2004. Social Security, retirement age, and optimal taxation. *Journal of Public Economics* 88, no. 11:2259–81.
- DiNatale, Marisa. 2005a. More on labor force participation. *DismalScientist*, March 8. <www.economy.com/dismal/pro/article.asp?cid=12672> (March 14, 2005).
- . 2005b. On labor force participation rates. *DismalScientist*, February 1. <www.economy.com/dismal/pro/article.asp?cid=11688> (March 14, 2005).
- Goldin, Claudia. 1995. The U-shaped female labor force function in economic development and economic history. In *Investment in human capital*, edited by T. Paul Schultz. Chicago: University of Chicago Press.
- Gruber, Jonathan. 2000. Disability insurance benefits and labor supply. *Journal of Political Economy* 108, no. 5:1162–83.
- Herrnstein, Richard J., and Charles Murray. 1994. *The bell curve: Intelligence and class structure in American life*. New York: Simon and Schuster.
- Hotchkiss, Julie L. 2003. *The labor market experience of workers with disabilities: The ADA and beyond*. Kalamazoo, Mich.: W.E. Upjohn.
- Hotchkiss, Julie L., Mary Mathewes Kassis, and Robert E. Moore. 1997. Running hard and falling behind: A welfare analysis of two-earner families. *Journal of Population Economics* 10, no. 3:237–50.
- Hotchkiss, Julie L., and M. Melinda Pitts. 2005. Female labour force intermittency and current earnings: Switching regression model with unknown sample selection. *Applied Economics* 37, no. 5:545–60.
- Kirkland, Katie. 2002. Declining teen labor force participation. *Issues in labor statistics*, Summary 02-06, September.
- Lumsdaine, Robin L., James H. Stock, and David A. Wise. 1997. Retirement incentives: The interaction between employer-provided pensions, Social Security, and retiree health benefits. In *The economic effects of aging in the United States and Japan*. Chicago: University of Chicago Press.
- Neal, Derek. 2004. The measured black-white wage gap among women is too small. *Journal of Political Economy* 112, no. 1, part 2:S1–S28.
- Rindfuss, Ronald R., Karin L. Brewster, and Andrew L. Kavee. 1996. Women, work, and children: Behavioral and attitudinal change in the United States. *Population and Development Review* 22, no. 3:457–82.

How Good Is What You've Got? DGSE-VAR as a Toolkit for Evaluating DSGE Models

MARCO DEL NEGRO AND FRANK SCHORFHEIDE

Del Negro is a research economist and assistant policy adviser in the Atlanta Fed's research department. Schorfheide is an associate professor at the University of Pennsylvania. The authors thank Tom Cunningham, Pedro Silos, and Ellis Tallman for helpful comments.

Dynamic stochastic general equilibrium (DSGE) models are becoming increasingly popular in central banking circles. The number of central bank-sponsored conferences on DSGE modeling and the amount of staff resources devoted to DSGE model development and estimation have risen dramatically over the past five years. This trend has affected monetary policy authorities around the globe, including the Federal Reserve System, the Bank of Canada, the European Central Bank, the Sveriges Riksbank, and the Reserve Bank of New Zealand. While few central banks are currently using DSGE models to generate forecasts and policy scenarios that provide the basis for interest rate decisions, many are contemplating doing so in the near future.

Part of the recent popularity of DSGE models is due to work by Smets and Wouters (2003), who document that a modified version of a New Keynesian model developed by Christiano, Eichenbaum, and Evans (2005) is able to track and forecast euro area time series as well as, if not better than, a vector autoregression (VAR) estimated with Bayesian techniques. While the empirical finding needs to be qualified, the results have had a considerable impact on how policymakers view DSGE models and have triggered efforts at many central banks to develop their own estimated DSGE model.¹

In the 1990s the prevailing view among some policymakers was that DSGE models provide "good theory" to sharpen the understanding of business cycle fluctuations and to address fundamental policy questions: How is the stabilization of output and inflation through monetary policy actions related to the maximization of aggregate welfare? Should the central bank react to asset market fluctuations? How should monetary policy be conducted if the nominal interest rate is close to its zero lower bound? Should central banks of small open economies respond to exchange rate movements? Despite these benefits, many policymakers were skeptical that DSGE models could be used for quantitative data analysis, especially short- and medium-term forecasting and the projection of macroeconomic aggregates under alternative interest rate scenarios.

At the same time most academic macroeconomists were still reluctant to use the kind of econometric techniques that enable careful documentation of the time series fit of a dynamic model. Many DSGE models impose very strong restrictions on actual time series and are rejected against less restrictive specifications such as VARs. Even though it has long been known that DSGE models can be estimated, their apparent misspecification was used as an argument in favor of informal calibration approaches, along the lines of Kydland and Prescott (1982).

In recent years econometricians have developed frameworks that formalize certain aspects of the calibration approach by taking the possibility of model misspecification explicitly into account without abandoning the tradition of probabilistic modeling

In the 1990s many policymakers were skeptical that DSGE models could be used for quantitative data analysis, especially short- and medium-term forecasting and projecting macroeconomic aggregates under alternative interest rate scenarios.

initiated by Haavelmo (1944). In particular, several authors, including DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), Otrok (2001), and Fernández-Villaverde and Rubio-Ramírez (2004), documented that Bayesian methods can be used in an insightful manner to estimate and evaluate DSGE models.

Smets and Wouters (2003) applied the newly developed Bayesian methods to a

DSGE model with enough nominal and real frictions that their specification had a good chance of fitting major aggregate time series in a traditional macroeconomic sense. In fact, the model development process was a productive synthesis of academic-style DSGE modeling and econometric model building. Technology and monetary policy shocks—the most common driving processes in theoretical models—were augmented by a list of shocks that to a large extent were chosen to pick up serial correlation of the wedges in intra- and intertemporal equilibrium conditions, the DSGE-model equivalent of regression residuals.

It is no surprise that central banks have paid a great deal of attention to Smets and Wouters's results and now devote significant resources to developing and estimating their own DSGE models. The best of two worlds appears within reach: a model that is well founded from a theoretical perspective and at the same time in tune with the empirical evidence so that it can deliver reliable forecasts and a coherent interpretation of past and current economic events.

Now that policy institutions are beginning to take the quantitative implications of DSGE models seriously, there is a need for robust evaluation procedures. The head of any central bank's research department that has just built a DSGE model for policy analysis and forecasting would want to know: How good is this model? Is it reliable enough so that it can be used for policy advice? Does the model need to be improved, for instance, by explicitly modeling labor market frictions, credit market imperfections, or information asymmetries? In principle, time will tell. As the model is used on a regular basis, analysts will discover *ex post* its strengths and weaknesses and point to directions for improvement. However, this real-time learning process is potentially slow and costly. Hence, it is important to subject the model to evaluation procedures that can signal deficiencies *ex ante*—that is, on the basis of the information currently available.

In an attempt to make the fit and forecasting performance of DSGE models comparable to a VAR, the structural models have been augmented by features that appear *ad hoc* and lack micro foundations. For instance, price stickiness is often introduced by assuming that only a fraction of firms are able to reoptimize their nominal prices. By itself this mechanism might not generate enough persistence to

explain the high autocorrelation of inflation rates in the data. Hence, researchers often add the assumption that those firms that do not reoptimize their prices can costlessly adjust their old prices by last period's inflation rate. Alternatively, the model could be augmented by serially correlated price markup shocks, which might reflect either time variation in the substitution elasticities between differentiated goods or the market structure of an industry. However, there is a trade-off for incorporating ad hoc propagation mechanisms or exogenous shocks into the model. On the one hand, model fit with respect to historical data is typically improved. On the other hand, it is questionable whether the ad hoc modifications are invariant to policy experiments. Although in the absence of large historical variation in monetary policy the invariance property is to some extent difficult to assess, the trade-off should serve as a word of caution and steer modelers toward parsimony and internal propagation mechanisms that are supported by microeconomic evidence.

In sum, there is a need for DSGE model evaluation procedures. This article reviews an evaluation procedure recently proposed in Del Negro and Schorfheide (2004) and Del Negro et al. (2004). The article first describes the DSGE model and the data used in the empirical application. The article next shows how the linear DSGE model can be nested in a VAR and reviews a procedure that is able to systematically relax the cross-coefficient restrictions imposed on the VAR by the DSGE model. The resulting DSGE-VAR specification is used as a tool to evaluate a version of the Smets and Wouters (2003) model. The analysis considers to what extent the DSGE model restrictions must be relaxed in order to optimize the fit of the DSGE-VAR and then uses the framework for comparisons of different DSGE model specifications. The article then describes some in- and out-of-sample results obtained with this procedure.²

The DSGE Model

The DSGE model used in the forecasting exercise is described in detail in Del Negro et al. (2004). The model is a slightly modified version of the DSGE model in Smets and Wouters (2003), which is in turn based on work of Christiano, Eichenbaum, and Evans (2005). Here we provide a brief and nontechnical overview of the model.

The model contains several nominal and real frictions. Nominal price and wage stickiness is modeled as in Calvo (1983). Firms (households) are monopolistic suppliers of a differentiated good (labor). In any period there is a chance that any given firm (household) may not be able to reset prices (wages). The prices (wages) of these firms (households) grow proportionally to the previous period's inflation. (This proportional growth is referred to as indexation in the remainder of the article.)

On the real side, the model features endogenous capital accumulation, adjustment costs to investment, and variable capital utilization. Households' preferences display habit persistence in consumption, and the utility function is separable in terms of consumption, leisure, and real money holdings. Fiscal policy amounts simply to balancing the budget in all periods. Monetary policy follows an interest feedback rule, in which the target federal funds rate depends on the rate of inflation and on the discrepancy between actual and trend output and adjustment to the target is gradual.

-
1. The empirical findings need to be qualified because Smets and Wouters worked with detrended data and thus did not use the most favorable prior for the Bayesian VAR and because VARs are not universally favored as a forecasting benchmark.
 2. The results shown in this article are variations or extensions of those discussed in Del Negro et al. (2004).

As in Smets and Wouters (2003), the model economy is subject to a large number of shocks: technology, discount rate, leisure preference, price markup, investment efficiency, monetary policy, and government spending. Technology shocks are assumed to be permanent and common to all firms. Discount rate and leisure preference shocks

In recent years econometricians have developed frameworks that formalize certain aspects of the calibration approach by taking the possibility of model misspecification explicitly into account without abandoning the tradition of probabilistic modeling.

shift households' utility; the first affects the household's willingness to substitute over time, and the latter the household's willingness to supply labor. So-called price markup shocks change the degree of substitutability among differentiated goods and in turn affect markups and the rate of inflation. Investment efficiency shocks alter the rate of transformation between

consumption and investment goods and serve as proxies for changes in the relative price of investment goods. Finally, both monetary policy and government spending shocks have a standard interpretation. All shocks are assumed to follow an autoregressive process of order one (in the case of technology, this assumption applies to the growth rate of technology) with the exception of monetary policy shocks, which are independently distributed over time.

The Data and the VAR Setup

The empirical analysis is based on quarterly U.S. observations that include both real and nominal series. The real variables are per capita real output, investment, consumption, hours per capita, and wages. The nominal variables are inflation and the interest rate. All data are obtained from Haver Analytics; Haver's abbreviations are in italics. Consistent with much of the real business cycle literature, this analysis treats consumption of durable goods (*CD*) as investment rather than consumption. Therefore, investment is defined as gross private domestic investment plus consumption of durables. Per capita real output, investment, and consumption are obtained by dividing the nominal series (*GDP*, *C - CD*, and *I + CD*, respectively) by the population sixteen years and older (*LN16N*) and deflating using the chained-price GDP deflator (*JGDP*). The real wage is computed by dividing compensation of employees (*YCOMP*) by total hours worked and the GDP deflator. Note that compensation per hour, which includes wages as well as employer contributions, accounts for both wage and salary workers and proprietors. The measure of hours worked is computed by taking total hours worked reported in the national income and product accounts (NIPA) (annual frequency) and interpolating it using growth rates computed from hours of all persons in the nonfarm business sector (*LXNFH*). Hours worked are divided by population to convert them into per capita terms. The analysis therefore uses a broad measure of hours worked that is consistent with its definition of both wages and output in the economy. Inflation rates are defined as log differences of the GDP deflator (*JGDP*) and converted into annualized percentages. The nominal rate corresponds to the effective federal funds rate (*FFED*), averaged within each quarter, also in percent.

As mentioned in the introduction, the DSGE model is nested in a more flexible vector autoregressive specification. The DSGE model features a stochastic trend, driven by the permanent technology shock. Real per capita output, consumption, investment, and the real wage are nonstationary and grow at the same rate in the long run. These nonstationary variables enter the VAR in growth rates, while the variables that are stationary according to the DSGE model—namely, per capita hours, inflation, and the nominal interest rate—enter the VAR in levels. All growth rates are computed

using quarter-to-quarter log differences and then multiplied by 100 to convert them into percentages. To take into account the fact that the nonstationary variables all move together in the long run according to the DSGE model, error-correction terms are also introduced into the VAR so that effectively we are estimating a vector error correction model (VECM). Importantly, to maintain the consistency between the VAR and the DSGE model, the coefficients in the cointegrating relationships are constrained to be those implied by the DSGE model—that is, real per capita output, consumption, investment, and the real wage are assumed to grow at the same rate in the long run in the VAR as well. The error correction terms are therefore given by consumption minus output, investment minus output, and the real wage minus output, respectively, all in logarithms.

The analysis uses observations from 1954Q4 to 2004Q1. The first four observations are used to initialize the lags of the VAR. The recursive estimation results and the pseudo-out-of-sample forecasts are based on a rolling sample of 120 observations (thirty years) starting in 1956Q1. Specifically, the rolling sample works as follows. We estimate the model on the sample 1956Q1–1985Q4, produce forecasts, and then shift the sample one quarter ahead and repeat the exercise. Therefore we have arguably a sample large enough to estimate the model as well as enough forecasts (fifty-eight) to assess the accuracy of out-of-sample predictions.

DSGE-VAR: A Brief Description of the Procedure

The short and informal description of the DSGE-VAR procedure in this section is intended to motivate the DSGE-VAR specification and to provide some intuition on how it can be used to estimate and evaluate DSGE models.³ It has long been recognized (for example, Sims 1980) that a tight relationship exists between dynamic equilibrium models and VARs. Imagine the following thought experiment, where for the moment the vector of DSGE model parameters is fixed. We generate 1 million observations from the DSGE model—that is, we generate a sequence of shocks (monetary policy, technology, etc.), feed them through the DSGE model, and obtain artificial data. Next, we estimate a VAR with p lags on these artificial data. If the DSGE model is covariance stationary, then the estimated VAR provides an approximation to the DSGE model with the property that its first p autocovariances are equivalent to the first p autocovariances of the DSGE model. By including more and more lags we can in principle match more and more autocovariances and increase the accuracy of the VAR approximation of the DSGE model. Now imagine that the data generation is repeated using different parameter values for the DSGE model. As long as the DSGE model parameter space is small compared to the VAR parameter space, a restriction function can be traced that maps the DSGE parameters into a VAR parameter subspace. Hence, estimating a DSGE model is (almost) like estimating a VAR with cross-equation restrictions.

Instead of dogmatically imposing the cross-coefficient restrictions implied by the DSGE model on the VAR, we will allow for deviations. The overall magnitude of these deviations is controlled by a hyperparameter, λ . Roughly speaking, if $\lambda = \infty$, then the restrictions are strictly enforced, whereas if $\lambda = 0$, the restrictions are completely

3. The section—and the whole article for that matter—purposely does not contain a formal treatment of the procedure. The latter is provided in Del Negro and Schorfheide (2004) and Del Negro et al. (2004). The appendix in Del Negro and Schorfheide (2004) discusses computational details for readers who are interested in implementing the procedure. Gauss and Matlab versions of the codes are available at www.econ.upenn.edu/~schorf/research.htm.

ignored in the estimation of the VAR parameters. To implement this idea formally, we use a Bayesian approach. In general terms, Bayesian methods are a collection of inference procedures that combine initial information about parameters with sample information in a logically coherent manner by use of Bayes's theorem. Both prior and postdata information are represented by probability distributions. In this particular application, the prior consists of a continuous probability distribution for the VAR

The best of two worlds appears within reach: a model that is well founded theoretically and can deliver reliable forecasts and a coherent interpretation of past and current economic events.

coefficients that is centered at the DSGE model implied restrictions.

The hyperparameter λ scales the covariance matrix of the prior: If λ is large the variance is small, and most of the prior mass on the VAR coefficients concentrates near the DSGE model restrictions. Vice versa, if λ is small the prior on the VAR

coefficients is diffuse. The prior is combined with the likelihood function to form the posterior distribution, which summarizes the postdata information about the VAR parameters. The larger λ is, the more the posterior shifts toward the DSGE model restrictions and the less the restrictions are relaxed in the estimation. We refer to the resulting vector autoregressive specification as DSGE-VAR. In the application the DSGE model depends on unknown parameters as well. It turns out that these parameters can be jointly estimated together with the VAR parameters by, loosely speaking, projecting the VAR coefficient estimates back onto the DSGE model restrictions.

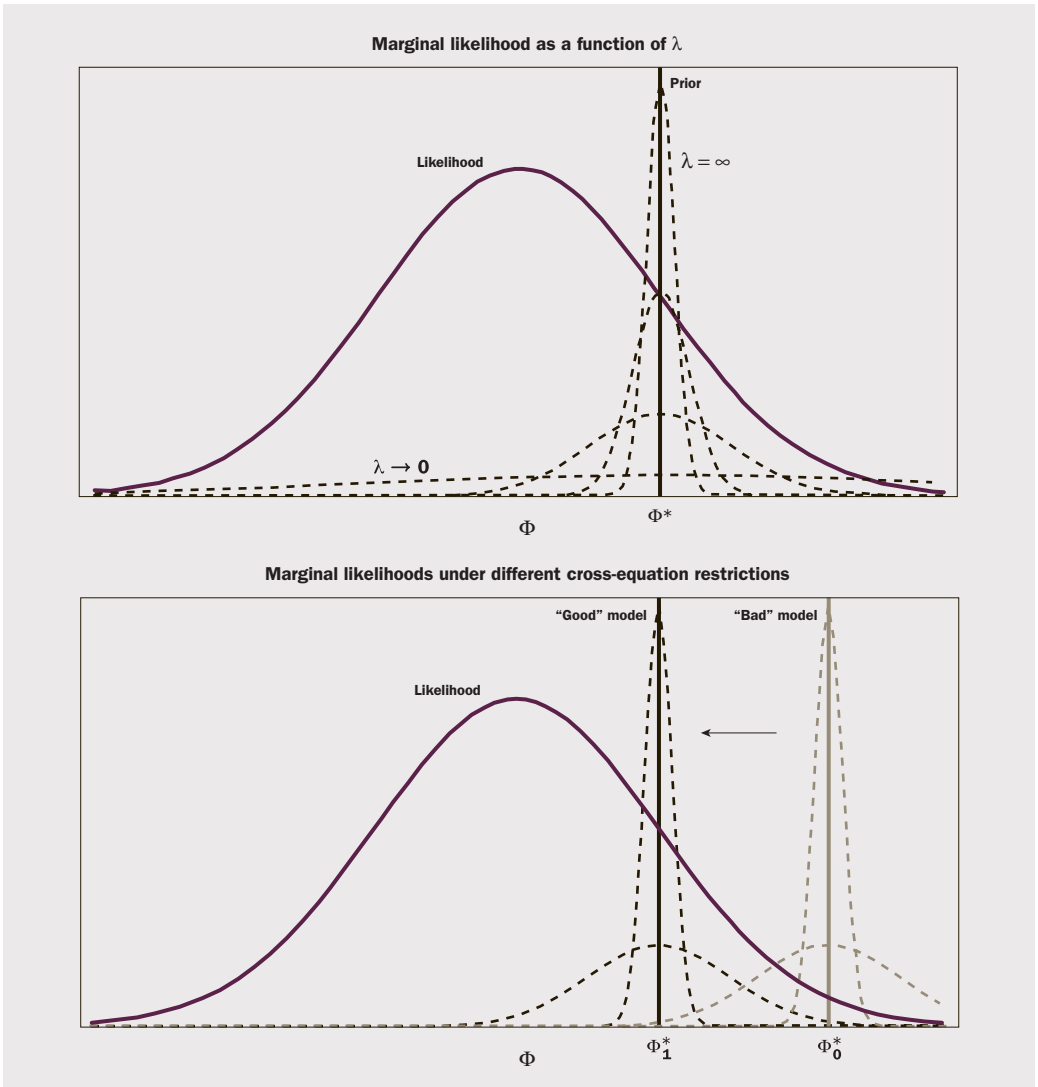
Both fit and forecasting performance suffer whenever the DSGE prior is either too tight or too loose. The fact that fit improves as the cross-equation restrictions are relaxed—that is, as λ decreases from infinity—indicates that these restrictions are at odds with the data in some dimensions. In the procedure proposed in Del Negro et al. (2004), an estimate of λ is used as a way to evaluate DSGE models. That is, the evaluation procedure hinges on the following question: How much must the cross-equation restrictions be relaxed to obtain the best-fitting model? The next section elaborates on why the answer to this question can shed light on some of the issues described in the introduction.

Why Does λ Tell Us How Good a DSGE Model Is?

In this section, a simple chart provides some intuition for the DSGE-VAR procedure. The first panel of Figure 1 plots the likelihood of the VAR as a function of the VAR parameter Φ . For the sake of exposition the multidimensional VAR parameter space is collapsed onto the real line. Assume that the DSGE model restrictions imply that the VAR parameter equals Φ^* . The remaining lines represent the DSGE prior for different values of λ . All these priors are centered at the cross-equation restrictions Φ^* . For $\lambda = \infty$ the prior puts all its mass on Φ^* . As λ decreases, the prior mass is spread out further away from the cross-equation restrictions. For λ approaching zero the prior becomes nearly flat. In a Bayesian setting a model consists of a likelihood function and a prior distribution. By varying the hyperparameter λ from infinity to zero we are essentially creating a continuum of models with the VAR approximation of the DSGE model at one end and an unrestricted VAR at the other end.

We adopt a measure of model fit that has two dimensions: goodness of in-sample fit on the one hand and a penalty for model complexity or degrees of freedom on the other hand. In a Bayesian framework such a measure is provided by the so-called marginal data density, which arises naturally in the computation of posterior model odds. The marginal data density is simply the integral of the likelihood taken according to the prior

Figure 1
Marginal Likelihoods and DSGE Priors



distribution—that is, the weighted average of likelihood where the weights are given by the prior. We then ask the following question: How does this measure of fit change as λ decreases from infinity to zero? We refer to the mapping from λ to the marginal data density as the posterior distribution of λ . Indeed, if we view λ as a hyperparameter and put a flat prior on it, this mapping characterizes a posterior distribution of λ .

Suppose one writes an oversimplified DSGE model, whose cross-equation restrictions are grossly at odds with the data. The first panel of Figure 1 clearly shows that if Φ^* is far in the tails of the likelihood, any prior that is very tight around Φ^* will have low marginal likelihood. As λ is decreased, the weight on parameters in the calculation of the data density that are associated with a high likelihood increases. Hence, small values of λ have large posterior weights. Notice, however, as λ approaches zero, the computation of the data density involves more parameter values for which

the likelihood function is essentially zero. Hence, one expects the posterior density of λ to fall eventually.

Now imagine improving the model by adding a number of frictions that generate more realistic cross-equation restrictions. One expects that the posterior distribution of λ will concentrate more mass on large values of the hyperparameter λ . The reasoning is as follows. Having better cross-equation restrictions means that Φ^* moves closer to the likelihood peak, as shown in the second panel of Figure 1. As a consequence, relatively tight priors will deliver a higher marginal likelihood than loose priors. As the posterior distribution of λ shifts to the right, its mode—the value $\hat{\lambda}$ that maximizes the marginal likelihood—will increase. The remainder of the article provides concrete examples of

how the posterior distribution of λ can shift as the underlying DSGE model changes.

What is the appeal of this procedure relative to the current practice in the literature? Following the work of Smets and Wouters (2003), a standard approach for evaluating the overall fit of a DSGE model

The fact that the DSGE model's fit improves as the cross-equation restrictions are relaxed indicates that these restrictions are at odds with the data in some dimensions.

is to compare its marginal data density (see definition above) with that of a Bayesian VAR (BVAR). Although most VARs used in practice are not equipped with a DSGE model prior—most researchers use either a version of the Minnesota prior (see Doan, Litterman, and Sims 1984) or a training sample prior—the problems arising in such a comparison can be discussed in the context of our framework. Current practice is to consider two extremes: On the one hand, $\lambda = \infty$ represents the DSGE model, and on the other hand, a small value $\lambda = \underline{\lambda}$ is a proxy for the BVAR that serves as a benchmark in the evaluation exercise. By using a very diffuse prior on one or more of the BVAR parameters—that is, choosing a low $\underline{\lambda}$ —one can make the marginal likelihood of the BVAR arbitrarily small. So one can always make the BVAR lose the horse race with the DSGE model by choosing, often unconsciously, a diffuse prior. At the same time, the VAR coefficient estimates simply converge to the maximum likelihood estimates as $\underline{\lambda}$ approaches zero.

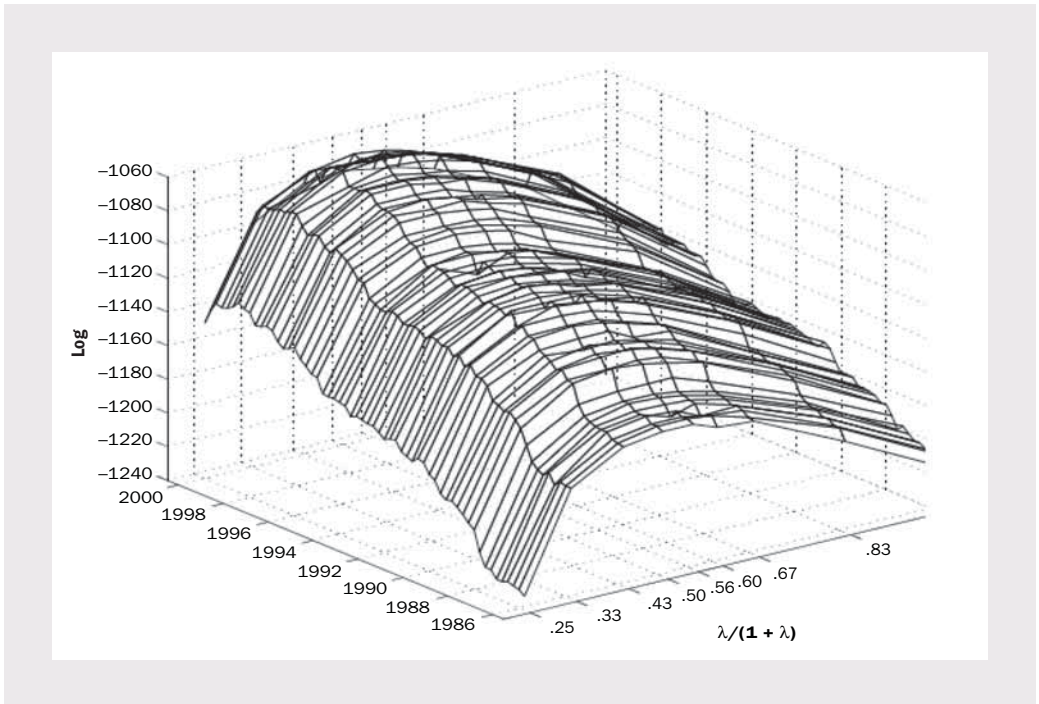
The sensitivity of posterior odds comparisons to seemingly innocuous changes in the prior distribution of the benchmark model implies that posterior odds of DSGE versus VAR models are not a robust way to address the question, How good is my DSGE model? Our DSGE-VAR framework imposes some rigor on the construction of the prior for the VAR; we emphasize that it is important to look at marginal data densities for an entire range of λ values instead of just two endpoints, one of which is typically chosen in a fairly arbitrary manner. As soon as we allow for intermediate values of λ , minor changes in model specifications are less likely to affect the answer to the question, Is there evidence of misspecification? Indeed, in Del Negro et al. (2004) and in the present article, we show that the overall shape of the posterior distribution of λ is quite a robust feature of the DSGE-VAR procedure.

A Look at the Data: The Posterior Distribution of λ over Time

The posterior distribution of λ is one of the main objects of interest in our empirical analysis: For any given sample it provides information on the degree of misspecification of the DSGE model. Since we estimate the DSGE-VAR in our empirical analysis not just once but essentially fifty-eight times based on the rolling samples, we can study how the posterior distribution of λ evolves over time.

Figure 2 shows the evolution of the posterior distribution of λ over time using a three-dimensional plot. For each of the fifty-eight rolling windows—the first ending

Figure 2
Marginal Likelihood as a Function of λ over Time



in 1985Q4 and the last one ending in 2000Q1—the marginal likelihood is computed for the following values of λ in the $[0.33, \infty]$ interval (where 0.33 is the smallest value of λ that generates a proper prior for the VAR parameters): 0.33, 0.5, 0.75, 1, 1.25, 1.5, 2, 5, ∞ . The x axis of Figure 2 shows the values of λ , which for expositional purposes are rescaled to be in the $[0, 1]$ interval—that is, the value of $\lambda/(1 + \lambda)$. The y axis shows the ending period of the rolling window, and the z axis shows the corresponding value of the logarithm of the marginal likelihood. Therefore, for any given rolling window ending between 1985Q4 and 2000Q1, the plot shows how the marginal likelihood of DSGE-VAR(λ) evolves as a function of λ .

The shape of the three-dimensional plot in Figure 2 is consistent with what we would expect. For any given window, the marginal likelihood initially increases with λ . Recall from Figure 1 that if the cross-equation restrictions are not too far in the tail of the likelihood—that is, if the DSGE model misspecification is not too large—tightening the DSGE prior leads to an improvement in the marginal likelihood. However, as the DSGE prior concentrates around the cross-equation restrictions, the marginal likelihood starts to decrease. We interpret this as evidence of misspecification because it suggests that relaxing the cross-equation restrictions improves the DSGE-VAR's fit.

The fact that the three-dimensional plot in Figure 2 looks like a tunnel indicates that the shape of the posterior distribution of λ is very robust over the sample period. Interestingly, this result is in contrast with other approaches to assessing the DSGE model's fit, as we will presently show. The tunnel is upward sloping in the time (y) dimension: For any given value of λ , the marginal likelihood tends to increase as the rolling window shifts forward. This phenomenon possibly reflects what has been dubbed the Great Moderation (Stock and Watson 2002): After the mid-eighties the volatility of

key macroeconomic variables has dropped sharply. Consequently, the predictability of these variables has increased, leading to an increase in the marginal likelihood. The ranking of fit as a function of λ is fairly stable over time with values of λ in the neighborhood of 1 (that is, $\lambda/(1+\lambda)$ around 0.5) always outperforming the two extremes of the interval—namely, DSGE-VARs with either a very loose or a very tight prior.

The relative ranking of the extremes of the λ interval—namely, the “loose prior” ($\lambda = 0.33$) versus the “degenerate prior” ($\lambda = \infty$) model—is not very robust, however. Comparing the fit of the DSGE model with that of a VAR with a loose prior—an

As the features of the DSGE model change, so do the cross-equation restrictions that the model imposes on its VAR representation.

approach that is often used in the literature—leads to conclusions that change dramatically over the sample period, even though the overall shape of the posterior distribution of λ is roughly the same. Since this pattern is difficult to assess from

Figure 2, the two charts in Figure 3 show slices of the tunnel, one taken at the beginning (1985Q4) and one at the end (2000Q1) of the rolling sample. The two charts in Figure 3 also contain a comparison across different models, which is the subject of the next section. For now, we focus on the Baseline model (the heavier black line), which plots the same numbers that are in Figure 2. The figure shows that at the beginning of the rolling sample, DSGE-VAR(∞) outperforms the VAR with a loose prior; the ranking is reversed at the end of the sample. The log difference between the marginal likelihood of DSGE-VAR(∞) versus DSGE-VAR(0.33) is 19 at the beginning of the rolling sample and -4 at the end. Taken literally, these differences imply posterior odds that are in one case decisively in favor of, and in the other case against, the DSGE model. Once again, the overall shape of the posterior distribution is roughly the same in both charts. In fact, in both cases the two extremes are in tails of the posterior distribution of λ , their posterior odds relative to the best-fitting DSGE-VAR being negligible.

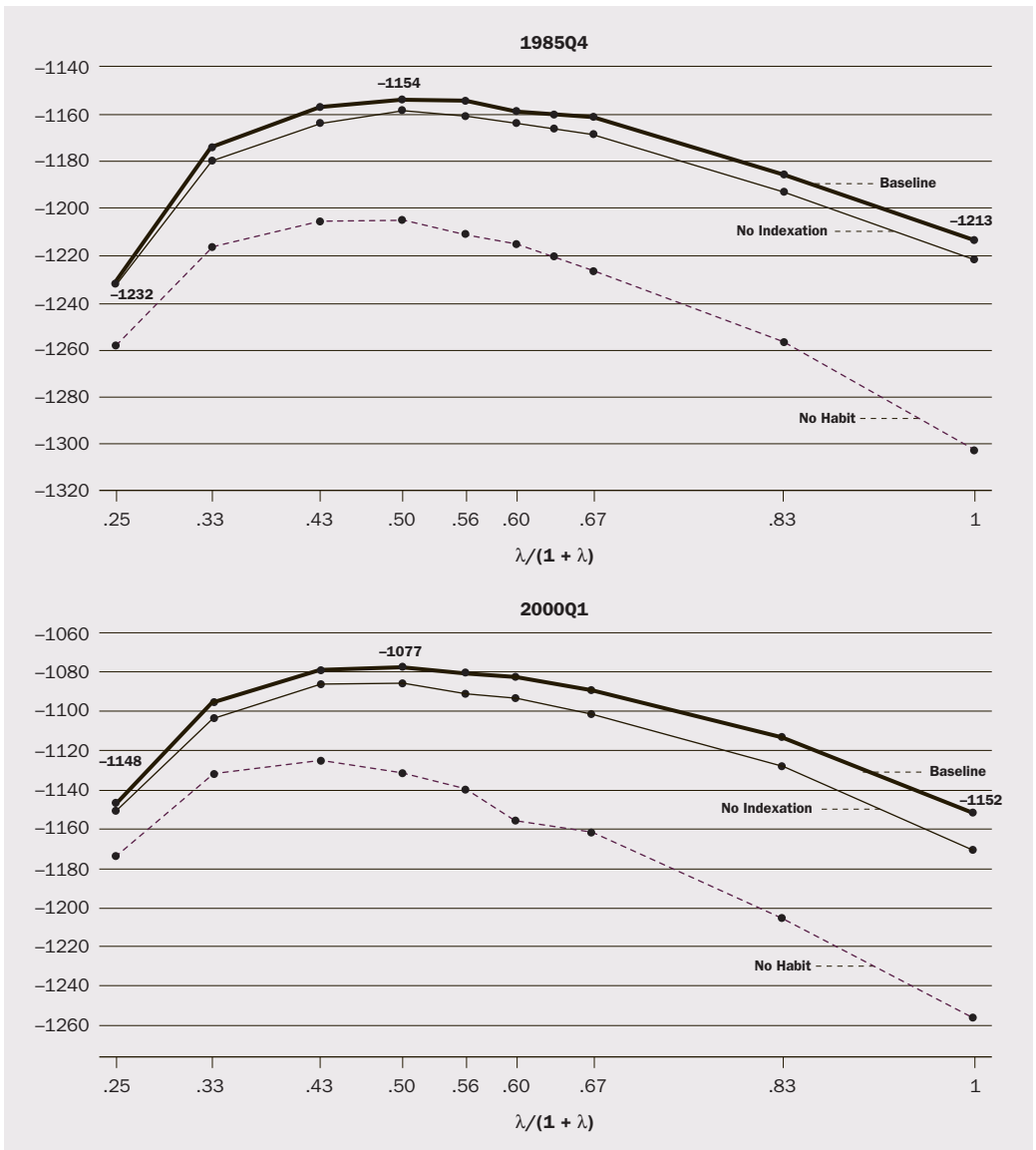
This last observation implies that the VAR with a loose prior may not be the right reference model to use for impulse response comparison because its fit is sometimes worse relative to that of the model being analyzed (the DSGE model). Certainly, both Figures 2 and 3 show that the fit of the VAR with a loose prior is always much worse than that of the best-fitting model DSGE-VAR($\hat{\lambda}$). This result suggests that the latter provides a more reliable benchmark.

Model Comparisons

In this section we use the DSGE-VAR procedure to compare across DSGE models. As the features of the DSGE model change, so do the cross-equation restrictions that the model imposes on its VAR representation. For the reasons discussed in the article’s introduction, we are interested in determining which model features are truly important and which are not.

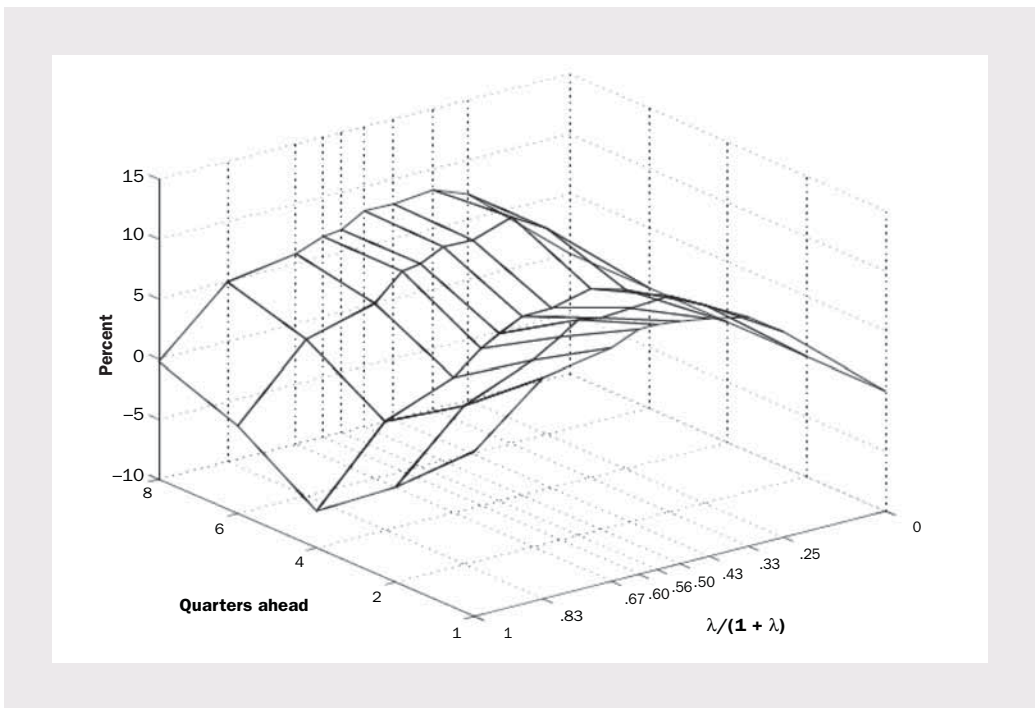
In Del Negro et al. (2004) we consider two alternatives, also shown in Figure 3, to the so-called Baseline model discussed so far. The No Indexation specification (the thinner black line) eliminates price and wage indexation to last period’s inflation. Price and wage indexation is often viewed as being somewhat at odds with microeconomic evidence and therefore considered not truly structural. The other specification, called No Habit (the dashed purple line), eliminates habit persistence in preferences. Some of the literature has argued that these features are needed in order to fit the data. Here we use the DSGE-VAR procedure to assess whether this is the case. We also want to learn whether the conclusions from the procedure are robust across samples.

Figure 3
Marginal Likelihood as a Function of λ for Different DSGE Models



We are interested in studying how the shape of the posterior distribution of λ changes across models. We argued previously that the posterior distribution of λ shifts to the left if the misspecification of the cross-equation restrictions (which we referred to as Φ') increases. Figure 3 shows that this shift indeed occurs for the No Habit model, regardless of the sample. Relative to the Baseline model, the marginal likelihood is much lower for any value of λ . Most importantly, the posterior mass clearly shifts to the left, toward lower values of λ (looser prior). This is not so much the case, however, for the No Indexation model. The marginal likelihood is slightly lower for the No Indexation relative to the Baseline model for any λ , indicating that the fit worsens, but the posterior distribution does not show any appreciable shift to the left.

Figure 4
Forecasting Accuracy as a Function of λ : Baseline Model



We interpret these findings as strong evidence that the habit persistence in preferences substantially improves the fit of the DSGE model. Therefore, those who believe that habit persistence is not a structural feature may have to introduce an alternative mechanism that delivers similar effects: Simply eliminating habit persistence comes at a significant cost in terms of fit. On the contrary, the evidence in favor of price and wage indexation is not nearly as strong.

Forecasting Results

This section presents some forecasting results obtained with the models discussed above. In particular, we want to find out to what extent the in-sample results shown so far carry over to the pseudo-out-of-sample comparison. Figure 4 shows a multivariate forecasting statistic for DSGE-VAR(λ) relative to the unrestricted VAR for forecasting horizons one through eight quarters ahead. The multivariate forecasting statistic is a summary measure of forecasting accuracy. Loosely speaking, this multivariate measure can be seen as a weighted average of the root mean square error for the individual variables. Differently from a simple weighted average, however, this measure also takes into account the correlation in the forecast errors. The z axis in Figure 4 reports the percentage gain in multivariate forecasting accuracy relative to the unrestricted VAR. As in Figure 2, we rescale λ to be in the [0, 1] interval. Thus, the x axis shows the value of $\lambda/(1 + \lambda)$, and the y axis shows the forecast horizon.

Just as in Figure 2, the three-dimensional plot in Figure 4 is also tunnel-shaped. In other words, forecasting accuracy is an inverted U-shaped function of λ for all forecast horizons. Consistently with the in-sample results, forecasting performance is maximized whenever the DSGE prior is neither too loose nor too tight. The magni-

Table
Out-of-Sample Root Mean Square Errors: Percentage Improvement Relative to VAR

		Forecast horizon (quarters)				
		1	2	4	6	8
Y	DSGE-VAR($\hat{\lambda}$)	16.3	14.1	12.5	13.5	13.6
	DSGE-VAR(∞)	0.9	-17.6	-56.5	-82.5	-102.9
	VAR, RMSE	0.67	0.97	1.68	2.38	2.98
C	DSGE-VAR($\hat{\lambda}$)	-6.8	-7.6	7.1	16.6	21.5
	DSGE-VAR(∞)	-15.7	-21.4	-0.8	11.3	12.0
	VAR, RMSE	0.42	0.62	1.06	1.56	2.03
I	DSGE-VAR($\hat{\lambda}$)	17.8	8.0	-5.0	-11.5	-17.2
	DSGE-VAR(∞)	-4.2	-41.2	-101.0	-135.3	-157.8
	VAR, RMSE	2.67	3.98	6.59	9.14	11.45
H	DSGE-VAR($\hat{\lambda}$)	10.0	10.9	-0.6	-0.0	0.7
	DSGE-VAR(∞)	-13.6	-37.9	-95.4	-116.5	-127.2
	VAR, RMSE	0.58	0.92	1.56	2.26	2.88
W	DSGE-VAR($\hat{\lambda}$)	8.2	11.7	11.1	14.9	18.4
	DSGE-VAR(∞)	6.7	12.7	18.1	27.0	36.6
	VAR, RMSE	0.65	1.06	1.72	2.28	2.82
Inflation	DSGE-VAR($\hat{\lambda}$)	10.7	10.9	22.9	31.0	36.6
	DSGE-VAR(∞)	8.4	4.2	10.4	21.1	29.6
	VAR, RMSE	0.25	0.47	0.98	1.68	2.42
R	DSGE-VAR($\hat{\lambda}$)	27.3	23.4	9.2	7.0	9.1
	DSGE-VAR(∞)	27.7	17.8	3.2	8.2	17.1
	VAR, RMSE	0.68	1.14	1.63	2.11	2.64
Multivariate statistic	DSGE-VAR($\hat{\lambda}$)	11.0	8.8	6.1	9.4	9.4
	DSGE-VAR(∞)	3.8	-2.1	-6.9	-2.7	-0.2
	VAR, RMSE	0.68	0.23	-0.18	-0.47	-0.65

tude of the relative improvement in forecasting accuracy varies over the forecasting horizon. But for any given forecasting horizon, the values of λ that maximize forecasting performance are those in the neighborhood of 1 (that is, $\lambda/(1 + \lambda) = 0.5$), and roughly correspond to those that maximize in-sample fit.

The table takes a closer look at the forecasting performance of the unrestricted VAR, DSGE-VAR($\hat{\lambda}$), and DSGE-VAR(∞). For each of the seven variables and for the multivariate statistic, the table shows the root mean square error (RMSE) of the unrestricted VAR as well as the percentage improvement in forecasting accuracy (whenever positive) of DSGE-VAR($\hat{\lambda}$) and DSGE-VAR(∞) relative to the VAR. For DSGE-VAR($\hat{\lambda}$) the value of $\hat{\lambda}$ is chosen ex ante for each rolling sample on the basis of the marginal likelihoods shown in Figure 2. The precise value changes from sample to sample but is always in the neighborhood of 1, as one can see from Figure 2. For those variables that enter the estimation in growth rates (output, consumption, investment, and the real wage), as well as for inflation, we focus on cumulative forecasts. Therefore, for forecast horizons beyond one quarter, the forecast errors measure the cumulative error in forecasting inflation over, say, the next two years, as opposed to the error in forecasting the variable two years ahead. For instance, an eight-quarter-ahead error of 2 percent in

forecasting consumption implies that the model makes a mistake of 50 basis points (annualized) in forecasting average consumption growth in the next two years.

The table shows that for most variables and forecasting horizons the DSGE-VAR($\hat{\lambda}$) improves over the unrestricted VAR. This is certainly the case for the multivariate statistic, as already mentioned in the discussion of Figure 4. Short-run consumption forecasts and long-run investment forecasts are an exception. Interestingly, there seems to be a trade-off between forecasting consumption and investment. This trade-off reflects the fact that all three models considered in the table are error-correction models with the same long-run cointegrating restrictions on output, consumption, investment, and the real wage. Since these cointegrating restrictions are at odds with the data, accurate forecasts for some of these variables result in inaccurate forecasts for others given that not all series grow proportionally in the long run as the model predicts. Another manifestation of this phenomenon is the fact that DSGE-VAR(∞) outperforms the other two models in forecasting the real wage but performs very poorly in forecasting both output and investment, especially in the long run. In summary, the fact that the DSGE model imposes these long-run cointegrating restrictions results in a serious limitation of its forecasting ability. To the extent that DSGE-VAR inherits the same long-run restrictions, its accuracy suffers as well.

For the remaining variables, DSGE-VAR($\hat{\lambda}$) is roughly as accurate as the unrestricted VAR in terms of hours per capita, while DSGE-VAR(∞) is far worse, especially in the long run. Conversely, DSGE-VAR(∞) performs well in terms of the nominal variables, inflation and the interest rate. For inflation its forecasting accuracy is slightly inferior to that of DSGE-VAR($\hat{\lambda}$) and far superior to that of the unrestricted VAR. For the nominal interest rate, DSGE-VAR(∞) outperforms DSGE-VAR($\hat{\lambda}$) for longer forecast horizons, but in the short run the two models have roughly the same forecasting performance.

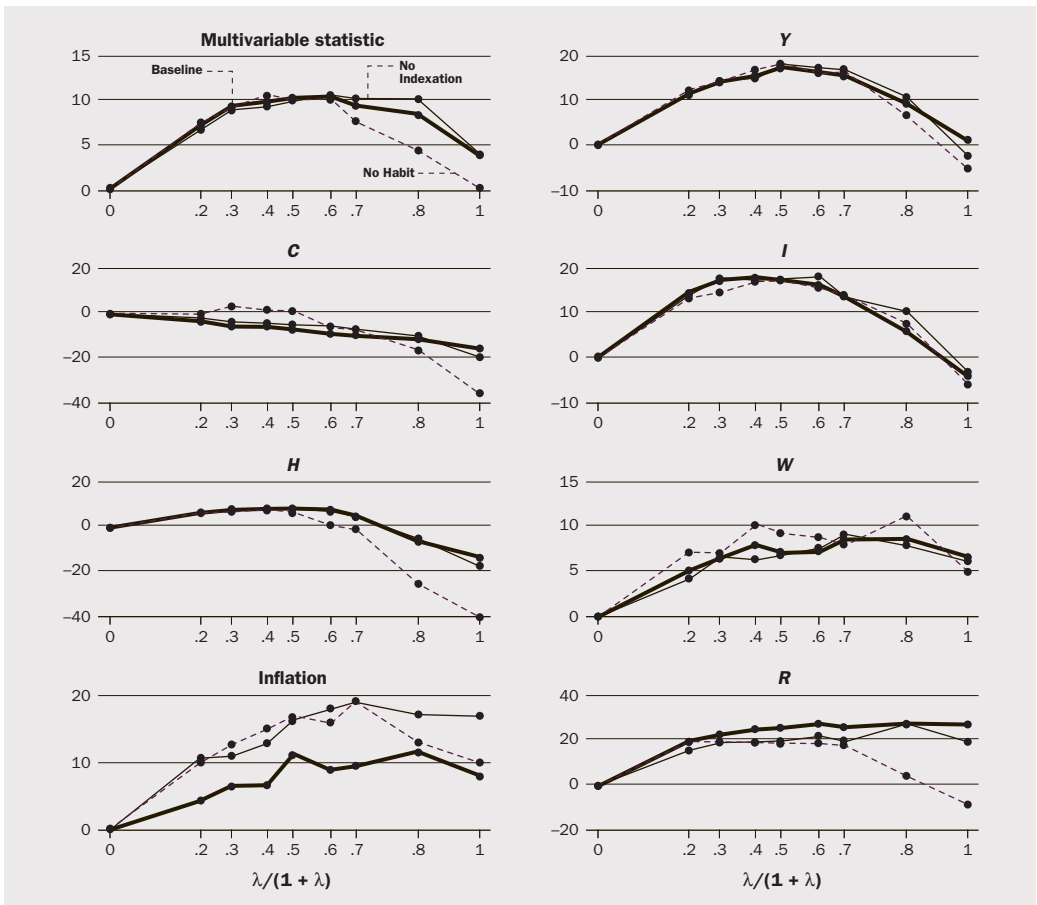
We conclude the section with a comparison of the out-of-sample forecasting performance across models. For each of the three models discussed so far (Baseline, No Indexation, No Habit), Figure 5 shows the one-quarter-ahead percentage improvement in RMSEs relative to the unrestricted VAR for all seven variables, as well as the improvement in the multivariate forecast statistic, as a function of λ . The focus on one-period-ahead forecasting accuracy facilitates the comparison with the results in Figure 3, which were based on the marginal likelihood.

The results in Figure 5 agree in a number of dimensions with those in Figure 3. The multivariate statistic plot, for instance, indicates that forecasting accuracy worsens considerably for the No Habit model as the DSGE prior becomes too tight. The plots for the individual variables show that for high values of λ the No Habit model performs worse than the other two models not only for consumption, as expected, but also for hours and the nominal interest rate. For other variables, however—notably, investment, real wages, and inflation—the No Habit model performs as well as the other two models.

Consistent with the overall message from the model comparison based on the marginal data densities, the No Indexation and Baseline models perform roughly as well in terms of the multivariate statistic. The forecasting performance of the two models is pretty much the same for most individual variables. One interesting exception is inflation, where the No Indexation model clearly forecasts better than the Baseline model. In summary, the out-of-sample exercise confirms the finding, consistent with our earlier discussion, that there is no strong evidence in favor of including wage and price indexation in the DSGE model.

In some other dimensions, however, the results in Figure 5 cast some doubt on the model comparison based on the marginal likelihood. For instance, the shape of the marginal likelihood curves as a function of λ for the No Indexation and Baseline models

Figure 5
Forecasting Accuracy as a Function of λ for Different DSGE Models



were very similar. Yet the marginal likelihood comparison, if taken literally, suggests that the No Indexation model should be at a loss relative to the Baseline model in terms of fit. This pattern does not emerge from the out-of-sample model comparison, however. Likewise, for low values of λ the difference in marginal likelihoods between the No Habit and the other two models is narrower than for large values of λ but still quite large. In the out-of-sample comparison, however, the three models seem to perform equally well in terms of the multivariate statistic for low values of λ .

DSGE-VAR as a Reference Model

In the previous sections, we argued that the posterior distribution of λ provides a robust measure of overall fit for the DSGE model. However, we often want to know more than just whether a DSGE model's fit is good or not. If the model fails—that is, if the λ posterior does not peak around a large value—we want to know in which dimensions it has to be improved. Although the forecast results provide us with information about the accuracy with which individual variables can be predicted by the model, they do not document how well the structural model captures comovements.

Comovements and the propagation of structural shocks can be illustrated with impulse response functions. While it is straightforward to compute impulse responses

from an estimated DSGE model, finding an appropriate benchmark to which these responses can be compared is more difficult. Many authors, including Nason and Cogley (1994), Rotemberg and Woodford (1997), Schorfheide (2000), and Christiano, Eichenbaum, and Evans (2005), have compared impulse responses from a DSGE model to responses obtained from a VAR. Such a comparison faces two challenges. First, for the VAR to be a meaningful benchmark, it has to fit the data better, in an econometric sense, than the DSGE model. Second, the VAR has to be expressed in terms of structural shocks—that is, technology shocks, monetary policy shocks, and so forth—rather than reduced-form one-step-ahead forecast errors. The identification of structural shocks in the context of a VAR requires auxiliary assumptions. Ideally, these auxiliary assumptions should satisfy

We find that the DSGE-VAR procedure delivers reasonably robust answers to the question, How good is my DSGE model?

the following coherency requirement: Supposing the identified VAR is fitted to artificial data from the DSGE model, then the VAR estimates of the structural shocks should coincide with the shocks that were fed into the DSGE model to generate the data.

The marginal data density analysis as well as the pseudo-out-of-sample forecasting exercise suggests that an unrestricted VAR does not master the first challenge: It fits and forecasts worse than the DSGE model and hence does not provide a credible benchmark.

On the other hand, the DSGE-VAR($\hat{\lambda}$) passes the first hurdle easily. Our procedure of selecting the hyperparameter ensures that we are using a benchmark specification that fits better than the DSGE model.

The second challenge lies in the derivation of a model-consistent VAR identification scheme. For instance, Altig et al. (2004) consider responses to two shocks: a permanent technology shock and a monetary policy shock. The authors identify technology shocks by assuming that these are the only shocks that can have a permanent effect on the long-run level of the real variables (output, consumption, etc.), as in the DSGE model. Monetary policy shocks are identified using the assumption that firms and households can observe them only with a one-period lag. Hence prices, output, and other macroeconomic quantities do not react instantaneously to monetary policy shocks.

It is well known, however, that in short samples long-run restrictions lead to imprecise estimates of impulse responses (see, for instance, Faust and Leeper 1997), making them a possibly unreliable benchmark. Monetary policy impulse responses, identified with short-run restrictions, do not suffer from this drawback. However, for the identification scheme to be model consistent, one typically has to introduce fairly ad hoc decision lags into the structural model. Finally, monetary policy and technology shocks combined explain only a fraction of the variation observed in the data. Yet it is typically not straightforward to construct model-consistent identification schemes for remaining shocks using traditional zero restrictions.

It turns out that the DSGE-VAR framework is rich enough to deliver an elegant solution to this identification problem, as originally discussed in Del Negro and Schorfheide (2004). Recall that a VAR is identified when there is a unique mapping between the forecast errors and economically interpretable shocks. In the DSGE-VAR procedure the mapping is chosen such that the DSGE and the DSGE-VAR impulse responses would coincide if the data were generated by the DSGE model ($\lambda = \infty$). By construction, the identification in the DSGE-VAR is therefore consistent with that in the DSGE model. Whenever $\hat{\lambda}$ is less than infinity, the impulse responses of DSGE-VAR($\hat{\lambda}$) and the DSGE model will differ. From the comparison between the two, one

can potentially learn in which dimensions the DSGE model is failing and how it can be improved. Although we do not discuss the empirical findings in Del Negro et al. (2004) in detail here, we note that the impulse response comparison confirms the results discussed so far. For the No Habit model there is clear evidence that something is amiss: For instance, consumption responds abruptly to both monetary policy and technology shocks for the DSGE model, while the response according to the reference model is smoother. Again, such strong evidence is absent in the case of indexation.

Conclusion

The article discusses DSGE-VAR—a procedure that can be used to evaluate and compare DSGE models. Drawing on existing work by Del Negro et al. (2004), the article also provides examples of how the procedure works in practice. We find that the DSGE-VAR procedure, unlike some of the current practices in the literature, delivers reasonably robust answers to the question, How good is my DSGE model?

REFERENCES

- Altig, David, Lawrence J. Christiano, Martin Eichenbaum, and Jesper Linde. 2004. Firm-specific capital, nominal rigidities, and the business cycle. Federal Reserve Bank of Cleveland Working Paper 04-16, December.
- Calvo, Guillermo. 1983. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* 12, no. 3:383–98.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles Evans. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113, no. 1:1–45.
- DeJong, David N., Beth F. Ingram, and Charles H. Whiteman. 2000. A Bayesian approach to dynamic macroeconomics. *Journal of Econometrics* 98, no. 2:203–23.
- Del Negro, Marco, and Frank Schorfheide. 2004. Priors from general equilibrium models for VARs. *International Economic Review* 45, no. 2:643–73.
- Del Negro, Marco, Frank Schorfheide, Frank Smets, and Raf Wouters. 2004. On the fit and forecasting performance of New Keynesian models. Federal Reserve Bank of Atlanta Working Paper 2004-37, December.
- Doan, Thomas, Robert Litterman, and Christopher Sims. 1984. Forecasting and conditional projections using realistic prior distributions. *Econometric Reviews* 3:1–100.
- Faust, Jon, and Eric M. Leeper. 1997. When do long-run identifying restrictions give reliable results? *Journal of Business & Economic Statistics* 15:345–53.
- Fernández-Villaverde, Jesús, and Juan Francisco Rubio-Ramírez. 2004. Comparing dynamic equilibrium models to data: A Bayesian approach. *Journal of Econometrics* 123, no. 1:153–87.
- Haavelmo, Trygve. 1944. The probability approach in econometrics. *Econometrica* 12 (supplement, July): iii–vi+1–115.
- Kydland, Finn E., and Edward C. Prescott. 1982. Time to build and aggregate fluctuations. *Econometrica* 50, no. 6:1345–70.
- Nason, James M., and Timothy Cogley. 1994. Testing the implications of long-run neutrality for monetary business cycle models. *Journal of Applied Econometrics* 9 (supplement, December): S37–S70.
- Otrok, Christopher. 2001. On measuring the welfare cost of business cycles. *Journal of Monetary Economics* 47, no. 1:61–92.
- Rotemberg, Julio, and Michael Woodford. 1997. An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual* 12:297–46.
- Schorfheide, Frank. 2000. Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics* 15, no. 6:645–70.
- Sims, Christopher A. 1980. Macroeconomics and reality. *Econometrica* 48, no. 1:1–48.
- Smets, Frank, and Raf Wouters. 2003. An estimated stochastic dynamic general equilibrium model of the euro area. *Journal of the European Economic Association* 1, no. 5:1123–75.
- Stock, James H., and Mark W. Watson. 2002. Has the business cycle changed and why? NBER Working Paper No. 9127, August.

Instability in U.S. Inflation: 1967–2005

JAMES M. NASON

The author is an economist in the macropolicy section of the Atlanta Fed's research department. He thanks Buz Brock, Tom Cunningham, Juan F. Rubio-Ramírez, Ellis Tallman, Mark Watson, and Tao Zha for comments and suggestions; Andy Bauer, Michael Hammill, and Annie Tilden for excellent research assistance; and Elaine Clokey for patient assistance in preparing the draft.

Concerns about price stability and high, persistent, and volatile inflation are universal among central bankers. These concerns are institutionalized in the United States by the Federal Reserve Act in its statement of monetary policy objectives:

The Board of Governors of the Federal Reserve System and the Federal Open Market Committee shall maintain long run growth of the monetary and credit aggregates commensurate with the economy's long run potential to increase production, so as to promote effectively the goals of maximum employment, stable prices, and moderate long-term interest rates. (Federal Reserve Act, sec. 2A)

Although the price stability goal is wedged between mandates to promote employment and restrain long-term interest rates, maintaining a stable price level has come to dominate discussions among academic economists and many central bankers. Statements by Federal Reserve policymakers have been remarkably consistent about what constitutes the price stability objective over the past twenty years. For example, in 1994 Alan Greenspan told a congressional subcommittee, “We will be at price stability when households and businesses need not factor expectations of changes in the average level of prices into their decisions” (1994). This statement suggests that price stability occurs when the only source of inflation dynamics is unpredictable shocks whose size does not vary “too much” over time.

This article studies U.S. inflation, inflation growth, and price level dynamics. The analysis is disciplined with autoregressive (AR), moving average (MA), and unobserved components (UC) models. The models produce mean inflation; inflation and inflation growth persistence; and inflation, inflation growth, and price level volatility estimates for a sample that begins in January 1967 (1967M01) and ends with September 2005 (2005M09). Although this article is silent on the success of policies aimed at price stability, these estimates reveal whether the persistence

and volatility of inflation, inflation growth, and the price level have changed during the past forty years.

The disinflation of the 1980s suggests that inflation became less persistent and volatile in the 1990s and early 2000s compared to the inflation of the 1970s. For example, Stock and Watson (2005) report that quarterly U.S. inflation became less persistent and volatile after 1984. This finding suggests that it is possible to better forecast inflation. However, lower inflation volatility also makes it more difficult to

choose the best inflation forecast from among a set of competing models. Stock and Watson (1999, 2005) verify the impact of lower persistence and volatility on post-1984 inflation forecasts.¹

The estimates in this article reveal whether the persistence and volatility of inflation, inflation growth, and the price level have changed during the past forty years.

Compared to the aims of Stock and Watson (2005), the goal of this article is modest. The article presents evidence about

inflation, inflation growth, and price level dynamics that complements Stock and Watson's evidence. Estimates of AR, MA, and UC models are reported in this article on the 1967M01–2005M09 sample and on two samples that roll through the 1970s, 1980s, and 1990s. The two rolling samples, described in a later section, produce AR, MA, and UC model estimates that provide information about instability in mean inflation and the persistence and volatility of inflation and inflation growth.

Four price level measures—different versions of the monthly consumer price index (CPI) and monthly personal consumption expenditure deflator (PCED)—are studied in this article. The CPI and PCED deflators are defined as CPI-CORE and PCED-CORE, which exclude food and energy items, and CPI-ALL and PCED-ALL, which include the relevant universe of consumer goods. These four series provide information on price level, inflation, and inflation growth.

This article reports AR persistence and volatility estimates that are sensitive to the choice of sample. For example, the first rolling sample yields AR persistent estimates that exhibit little change after the 1973–75 recession for the four inflation rates. When the second rolling sample drops observations from the 1970s for CPI inflation and from the 1970s and 1980s for PCED inflation, drift in the AR coefficients suggests instabilities in inflation persistence.

The MA and UC model estimates appear consistent with instabilities in the persistence and volatility of CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation growth and levels. Especially striking are MA and UC model estimates on the second rolling sample that suggest PCED-CORE inflation is serially uncorrelated subsequent to the early 1990s recession, a result that is affirmed by the AR persistence estimates. Thus, a reasonable current forecast of PCED-CORE inflation might be its average of the past fifteen years. Whether such a forecast is consistent with the Greenspan (1994) notion of price stability is outside the scope of this article.

The Models

This section reviews the empirical models: a p th-order autoregression, $AR(p)$; a first-order moving average, $MA(1)$; and an unobserved components–local level (UC-LL) model. These models are employed to study inflation, inflation growth, and price level dynamics; the choice of these models is guided by the literature on inflation dynamics. For example, Stock and Watson (2005) report estimates of $AR(p)$, $MA(1)$, and UC models on quarterly inflation and inflation growth. This article employs similar models but engages monthly samples of CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE.

The AR(p) model yields estimates of average or mean inflation, inflation persistence, and inflation volatility. These estimates are generated by AR(p) models written in deviations from mean inflation,

$$\pi_t - \pi_0 = \sum_{j=1}^p \gamma_j (\pi_{t-j} - \pi_0) + \varepsilon_t,$$

where inflation, π_t , is defined as the difference between the (natural) log of the month t price level, P_t , and month $t - 1$ price level, $\pi_t \equiv 1200 \times (\ln P_t - \ln P_{t-1})$, π_0 is mean inflation, and ε_t is the inflation forecast innovation or shock. Maximum likelihood estimates (MLEs) of the AR(p) are generated from Kalman filter iterations. (See the appendix for details.)

Information about inflation persistence is contained in the γ_j s. One measure of inflation persistence is the sum of the γ s, defined by $\gamma(\mathbf{1}) \equiv \sum_{j=1}^p \gamma_j$. This sum represents the cumulative response of inflation to its own shock, ε_t . Another metric of inflation persistence is the largest AR root of the γ s, Λ .² The largest eigenvalue Λ of the γ s captures the speed at which inflation returns to its long-run average in response to ε_t .³ Since $\gamma(\mathbf{1})$ and Λ are functions of the AR coefficients, $\gamma_1 \dots \gamma_p$, these statistics reveal different aspects of inflation persistence. The length of time inflation takes to return halfway to its long-run mean is a function of the largest eigenvalue Λ , $\ln 0.5 / \ln \Lambda$. Inflation persistence rises as $\gamma(\mathbf{1})$ and Λ approach 1 (from below).

As inflation persistence rises, it takes on a unit root and becomes nonstationary. This condition arises when, for example, $\gamma(\mathbf{1}) \geq 1$. Stock and Watson (2005) report estimates of $\gamma(\mathbf{1})$ larger than 1 that point to a unit root in quarterly U.S. inflation since 1970. A lesson they draw is that it is better to study models of inflation growth, $\Delta\pi_t = \pi_t - \pi_{t-1}$, rather than the level of inflation. One such model is the MA(1) process,

$$\Delta\pi_t = \eta_t - \theta\eta_{t-1},$$

where θ is the MA1 coefficient of inflation growth and η_t is the MA(1) mean zero forecast innovation or shock, with homoskedastic standard deviation σ_η .⁴

Estimates of the MA(1) coefficient θ contain information about inflation growth persistence. The MA(1) yields the AR(∞), $\Delta\pi_t = \sum_{j=1}^{\infty} \vartheta_j \Delta\pi_{t-j} + \eta_t$, where $\vartheta_j = \theta^j$, given $|\theta| \in (-1, 1)$. The sum $\vartheta(\mathbf{1})$ equals $-\theta/1 - \theta$. Therefore, the long-run response of inflation growth to its shock η_t increases as $\theta \rightarrow 1$. At $\theta = 1$, the speed of adjustment of inflation to an own shock is instantaneous.

It is interesting to explore the impact of $\theta = 1$ on the MA(1) of inflation growth. In this case, $\Delta\pi_t = \Delta\eta_t$. Since the difference operator Δ appears on either side of the equality, the Δ operators cancel. The result is that inflation collapses to the white noise process, $\pi_t = \eta_t$.⁵ When $\theta = 1$, inflation is unforecastable because it is driven only by the unpredictable shock η_t .

-
1. Hansen, Lunde, and Nason (2005) provide similar evidence. They apply their metric for choosing the best forecasting models on pre- and post-1984 samples. The Hansen, Lunde, and Nason metric finds it more difficult to distinguish between competing inflation forecasting models in the post-1984 sample. Nonetheless, that study is able to identify several Phillips curve models that outperform a random walk model in out-of-sample inflation forecasting exercises across the two samples. This result stands in contrast to results in Atkeson and Ohanian (2001) and Fisher, Liu, and Zhou (2002).
 2. Computation of Λ is described in the appendix.
 3. Another way to describe Λ is that it is the speed of adjustment of inflation along its transition path.
 4. Nelson and Schwert (1977) and Pearce (1979) also find that an integrated MA(1) best fits U.S. inflation.
 5. This white noise process for inflation ignores π_t .

Another implication of $\theta = 1$ is that the price level is a random walk, $\ln P_t = \ln P_{t-1} + \eta_t$. A random walk forces persistence onto the price level because an increase in η_t never decays regardless of the length of the forecast horizon.⁶ For example, the forecast of $\ln P_{t+j}$, $j > 1$, is $\ln P_{t-1} + \eta_t$, according to the random walk. A random walk in the price level also sets its trend to the sum of the shocks, $\ln P_t = \sum_{j=0}^{\infty} \eta_{t-j}$.

The UC-LL model imposes random walks on $\ln P_t$ and π_t . Besides placing a random walk trend in the price level, the UC-LL model endows inflation with a random walk that measures deviations from the price level trend. A convenient way to write the UC-LL model is

$$\begin{aligned} \ln P_t &= \mu_{1,t}, \\ \mu_{1,t+1} &= \mu_{1,t} + \mu_{2,t} + \delta_{t+1}, \quad \delta_{t+1} \sim \mathbf{N}(0, \sigma_\delta^2), \\ \mu_{2,t+1} &= \mu_{2,t} + \psi_{t+1}, \quad \psi_{t+1} \sim \mathbf{N}(0, \sigma_\psi^2), \end{aligned}$$

where $\mu_{1,t}$ denotes the price level trend, δ_t is its forecast innovation, $\mu_{2,t}$ represents trend deviations from the price level, and ψ_t is its forecast innovation.⁷ When δ_{t+1} rises, the impact on $\mu_{1,t+j}$ ($j \geq 1$) and $\ln P_{t+j}$ is permanent because it never decays. The same response is generated by the shock to trend deviations from the price level, ψ_{t+1} .

The UC-LL model provides estimates of expected inflation, $\mathbf{E}_t \pi_{t+1}$.⁸ Recognize that $\Delta \ln P_t = \mu_{1,t} - \mu_{1,t-1} = \mu_{2,t} + \delta_{t+1}$. Next, use the expectations operator, $\mathbf{E}_t \{\cdot\}$, to find $\mathbf{E}_t \Delta \ln P_{t+1} \equiv \mathbf{E}_t \pi_{t+1} = \mu_{2,t}$. Thus, deviations from the price level trend provide estimates of expected inflation. These deviations are persistent—a random walk, in fact—and have innovations, ψ_{t+1} , whose impact on $\mathbf{E}_t \pi_{t+j}$ is permanent. Given MLEs of the UC-LL model, $\mathbf{E}_t \pi_{t+1}$ can be computed using the Kalman filter or smoother.⁹

Data and Sample Construction

The four series studied are CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE.¹⁰ The sample begins with 1967M01 and runs to 2005M09, providing 465 observations.

Evidence of instability in inflation, inflation growth, and price level dynamics on the MLEs of AR(p), MA(1), and UC-LL models is explored with two samples that move or roll through the entire sample. The process involves the following steps:

1. The first rolling sample always starts with 1967M01, and its initial pass through the data sets its last observation to $J = 1972M08$, which covers 15 percent of the entire sample.
2. The second rolling sample starts where the first rolling ends—at the next observation $J + 1 = 1972M09$ —and ends with 2005M09, which is the remaining 85 percent of the sample.
3. Next, the first rolling sample is extended one observation to $J = 1972M09$, which forces the second rolling sample to commence with $J + 1 = 1972M10$, but the second sample retains 2005M09 as its final observation.
4. The procedure is complete when the last observation of the first rolling sample reaches $J = 1999M09$, which is 85 percent of the entire sample, and $J + 1 = 1999M10$ is the initial observation of the second rolling sample that ends with 2005M09, which represents the other 15 percent of the sample.

Steps 1–4 create two rolling samples on the four price indexes from which 326 sets of MLEs of the AR(p), MA(1), and UC model parameters are taken.

Table 1
Summary of Empirical Results

	CPI-ALL	CPI-CORE	PCED-ALL	PCED-CORE
Mean π				
First rolling sample	Falls at end of second oil price shock	Same	Same	Same
Second rolling sample	Falls at end of first oil price shock	Same	Same	Same
Persistence				
First rolling sample	Near 1 after 1973–75 recession	Same	Same	Same
Second rolling sample	Falls during 1980 recession	Falls at 1990–91 recession	Same as CPI-ALL	Same as CPI-CORE
Volatility				
First rolling sample	Rises at first oil price shock	Same	Same	Same
Second rolling sample	Rises at 1990–91 recession	Falls at 1980 recession	Same as CPI-ALL	Falls at 1981–82 recession

Results

This section reports MLEs of AR(p), MA(1), and UC-LL models on the CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE indexes. The entire 1967M01–2005M09 sample and the two rolling samples are used to produce estimates. Table 1 summarizes the results.

AR(p) model estimates. Table 2 presents MLEs of AR(p)s. Lag lengths of the AR(p)s are set by the Schwarz information criterion (SIC), where $p = 1, \dots, 18$.¹¹ Compared to CPI-ALL and PCED-ALL inflation, CPI-CORE and PCED-CORE inflation have smaller estimated means, $\hat{\pi}_0$, and are less volatile as measured by estimates of the standard deviation of regression residuals, $\hat{\sigma}_\varepsilon$. Also, CPI inflation is higher on average and more volatile than PCED inflation given the MLEs of $\hat{\pi}_0$ and $\hat{\sigma}_\varepsilon$.

The estimates of the AR(p)s yield evidence that CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation are persistent on the 1967M01–2005M09 sample. Sums of the estimated γ_i s, $\hat{\gamma}(\mathbf{1})$ s, are all greater than 0.8. The estimates of the largest eigenvalue,

-
6. The price level random walk ignores drift produced by π_t .
 7. Harvey (1990) and Gourieroux and Monfort (1997) sketch the UC-LL model. The appendix outlines methods to estimate it and the MA(1) model.
 8. $\mathbf{E}_t\{\cdot\}$ is the mathematical expectations operator conditional on date t information.
 9. Hamilton (1994) presents methods to compute the Kalman filter and smoother; also see the appendix.
 10. The CPI equals the ratio of the date t value of a fixed market basket of goods to the same fixed market basket valued at the base year, which is 1982–84 at the moment. The PCED weights are chained to the 2000 base year. This article uses data available on December 1, 2005.
 11. Besides the SIC, the Akaike information criterion (AIC) and likelihood ratio (LR) tests are computed. It is no surprise that the AIC picks a longer lag length than the SIC for CPI-ALL ($p = 13$), CPI-CORE ($p = 9$), and PCED-ALL ($p = 13$) but $p = 6$ for PCED-CORE inflation. LR tests select p s that fall between SIC and AIC choices except for PCED-CORE inflation ($p = 13$).

Table 2

Estimates of AR(p)s of Inflation: Sample, 1967M01–2005M09

	CPI-ALL	CPI-CORE	PCED-ALL	PCED-CORE
$\hat{\pi}_0$	4.64 (0.71)	4.36 (0.73)	4.10 (0.87)	3.97 (0.92)
$\hat{\gamma}_1$	0.40 (0.07)	0.31 (0.05)	0.41 (0.06)	0.29 (0.08)
$\hat{\gamma}_2$	0.14 (0.06)	0.31 (0.07)	0.04 (0.06)	0.07 (0.05)
$\hat{\gamma}_3$	0.06 (0.05)	0.05 (0.06)	0.09 (0.05)	0.18 (0.05)
$\hat{\gamma}_4$	0.08 (0.05)	-0.06 (0.06)	0.06 (0.05)	0.10 (0.05)
$\hat{\gamma}_5$	0.15 (0.06)	0.10 (0.05)	0.15 (0.06)	0.07 (0.05)
$\hat{\gamma}_6$	—	0.17 (0.06)	0.16 (0.05)	0.22 (0.05)
$\hat{\gamma}(\mathbf{1})$	0.83 (0.05)	0.89 (0.06)	0.90 (0.04)	0.92 (0.04)
$\hat{\Lambda}$	0.94 [11.7]	0.96 [16.3]	0.97 [20.5]	0.98 [29.4]
$\hat{\sigma}_\epsilon$	2.65 (0.15)	2.03 (0.13)	1.89 (0.07)	1.50 (0.09)

Notes: Heteroskedastic-consistent asymptotic standard errors appear in parentheses. The persistence measures are $\hat{\gamma}(\mathbf{1}) = \sum_{j=1}^p \hat{\gamma}_j$ and $\hat{\Lambda}$, which is the largest eigenvalue of the estimated AR coefficients $\hat{\gamma}_j, j = 1, \dots, p$. Numbers in brackets are estimates of the half-life (in years) of the response of inflation to an own shock, according to the persistence measure $\ln 0.5 / \ln \hat{\Lambda}$.

$\hat{\Lambda}$, of the $\hat{\gamma}_j$ s are no smaller than 0.94. The $\hat{\Lambda}$ s translate into measures of the half-life of the response to an own shock of about twelve and sixteen years for CPI-ALL and CPI-CORE inflation, respectively. The PCEDs are more persistent than the CPIs because the former's $\hat{\Lambda}$ s predict between twenty and twenty-nine years for the half-life of the response to an own shock. Also, note that CORE inflation is more persistent than ALL inflation.

Table 2 presents MLEs that suggest monthly CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation are persistent. However, this is not evidence that U.S. inflation has a unit root. Rather than report unit root tests, estimates of the largest AR root of these inflation measures, along with 95 percent confidence intervals, are reported next.

Andrews and Chen (1993) develop an (approximate) median-unbiased estimator of the largest AR root of a time series.¹² The largest monthly median-unbiased AR root of CPI-ALL inflation is close to but less than 1 at 0.9965, which fails to support

the unit root hypothesis on the 1967M01–2005M09 sample. However, this estimate predicts a half-life of an own shock to CPI-ALL inflation of 16.5 years, which is long-lived relative to a sample of nearly thirty-nine years.¹³ The evidence supports a unit root in the three other inflation rates because the Andrews-Chen median-unbiased estimator yields 95 percent confidence intervals of the largest AR root of CPI-CORE, PCED-ALL, and PCED-CORE inflation equal to [0.9863 1.0004], [0.9868 1.0007], and [0.9900 1.0017], respectively.

Part of the puzzle of U.S. inflation dynamics is whether it suffers from instability. For example, Cogley and Sargent (2001) argue that shifts in the structure of monetary policy alter the process generating inflation. Since such changes can force inflation to appear nonstationary, which can be confused for a unit root, they raise questions about the stability of persistence estimates and unit root tests of inflation.

Figure 1 plots mean inflation estimates, $\hat{\pi}_{0,t}$, constructed on the two rolling samples and four inflation measures.¹⁴ The first rolling sample produces $\hat{\pi}_{0,t}$ on observations that always begin with 1967M01 and end with $J = 1972M08, \dots, 1999M09$. For example, the first element of the line plotting the first rolling sample is estimated on a 1967M01–1972M08 sample, the second element on a 1967M01–1972M09 sample, and so on. The figure also includes plots of $\hat{\pi}_{0,t}$ estimated on the second rolling sample, which runs from $J + 1 = 1972M09, \dots, 1999M10$ to 2005M09. Thus, plots of $\hat{\pi}_{0,t}$ are obtained from the first rolling sample by adding an observation to its end at each date J , while the second rolling sample sequentially eliminates the initial observation as $J + 1$ advances from 1972M09 to 1999M10.

The four windows of Figure 1 show that $\hat{\pi}_{0,t}$ fell during several of the recessions of the last forty years. The largest drop in $\hat{\pi}_{0,t}$ for the four inflation measures occurs in the 1973–75 recession. However, $\hat{\pi}_{0,t}$ rises for CPI inflation in the 1980 recession in the first rolling sample, while $\hat{\pi}_{0,t}$ shows little change for PCED inflation in the same period for this sample. The recessions of 1980 and 1981–82 see lower $\hat{\pi}_{0,t}$ for the four inflation measures on the second rolling sample. The first rolling sample also finds $\hat{\pi}_{0,t}$ drops in the recession of 1981–82. Subsequent to this recession, $\hat{\pi}_{0,t}$ falls, continuing for the four inflation rates on the rest of the two rolling samples. By the end of the first (second) rolling sample, CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE $\hat{\pi}_{0,t}$ equal 4.8 (2.8), 4.8 (2.1), 4.1 (2.3), and 4.0 (1.7 percent), respectively.

Figure 1 also shows that prior to mid-1974 the second rolling sample yields larger $\hat{\pi}_{0,t}$ than the first rolling sample across all inflation measures. This pattern is reversed subsequent to the end of the 1973–75 recession. The first rolling sample also reveals that $\hat{\pi}_{0,t}$ begins to drift higher after 1975 and peaks prior to the 1980 recession at about 7 percent for CPI inflation, 6 percent for PCED-ALL inflation, and 5.5 percent for PCED-CORE inflation. Since the second rolling sample produces a fall in mean inflation prior to (or around) the 1980 recession, it points to possible instability in $\hat{\pi}_{0,t}$ for the four inflation rates toward the end of the 1970s.

-
12. Estimates of the largest AR root rely on AR(6)s that contain an intercept but not a time trend.
13. Stock (1991) and Andrews and Chen (1993) discuss that a least squares estimate of the largest AR root of a unit root process is biased downward. This result explains the smaller root of CPI-ALL inflation to an own shock reported in Table 2 compared to the estimate of 0.9965 of the median-unbiased estimator.
14. The two rolling samples yield MLEs that suggest using tests, say, by Andrews (1993), of parameter instability given an unknown break date. The problem is that the Hansen (1997) and Andrews (2003) critical values cannot always be used because the two rolling samples produce MLEs of the AR(p)s, MA(1), and UC-LL models that are often on the boundary of the permissible parameter space, which implies that critical values would have to be constructed on a case-by-case basis.

Figure 1
Time Variation in Mean Inflation

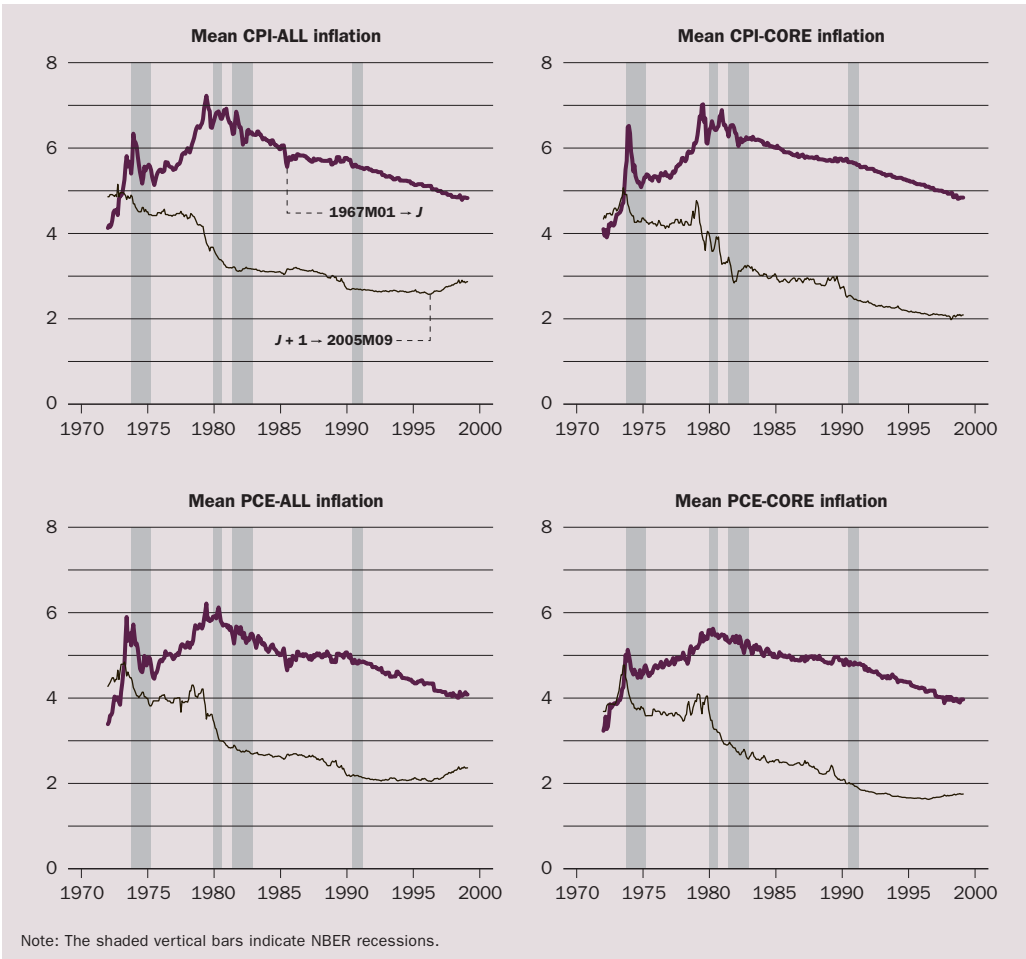
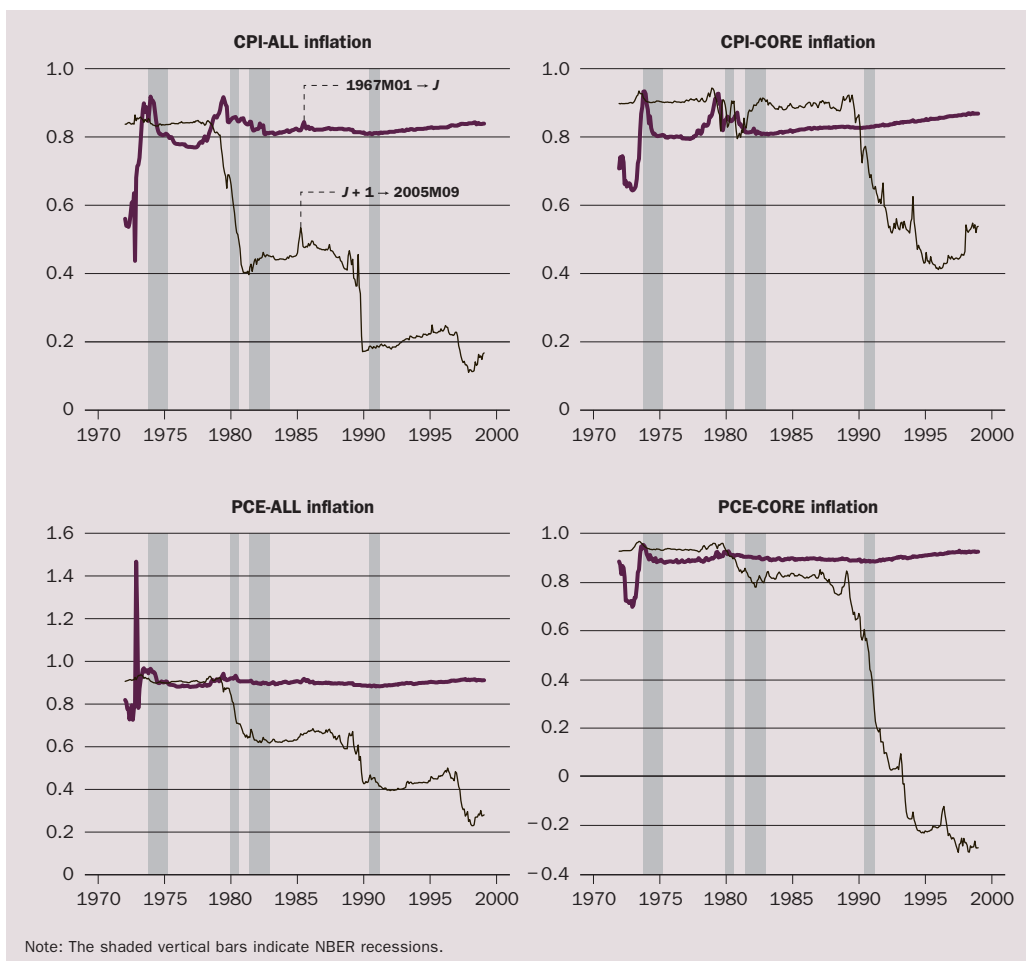


Figure 2 presents estimates of time variation in inflation persistence, $\hat{\gamma}(\mathbf{1})_t$.¹⁵ The first rolling sample yields plots of $\hat{\gamma}(\mathbf{1})_t$ that remain close to but below 1 from just before the 1973–75 recession to 1999M09. The second rolling sample generates $\hat{\gamma}(\mathbf{1})_t$ that are close to 1 until the 1980 recession for the ALL inflation rates and until the 1990–91 recession for the CORE inflation series. Prior to the latter recession, the second rolling sample generates $\hat{\gamma}(\mathbf{1})_t$ that drop from 0.9 to nearly 0.4 in 1996 for CPI-CORE inflation before rising to about 0.55 in 1999. PCED-CORE inflation persistence exhibits the same behavior except that its $\hat{\gamma}(\mathbf{1})_t$ turns negative in the mid-1990s and remains negative for the remainder of the second rolling sample; this downturn suggests either a negatively serially correlated or a serially uncorrelated process once observations from the 1970s and 1980s are dropped. Thus, eliminating observations from the 1970s and 1980s leads to smaller persistence estimates for the four inflation series.

Sargent (1999) provides an interpretation of the first and second rolling sample $\hat{\gamma}(\mathbf{1})_t$ s found in Figure 2. In his analysis, a key element is the interaction of beliefs about monetary policy and the discount applied to past observations, say, on inflation. For example, discounting past observations can lead to less inflation persistence

Figure 2
Time Variation in AR Coefficient Sums



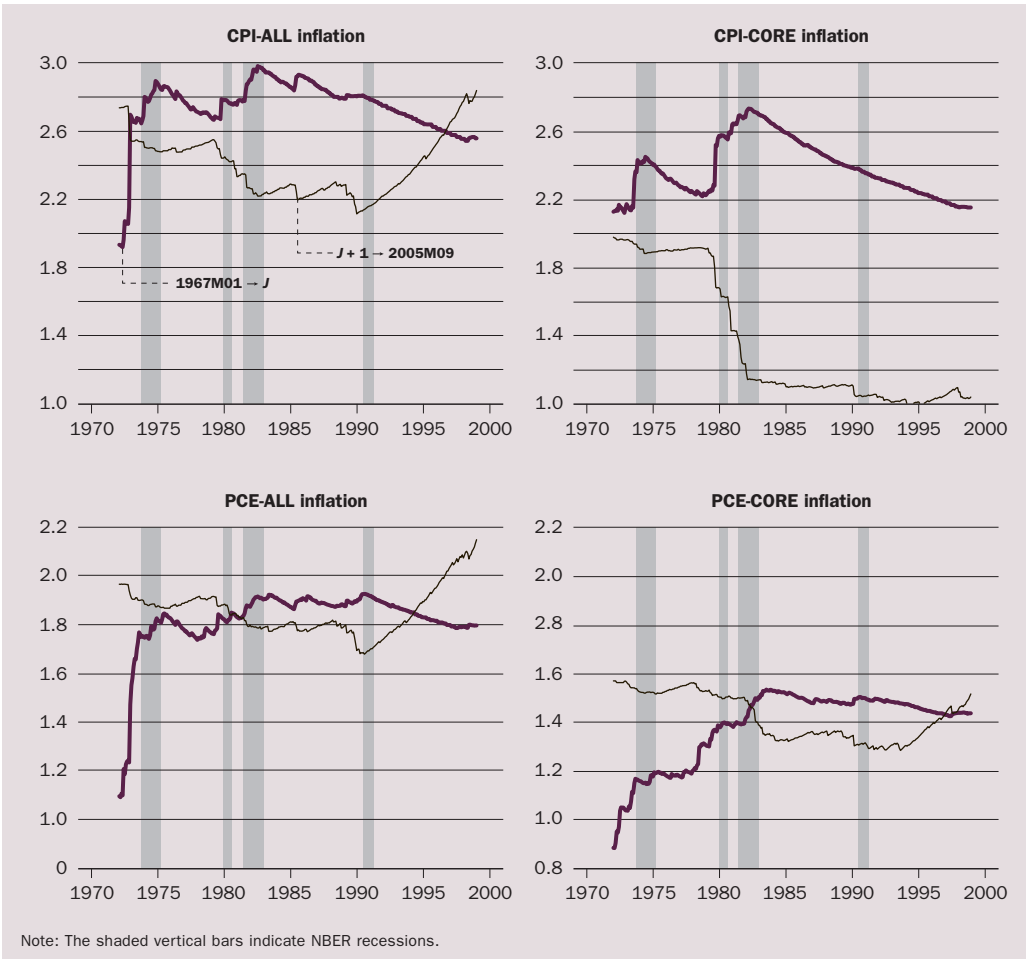
because, according to Sargent, discounting is a reasonable response by monetary policy if inflation dynamics are suspected of being unstable. Whether this explanation accounts for the past forty years of U.S. inflation and monetary policy is not addressed by this article, but Cogley and Sargent (2005) and Sargent, Williams, and Zha (forthcoming) provide useful analyses.

Figure 3 presents plots of the volatility of the four inflation series measured by $\hat{\sigma}_{\epsilon,t}$. Given the first rolling sample, plots of $\hat{\sigma}_{\epsilon,t}$ support the view that CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation volatility began to increase around the first oil price shock of the mid-1970s, continuing until the end of the 1981–82 recession before leveling off or declining in the early to mid-1980s.

The second rolling sample generates $\hat{\sigma}_{\epsilon,t}$ plots that give a different view of inflation volatility. Figure 3 shows that volatility in the four inflation measures began to drop off subsequent either to the second oil price shock or to the 1981–82 recession based

15. Plots of Λ yield the same qualitative evidence about persistence for CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation using the two rolling samples. These plots are available on request.

Figure 3
Time Variation in Standard Deviation of Inflation AR Residuals



on the second rolling sample. CPI-CORE inflation has the largest fall in $\hat{\sigma}_{\epsilon,t}$, while the other three inflation measures decline less.

There is no consistent pattern to inflation volatility instability according to the $AR(p)$ model estimates. For CPI-ALL inflation, instability in $\hat{\sigma}_{\epsilon,t}$ possibly exists between the 1980 recession and the late 1990s. The 1980 recession begins a period of instability in $\hat{\sigma}_{\epsilon,t}$ for CPI-CORE inflation. The beginning and end of the two rolling samples most likely suggest when instability in $\hat{\sigma}_{\epsilon,t}$ of PCE-ALL inflation can be found. For PCE-CORE inflation, instability in $\hat{\sigma}_{\epsilon,t}$ appears to occur from 1972M08 to the 1980 recession. Thus, there seems to be no consistent pattern of instability in $\hat{\sigma}_{\epsilon,t}$ for the four inflation rates.

MA(1) model estimates. Table 3 reports estimates of the MA(1) coefficient, $\hat{\theta}$, and the standard deviation of the MA(1) residual, $\hat{\sigma}_{\eta}$, on the 1967M01–2005M09 sample. Estimates of $\hat{\theta}$ are similar across the four inflation growth measures. The point estimates range from 0.72 to 0.78, which predict that within one month inflation growth loses about three-fourths of the increase caused by an own-unit shock. However, inflation growth is lower by about 2.5 to 3.5 percent in the long run given such a shock,

Table 3
Estimates of Inflation Growth MA(1): Sample, 1967M01–2005M09

	CPI-ALL	CPI-CORE	PCED-ALL	PCED-CORE
$\hat{\theta}$	0.75 (0.05)	0.72 (0.06)	0.75 (0.05)	0.78 (0.05)
$\hat{\sigma}_\eta$	2.67 (0.15)	2.07 (0.15)	1.94 (0.08)	1.53 (0.09)
$\hat{\theta}(1)$	-2.97 (0.39)	-2.56 (0.34)	-2.93 (0.37)	-3.45 (0.49)

Note: Heteroskedastic-consistent asymptotic standard errors appear in parentheses.

as measured by $\hat{\theta}(1)$. This result indicates inflation growth is subject to large low-frequency fluctuations. Not unexpectedly, $\hat{\sigma}_\eta$ shows that CPI-ALL (PCED-ALL) inflation growth is more volatile than CPI-CORE (PCED-CORE) inflation growth. PCED-ALL and PCED-CORE inflation growth are also less volatile than their CPI counterparts.

Figure 4 suggests instability in $\hat{\theta}_t$ across the four inflation growth measures and the two rolling samples. Instability in $\hat{\theta}_t$ appears to arise between the recessions of 1980 and 1981–82 for CPI-ALL inflation growth and the first oil price shock for CPI-CORE, PCED-ALL, and PCED-CORE inflation growth. Also, the first rolling sample generates $\hat{\theta}_t$ that are close to 0.70 and stable subsequent to the 1981–82 recession. The plots of $\hat{\theta}_t$ drift toward 1 for the four inflation growth measures on the second rolling sample around the 1980 recession for the CPI inflation growth rates and the 1973–75 recession for the PCED inflation growth rates.

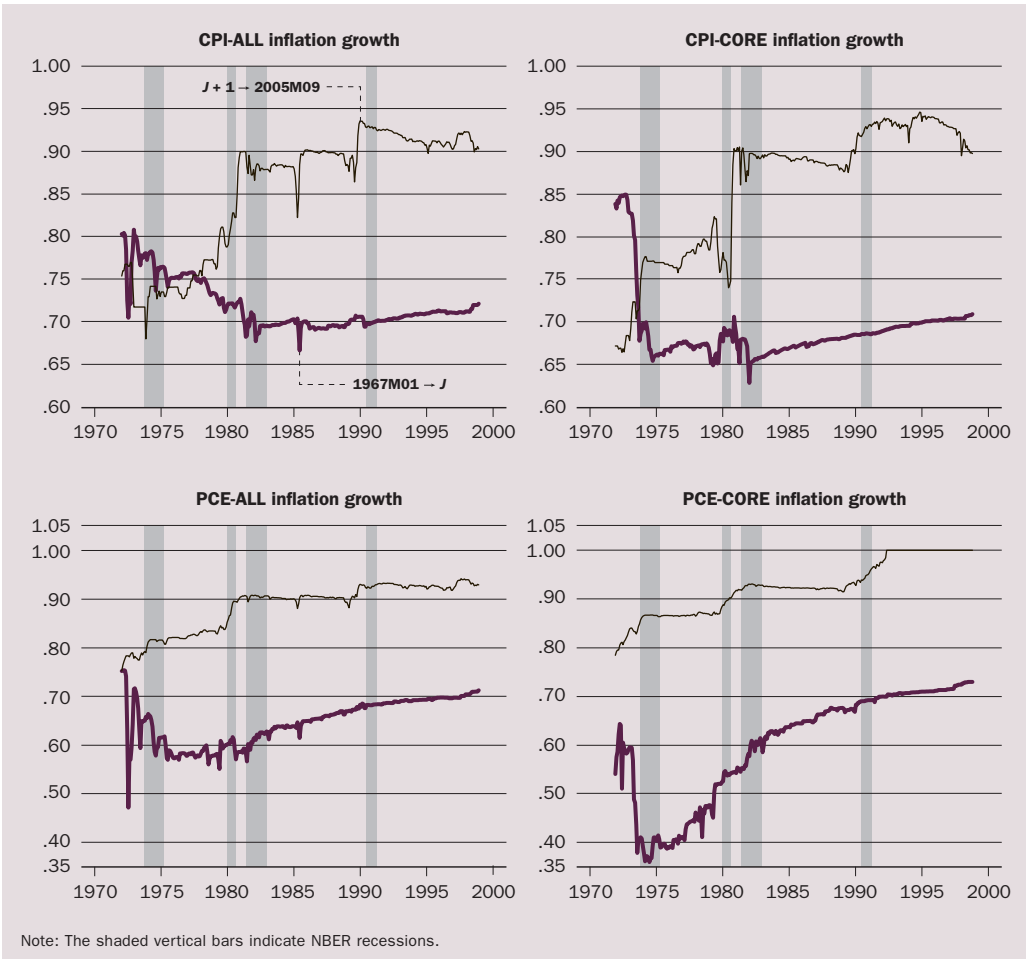
A prominent feature of the bottom right window of Figure 4 is that $\hat{\theta}_t = 1$ for PCED-CORE inflation growth on the second rolling sample from 1992M04 to 1999M10. The impact on inflation dynamics is that the MA(1) collapses to $\pi_t = \eta_t$ (ignoring mean inflation) when $\theta = 1$. Thus, the second rolling sample yields $\hat{\theta}_t$ that predict PCED-CORE inflation is driven only by white noise shocks from the recovery of the early 1990s to 1999M10. This result matches the small AR persistence estimates reported in this article for PCED-CORE inflation and evidence reported by Stock and Watson (2005).

Figure 5 contains plots of $\hat{\sigma}_{\eta,t}$ for the four inflation growth series. CPI-ALL inflation growth and the first rolling sample yield $\hat{\sigma}_{\eta,t}$ that are below those of the second rolling sample except around the first oil price shock. CPI-CORE inflation growth and the first rolling sample produce $\hat{\sigma}_{\eta,t}$ that are always above those of the second rolling samples. The second oil price shock matters for CPI-CORE inflation growth volatility because around this time $\hat{\sigma}_{\eta,t}$ falls by 45 and 70 percent on its first and second rolling samples.

The bottom row of graphs in Figure 5 includes plots of $\hat{\sigma}_{\eta,t}$ that qualitatively resemble plots of $\hat{\sigma}_{\varepsilon,t}$ for PCED-ALL and PCED-CORE inflation in Figure 3. Thus, the MA(1) and AR(p) models produce PCED inflation growth and inflation volatility estimates that are similar.

UC-LL model estimates. Table 4 provides estimates of the standard deviations of innovations to the price level trend, $\hat{\sigma}_\delta$, and to price level trend deviations, $\hat{\sigma}_\psi$. The largest estimates of $\hat{\sigma}_\delta$ and $\hat{\sigma}_\psi$ are obtained from the CPI indexes. The last row of the table shows that shocks to the price level trend dominate fluctuations in the four

Figure 4
Time Variation in MA(1) Coefficient of Inflation Growth

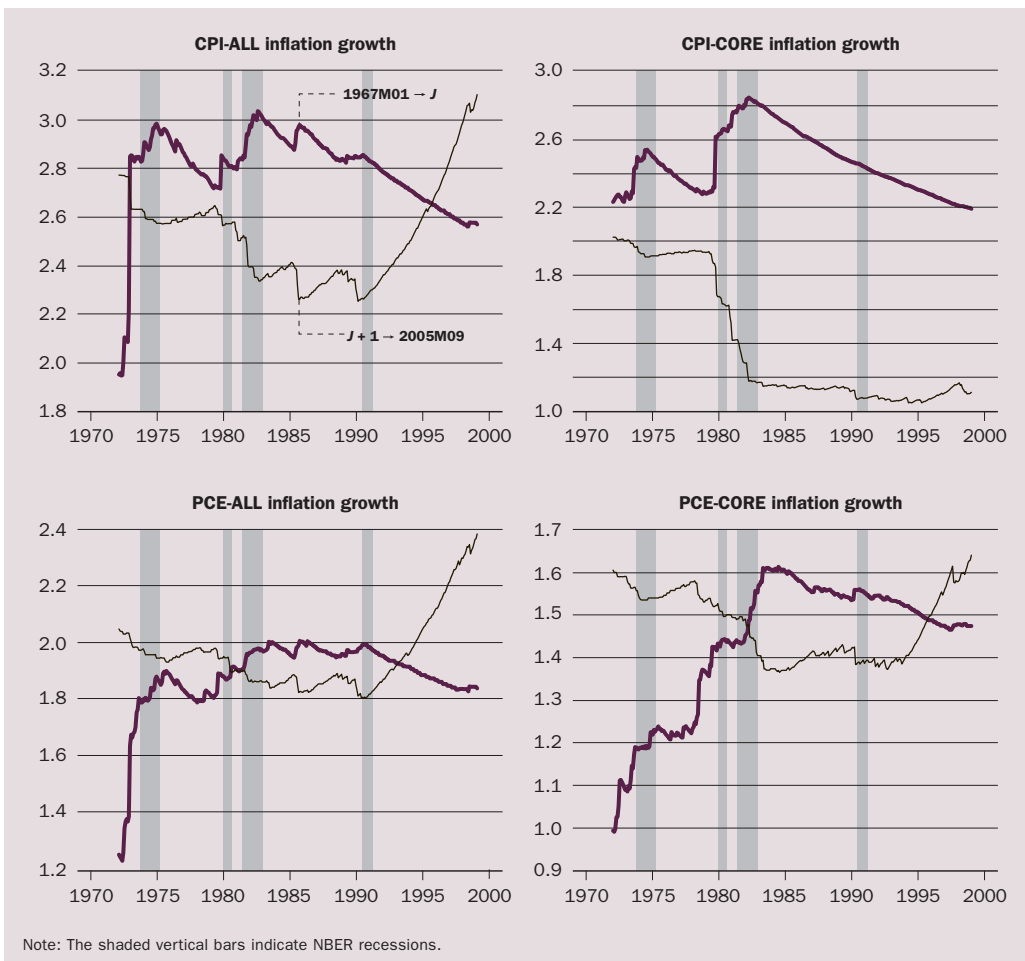


price indexes because $\hat{\sigma}_\delta$ is always larger than $\hat{\sigma}_\psi$ by a factor of almost three to four. Also, note that the largest estimated ratio of $\hat{\sigma}_\delta$ to $\hat{\sigma}_\psi$ is for PCED-CORE.

The UC-LL model predicts that inflationary expectations equal trend price level deviations, $\mathbf{E}_t \pi_{t+1} = \mu_{2,t}$. Figure 6 contains smoothed and filtered estimates of $\hat{\mu}_{2,t}$ computed from MLEs of the UC-LL model for the four price indexes on the 1967M01–2005M09 sample. Although the filtered $\hat{\mu}_{2,t}$ are more volatile and “choppier” than smoothed $\hat{\mu}_{2,t}$, the latter have earlier turning points because smoothing employs information in the entire 1967M01–2005M09 sample. Only observations from 1967M01 to date t are available to compute filtered $\hat{\mu}_{2,t}$.¹⁶

Estimates of $\mathbf{E}_t \pi_{t+1}$ reveal that the relatively small $\hat{\sigma}_\psi$ of Table 3 generates economically important fluctuations in inflationary expectations. For example, CPI estimates of $\mathbf{E}_t \pi_{t+1}$ peak during every recession during the 1967M01–2005M09 sample except the 2001 recession, as found in the top row of graphs in Figure 6. These plots show peaks in CPI-ALL and CPI-CORE filtered (smoothed) expected annual inflation rates of 12.4 and 12.7 (10.9 and 11.6) percent at 1974M08 and 1974M07 (1974M06 and 1974M06) and 14.8 and 14.2 (13.4 and 12.8) percent at 1980M03 (1979M12).

Figure 5
Time Variation in Standard Deviation of Inflation Growth MA(1) Residuals



Subsequently, filtered (smoothed) $\mathbf{E}_t \pi_{t+1}$ falls to -0.9 (0.7) percent by 1986M04 (1986M02) for CPI-ALL and to 3.4 (3.6) percent by 1986M05 (1986M04) for CPI-CORE. At 1990M09 and 1990M07 (1990M06 and 1990M05), filtered (smoothed) CPI-ALL and CPI-CORE $\mathbf{E}_t \pi_{t+1}$ peak at about 7 and 6 (6 and 5) percent. From 1992 to 2004, filtered (smoothed) CPI-ALL and CPI-CORE $\mathbf{E}_t \pi_{t+1}$ are no higher than 3.5 and 3.7 percent and no smaller than -0.02 and 0.08 percent before reaching 6.7 and 1.5 percent by 2005M09.

PCED-ALL and PCED-CORE $\mathbf{E}_t \pi_{t+1}$ appear in the bottom two rows of Figure 6. These measures of inflation are qualitatively similar to those for the CPI indexes in the top row of graphs, but estimates of $\mathbf{E}_t \pi_{t+1}$ peak only during the 1973–75, 1980, and 1990–91 recessions. Peaks in PCED-ALL and PCED-CORE $\mathbf{E}_t \pi_{t+1}$ are successively lower at each recession irrespective of the filtered or smoothed estimates. These estimates

16. Filtered and smoothed $\hat{\mu}_{2,t}$ are generated with the Kalman filter, as discussed by Hamilton (1994). Smoothed $\hat{\mu}_{2,t}$ involves a two-sided (in-sample) forecast (using the full data set), while filtered $\hat{\mu}_{2,t}$ is a one-sided forecast given observations $1, \dots, t$. Filtered $\hat{\mu}_{2,t}$ is initialized with $\mathbf{E}_0 \pi_{1967M01} = 0$.

Table 4
Estimates of the UC-LL Model: Sample, 1967M01–2005M09

	CPI-ALL	CPI-CORE	PCED-ALL	PCED-CORE
$\hat{\sigma}_{\delta}$	2.31 (0.16)	1.74 (0.15)	1.67 (0.08)	1.34 (0.05)
$\hat{\sigma}_{\psi}$	0.68 (0.14)	0.61 (0.14)	0.51 (0.10)	0.36 (0.07)
$\hat{\sigma}_{\delta}/\hat{\sigma}_{\psi}$	3.38 (0.76)	2.85 (0.73)	3.29 (0.74)	3.77 (0.86)

Note: Heteroskedastic-consistent asymptotic standard errors appear in parentheses.

are between 9 and 10.5 percent during the 1973–75 recession and 1974M06, 9.5 and 12 percent at the 1980 recession, and 4.5 to 5 percent for the 1990–91 recession. From 1992 through 2004, PCED-ALL and PCED-CORE $\mathbf{E}_t\pi_{t+1}$ range from 0.5 to 3 percent. By 2005M09, PCED-ALL and PCED-CORE $\mathbf{E}_t\pi_{t+1}$ equal 5.2 and 1.8 percent, respectively.

Information about parameter instability in the MLEs of σ_{δ} and σ_{ψ} appears in Figures 7 and 8. Parameter instability in the UC-LL model garners information about changing CPI-ALL, CPI-CORE, PCED-ALL, and PCI-CORE price dynamics. This information is useful to understanding whether the declines in $\mathbf{E}_t\pi_{t+1}$ subsequent to the recession of the early 1980s that appear in Figure 6 are related to small shock realizations to trend deviations from the price level, ψ_t , or to instability in the volatility of this shock, σ_{ψ} .

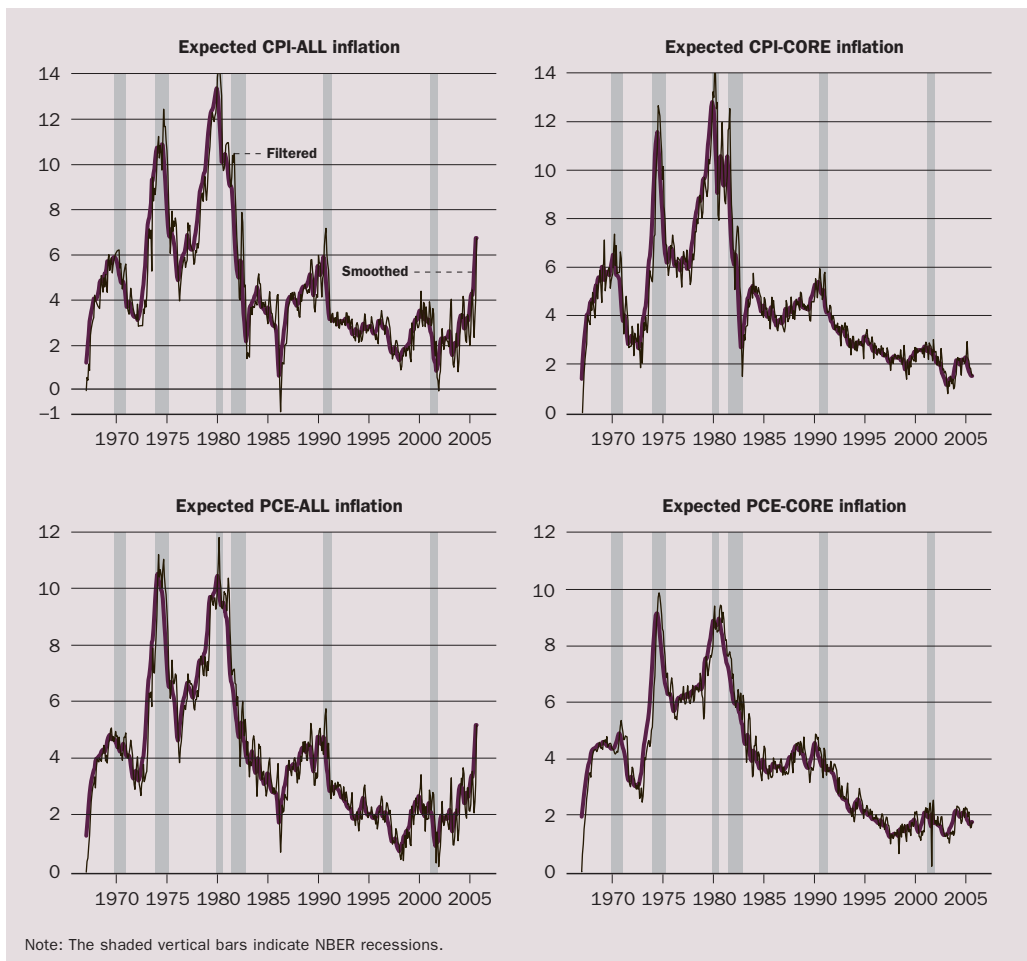
Figure 7 contains four graphs that plot the first and second rolling sample estimates of the standard deviation of the price level trend shock innovation, $\hat{\sigma}_{\delta,t}$. These estimates suggest instability in $\hat{\sigma}_{\delta,t}$. The instability in $\hat{\sigma}_{\delta,t}$ appears to arise in the late 1990s for the ALL price indexes. Evidence of a break in $\hat{\sigma}_{\delta,t}$ for CPI-CORE is suggested by its drop in the second rolling sample at the end of the 1980 recession. For PCED-CORE, the instability in $\hat{\sigma}_{\delta,t}$ possibly occurs during the first oil price shock.

Another feature of Figure 7 is that the second rolling sample generates $\hat{\sigma}_{\delta,t}$ with little movement until the early 1990s, when it begins to rise steadily for CPI-ALL, PCED-ALL, and PCED-CORE. CPI-CORE is the exception because for the second rolling sample $\hat{\sigma}_{\delta,t}$ falls from around 1.75 in mid-1979 to slightly greater than 1 from late 1983 to late 1999. Also note that $\hat{\sigma}_{\delta,t}$ rises, for the most part, from 1972M07 to 1999M10 for PCED-ALL and PCED-CORE for the two rolling samples.

Figure 8 replicates Figure 7 except that $\hat{\sigma}_{\psi,t}$ replaces $\hat{\sigma}_{\delta,t}$.¹⁷ Instability in $\hat{\sigma}_{\psi,t}$ is possible for CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE according to Figure 8. The CPIs and the two rolling samples suggest that instability begins at the 1980 recession. For PCED-ALL and PCED-CORE, instability seems to start with the 1973–75 recession.

The two rolling samples have different implications for the paths of $\hat{\sigma}_{\psi,t}$. The first rolling sample yields $\hat{\sigma}_{\delta,t}$ that range from about 0.4 to 0.8 for CPI-ALL, PCED-ALL, and PCED-CORE and between 0.5 and 0.9 for CPI-CORE. Thus, adding observations to the first rolling sample produces $\hat{\sigma}_{\psi,t}$ that do not fall by much. The second rolling sample shows that $\hat{\sigma}_{\psi,t}$ falls to around 0.2 for CPI-ALL, CPI-CORE, and PCED-ALL

Figure 6
UC-LL Estimates of Expected Inflation

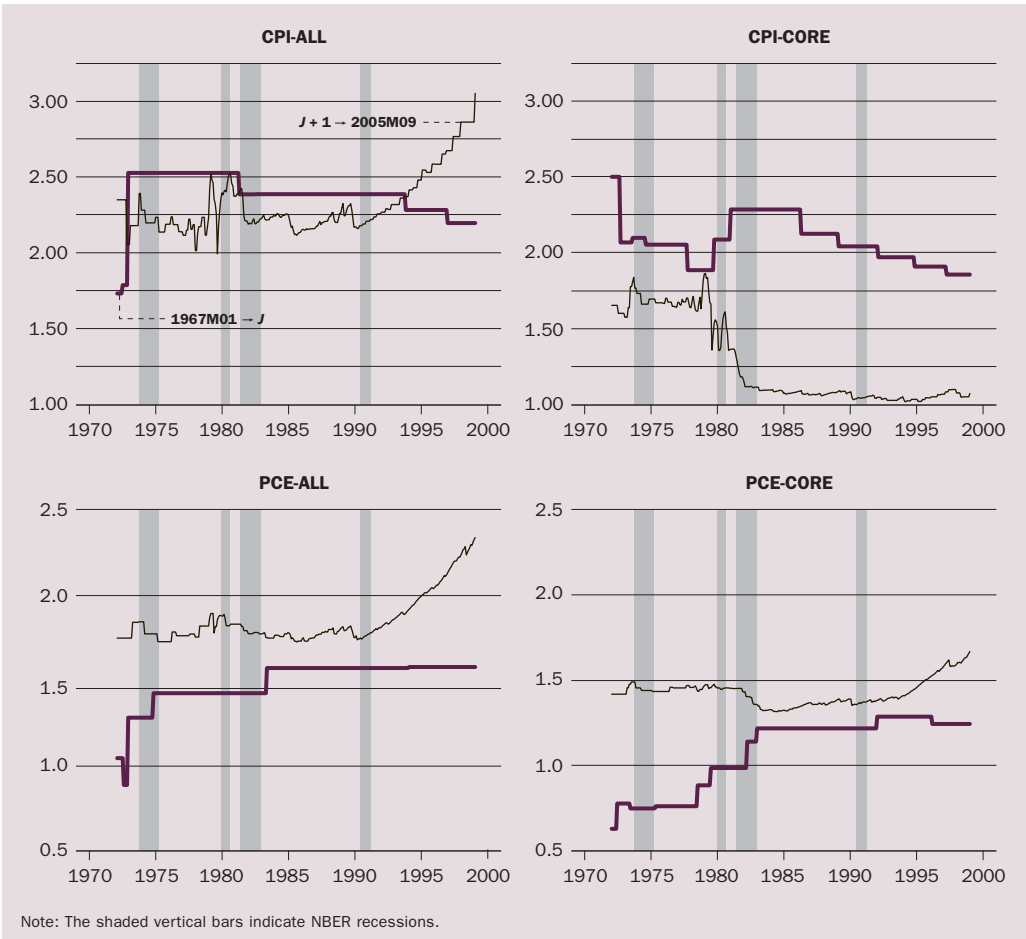


subsequent to the 1980 recession. For PCED-CORE, the drop in $\hat{\sigma}_{\psi,t}$ to 0.2 occurs around the 1973–75 recession. These estimates suggest that smaller realizations of $\hat{\psi}_t$ (that imply smaller $\hat{\sigma}_{\psi,t}$) are responsible for the fall in $\mathbf{E}_t\pi_{t+1}$ subsequent to the 1980 recession, as plotted in Figure 6.

The lower right graph of Figure 8 reveals that PCED-CORE and the second rolling sample drive $\hat{\sigma}_{\psi,t}$ to 0 by 1992M08, where it remains through 1999M10. If $\sigma_{\psi} = 0$, the UC-LL model predicts that the price level is a random walk driven by δ_t . An implication is that PCED-CORE inflation resembles a white noise process when observations from the 1970s, 1980s, and early 1990s are eliminated from the second rolling sample. A similar result is reported in this study for the $AR(p)$ and $MA(1)$ models and by Stock and Watson (2005).

17. Figures 7 and 8 contain plots of $\hat{\sigma}_{\psi,t}$ and $\hat{\sigma}_{\delta,t}$ that appear to be step functions on the first rolling sample. The mapping from the $MA(1)$ coefficients θ and σ_{η} to σ_{ψ} and σ_{δ} is one explanation for this observation. The recursive mapping is $(1 + \theta^2) \sigma_{\eta}^2 = 2\sigma_{\delta}^2 + \sigma_{\psi}^2$ and $-\theta\sigma_{\eta}^2 = -\sigma_{\delta}^2$. Watson (1986) and Morley, Nelson, and Zivot (2003) review the link between UC and ARMA models.

Figure 7
Time Variation in Standard Deviation of Price Level Trend



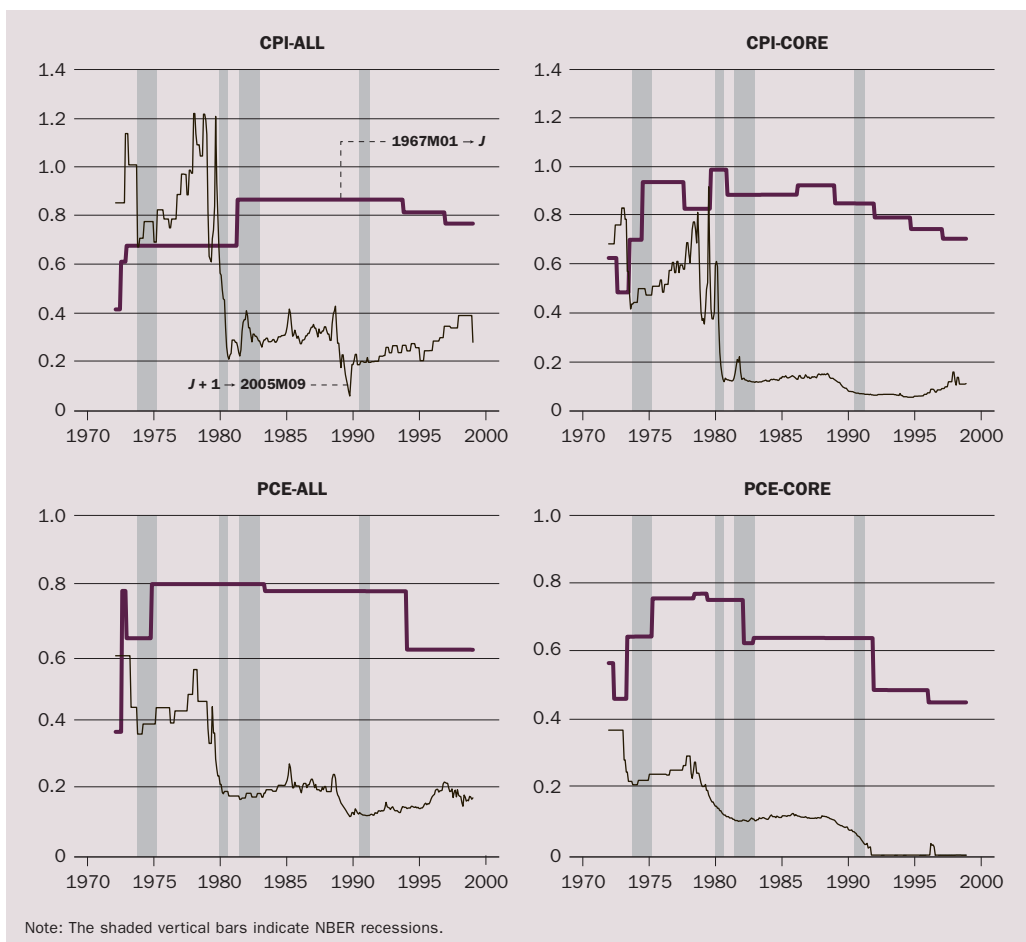
Conclusions

This article studies U.S. inflation, inflation growth, and price level dynamics with the CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE on a sample that runs from 1967M01 to 2005M09. Two rolling samples are constructed to uncover evidence about instability in inflation, inflation growth, and price level dynamics.

Autoregressive models produce persistence and volatility estimates that vary with different combinations of the two rolling samples and four price indexes. For example, inflation and inflation growth persistence estimates differ across CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE and are sensitive to observations from the 1970s, 1980s, and early 1990s. For example, inflation persistence appears to be large and stable if these observations are included in the sample. However, instabilities appear to arise when the observations are discounted. Equally striking is that PCED-CORE inflation approximates serially uncorrelated white noise when observations up to and including the recession of 1990–91 are eliminated.

Inflation, inflation growth, and price level volatility estimates behave similarly across CPI-ALL, CPI-CORE, PCED-ALL, PCED-CORE, and the two rolling samples. An impor-

Figure 8
Time Variation in Standard Deviation of Price Level Trend Deviations



tant example is that the volatility of shocks to expected inflation has fallen substantially for the four price indexes either prior to or during the 1980 recession. Especially striking is the lack of volatility in these shocks for PCED-CORE subsequent to the 1990–91 recession. Along with the AR persistence estimates, it suggests that at the moment a sensible forecast for PCED-CORE inflation is its mean on the 1992–2005 sample.

Another way to summarize the empirical results of this article is that instability in the persistence and volatility of CPI-ALL, CPI-CORE, PCED-ALL, and PCED-CORE inflation, inflation growth, and levels coincides with different economic events. An unresolved question is whether such changes are one-time events or can be expected to be repeated systematically in the future. For example, was the decline in PCED-CORE inflation persistence around the end of the 1990–91 recession caused by changes in beliefs about the systematic engineering of monetary policy, or did it reflect technology innovations, changes in market structure, or changes in the composition of the economy (that is, away from manufacturing to the service sector)? Such questions pose a challenge to economic research, forecasting, and monetary policy. Any response should find useful the tools developed by Sargent (1999), Cogley and Sargent (2005), Sims and Zha (2006), Sargent, Williams, and Zha (forthcoming), and Brock, Durlauf, and West (forthcoming).

Appendix
The Models

The models studied in this article are a p th-order autoregression, AR(p); a first-order moving average, MA(1); and an unobserved components (UC) structure. The choice of these models to study inflation, inflation growth, and price level dynamics is guided by the literature on inflation dynamics. For example, Stock and Watson (2005) report estimates of AR(p), MA(1), and UC models on quarterly gross domestic product (GDP) deflator inflation, PCED-ALL, and PCED-CORE inflation and inflation growth. This article employs similar models but includes estimates on monthly samples of PCED-ALL and PCED-CORE as well as the CPI-ALL and CPI-CORE.

The univariate AR(p) yields estimates of mean inflation, inflation persistence, and inflation volatility. In deviations from mean inflation, π_0 , the AR(p) model is

$$(A.1) \quad \pi_t - \pi_0 = \sum_{j=1}^p \gamma_j (\pi_{t-j} - \pi_0) + \varepsilon_t,$$

where inflation, π_t , is defined as the difference between the (natural) log of the month t price level, P_t , and month $t - 1$ price level, $\pi_t \equiv 1200 \times (1 - \mathbf{L}) \ln P_t$; the lag operator \mathbf{L} produces $\mathbf{L} \ln P_t \equiv \ln P_{t-1}$; and ε_t is the Gaussian inflation forecast innovation with standard deviation σ_ε . The article reports MLEs of π_0 , $\gamma_1, \dots, \gamma_p$, and σ_ε from Kalman filter iterations of the state space model

$$(A.2) \quad \begin{aligned} \pi_t &= \pi_0 + \mathbf{e}_1 \Xi_t, \\ \Xi_{t+1} &= \mathbf{\Gamma} \Xi_t + \mathbf{\varepsilon}_{t+1}, \end{aligned}$$

where the first equation is the observer equation, the second equation is the state equation, $\Xi_t = [\xi_{1,t} \dots \xi_{p,t}]'$ is the $p \times 1$ state vector, the row vector $\mathbf{e}_1 = [1 \ \mathbf{0}_p]$, $\mathbf{0}_p$ is a $1 \times (p - 1)$ vector of zeros, $\mathbf{\varepsilon}_t = [\varepsilon_t \ \mathbf{0}_p]'$, and $\mathbf{\Gamma}$ is the companion matrix of the γ_j s:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \dots & \gamma_{p-1} & \gamma_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

The state vector Ξ_t is initialized with a vector of zeros because under the null of the AR(p) the eigenvalues of $\mathbf{\Gamma}$ are inside the unit circle. At date $t = 0$, the mean square error of Ξ_t is set to $[\mathbf{I}_p - \mathbf{\Gamma} \otimes \mathbf{\Gamma}]^{-1} \text{vec}(\mathbf{E}\{E_t E_t'\})$ where $\text{vec}(\cdot)$ denotes placing the second column below the first, the third column below the previous two, and so on. Hamilton (1994) discusses in detail the Kalman filter approach to MLE of ARMA models.

Information about inflation persistence is contained in the γ_j s. One measure of inflation persistence is denoted with the sum $\gamma(\mathbf{1}) \equiv \sum_{j=1}^p \gamma_j$. Another is the largest AR root of inflation, which can be found by computing the largest eigenvalue of $\mathbf{\Gamma}$, $\Lambda(\mathbf{\Gamma})$. Since $\gamma(\mathbf{1})$ and $\Lambda(\mathbf{\Gamma})$ are functions of the AR coefficients, $\gamma_1 \dots \gamma_p$, these statistics reveal different aspects of inflation persistence. A metric of the cumulative response of inflation to an own shock is $\gamma(\mathbf{1})$. The largest eigenvalue of the companion matrix $\mathbf{\Gamma}$ measures the speed of adjustment of inflation to an own shock along the transition path. The speed of adjustment is translated into the length of time inflation takes to return halfway to its (long-run) mean with $\ln 0.5 / \ln \Lambda(\mathbf{\Gamma})$. Inflation persistence rises as $\gamma(\mathbf{1})$ and $\Lambda(\mathbf{\Gamma})$ approach 1 (from below).¹

As $\gamma(\mathbf{1}) \rightarrow 1$, inflation persistence increases. In the limit, inflation takes on a unit root and becomes nonstationary. Stock and Watson (2005) report estimates of $\gamma(\mathbf{1}) \geq 1$ that point to a unit root in quarterly inflation since 1970. A lesson they draw is that it is better to work with inflation growth instead of the level of inflation. This conclusion leads them to advocate a model in which inflation growth is decomposed into unobserved permanent (that is, a unit root or random walk) and transitory components

$$\begin{aligned} \pi_t &= \mu_{\pi,t} + v_t, \quad v_t \sim \mathbf{N}(0, \sigma_v^2), \\ \mu_{\pi,t+1} &= \mu_{\pi,t} + \tau_{t+1}, \quad \tau_{t+1} \sim \mathbf{N}(0, \sigma_\tau^2), \end{aligned}$$

where $\mu_{\pi,t}$ is the permanent component of inflation, its forecast innovation is τ_t , and v_t denotes the transitory shock innovation.² Also assume $\mathbf{E}\{v_{t+j} \tau_{t+j}\} = 0, \forall_{q,j}$.

The reduced form of the Stock and Watson (SW-)UC model is a first-order MA. The

reduced-form MA(1) is constructed by passing the first difference operator, $1 - \mathbf{L}$, through $\pi_t = \mu_{\pi,t} + \nu_t$ and substituting for $(1 - \mathbf{L})\mu_{\pi,t} = \tau_t$ to find $(1 - \mathbf{L})\pi_t = \tau_t + \nu_t - \nu_{t-1}$, with the one-step-ahead forecast error $\eta_t \equiv \tau_t + \nu_t + \mu_{\pi,t} - \mu_{\pi,t-1|t-1}$, where $\mu_{\pi,t-1|t-1}$ is conditional on observations through date $t - 1$. The first-order moving average dynamics of inflation growth motivates studying it with the fixed-coefficient MA(1),

$$(A.3) \quad (1 - \mathbf{L})\pi_t = (1 - \theta\mathbf{L})\eta_t,$$

to obtain evidence of changes in inflation growth persistence, as measured by θ . In this case, time variation in σ_ν and σ_τ drives changes in inflation growth persistence and volatility. The map between the SW-UC model and the MA(1) of equation (A.3) consists of $(1 + \theta^2)\sigma_\eta^2 = 2\sigma_\nu^2 + \sigma_\tau^2$ and $-\theta\sigma_\eta^2 = -\sigma_\nu^2$.

MLEs of θ and σ_η are obtained from iterating the Kalman filter. The filter is initialized following the procedure outlined for the AR(p) of equation (A.1). The article reports estimates of θ and σ_η on the two rolling samples corrected for Blaschke factors when necessary. Hansen and Sargent (1980) and Hamilton (1994) show how to extract Blaschke factors of noninvertible MA processes to adjust MLEs of θ and σ_η to obtain an invertible MA.

Estimates of the MA(1) coefficient θ of equation (A.3) contain information about inflation growth persistence. The MA(1) of equation (A.3) yields the AR(∞), $(1 - \mathbf{L})\pi_t = \sum_{j=1}^{\infty} \vartheta_j(1 - \mathbf{L})\pi_{t-j} + \eta_t$, where $\vartheta_j = \theta^j$, given $|\theta| \in (-1, 1)$. The sum $\vartheta(\mathbf{1})$ equals $-\theta/1 - \theta$. Therefore, the long-run response of inflation growth to an own shock increases as $\theta \rightarrow 1$. At $\theta = 1$, the speed of adjustment of inflation to an own shock is instantaneous.

Price level dynamics are not directly studied by the SW-UC model. Rather than define the observer equation with inflation, expressing it in (the log of) the price level, $\ln P_t$, gives the UC-LL model,

$$(A.4) \quad \begin{aligned} \ln P_t &= \mu_{1,t}, \\ \mu_{1,t+1} &= \mu_{1,t} + \mu_{2,t} + \delta_{t+1}, \quad \delta_{t+1} \sim \mathbf{N}(0, \sigma_\delta^2), \\ \mu_{2,t+1} &= \mu_{2,t} + \psi_{t+1}, \quad \psi_{t+1} \sim \mathbf{N}(0, \sigma_\psi^2), \end{aligned}$$

where $\mu_{1,t}$ denotes the price level trend, δ_t is its forecast innovation, $\mu_{2,t}$ represents price level trend deviations, and ψ_t is its forecast innovation. When δ_{t+1} rises, the impact on $\mu_{1,t+j}$ ($j \geq 1$) and $\ln P_{t+j}$ is permanent because it never decays. The same response is generated by the shock to price level trend deviations, ψ_{t+1} . Details about UC models are found in Harvey (1990) and Gourieroux and Monfort (1997). Harvey suggests that mean square error estimates of the state vector distinguish the SW-UC and UC-LL models.

The reduced-form MA(1) of the UC-LL model is $(1 - \mathbf{L})\pi_t = (1 - \mathbf{L})\delta_t + \psi_t$. Since the reduced form of the UC-LL model is a first-order moving average, the results discussed above about the connection between the SW-UC model and the MA(1) of equation (A.3) are applicable. For the UC-LL model, the mapping is $(1 + \theta^2)\sigma_\eta^2 = 2\sigma_\delta^2 + \sigma_\psi^2$ and $-\theta\sigma_\eta^2 = -\sigma_\delta^2$. The UC-LL also draws out implications for inflation of price level trend shocks, δ_t . This aspect of the UC-LL model ties inflation persistence and volatility, in part, to price level shocks as predicted by the monetary growth model Brock (1974) analyzes.

Harvey (1990) and Gourieroux and Monfort (1997) show how to obtain MLEs of the UC-LL model from the Kalman filter. An issue is that the Kalman filter cannot be initialized using standard approaches because the state space includes unit root processes. Instead, an algorithm Koopman (1997) develops is employed to initialize the nonstationary components of the state vector. These procedures impose a diffuse prior on the initial state vector to compute an exact initialization of the Kalman filter.

The UC-LL model (A.4) and Kalman filter yield estimates of expected inflation. Let $\mathbf{E}_t\pi_{t+1}$ denote expected inflation, where $\mathbf{E}_t\{\cdot\}$ is the mathematical expectations operator conditional on

1. A priori, there is no restriction that $\gamma(\mathbf{1}) \leq 1$ or $0 \leq \gamma(\mathbf{1})$, but (in modulus) $\Lambda(\Gamma) \in [0, 1]$.
 2. The SW-UC model implies that the mean of inflation growth is zero.

Appendix (continued)

date t information. Pass the first difference operator $1 - \mathbf{L}$ through the first line of equation (A.4) followed by the expectations operator, $\mathbf{E}_t\{\cdot\}$, to obtain $\mathbf{E}_t(1 - \mathbf{L}) \ln P_{t+1} \equiv \mathbf{E}_t \pi_{t+1} = \mu_{2,t}$. Thus, expected inflation equals deviations from the price level trend. These deviations are persistent—a random walk, in fact—and have

innovations, ψ_{t+1} , whose impact on $\mathbf{E}_t \pi_{t+1+j}$ is permanent. Given MLEs of the UC-LL model, $\mathbf{E}_t \pi_{t+1}$ is computed using the Kalman filter or smoother. Hamilton (1994) describes these procedures. The initialization of the filtered estimates of $\mathbf{E}_t \pi_{t+1}$ follows Koopman (1997).

REFERENCES

- Andrews, Donald W.K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, no. 4:821–56.
- . 2003. Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica* 71, no. 1:395–97.
- Andrews, Donald W.K., and Hong-Yuan Chen. 1993. Approximately median-unbiased estimation of autoregressive models. *Journal of Business & Economic Statistics* 12, no. 2:187–204.
- Atkeson, Andrew, and Lee E. Ohanian. 2001. Are Phillips curves useful for forecasting inflation? Federal Reserve Bank of Minneapolis *Quarterly Review* 25, no. 1:2–11.
- Brock, William A. 1974. Money and growth: The case of long-run perfect foresight. *International Economic Review* 15, no. 3:750–77.
- Brock, William A., Steven N. Durlauf, and Kenneth D. West. Forthcoming. Model uncertainty and policy evaluation: Some empirics and theory. *Journal of Econometrics*.
- Cogley, Timothy, and Thomas J. Sargent. 2001. Evolving post-World War II U.S. inflation dynamics. *NBER Macroeconomics Annual 2001* 16, no. 1:331–73.
- . 2005. The conquest of U.S. inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics* 8, no. 2:528–63.
- Fisher, Jonas D.M., Chin Te Liu, and Ruilin Zhou. 2002. When can we forecast inflation? Federal Reserve Bank of Chicago *Economic Perspectives* 26, no. 1:30–42.
- Gourieroux, Christian, and Alain Monfort. 1997. *Time series and dynamic models*. New York: Cambridge University Press.
- Greenspan, Alan. 1994. Statement before the Subcommittee on Economic Growth and Credit Formulation of the Committee on Banking, Finance, and Urban Affairs, U.S. House of Representatives, February 22.
- Hamilton, James D. 1994. *Time series analysis*. Princeton, N.J.: Princeton University Press.
- Hansen, Bruce E. 1997. Approximate asymptotic P values for structural-change. *Journal of Business & Economic Statistics* 15, no. 1:60–67.
- Hansen, Lars P., and Thomas J. Sargent. 1980. Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control* 2, no. 1:7–46.
- Hansen, Peter R., Asger Lunde, and James M. Nason. 2005. Model confidence sets for forecasting models. Federal Reserve Bank of Atlanta Working Paper 2005-7, March.
- Harvey, Andrew C. 1990. *Forecasting, structural time series models and the Kalman filter*. New York: Cambridge University Press.
- Koopman, Siem Jan. 1997. Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association* 92, no. 440:1630–38.
- Morley, James C., Charles R. Nelson, and Eric Zivot. 2003. Why are Beveridge-Nelson and unobserved component decompositions of GDP so different? *Review of Economics and Statistics* 85, no. 2:235–43.
- Nelson, Charles R., and G. William Schwert. 1977. Short-term interest rates as predictors of inflation: On testing the hypothesis that the real rate of interest is constant. *American Economic Review* 67, no. 3:478–86.
- Pearce, Douglas K. 1979. Comparing survey and rational measures of expected inflation. *Journal of Money, Credit, and Banking* 11, no. 4:447–56.
- Sargent, Thomas J. 1999. *The conquest of American inflation*. Princeton, N.J.: Princeton University Press.
- Sargent, Thomas J., Noah Williams, and Tao Zha. Forthcoming. Shocks and government beliefs: The rise and fall of American inflation. *American Economic Review*.

Sims, Christopher A., and Tao Zha. 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* 96, no. 1:54–81.

Stock, James H. 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, no. 3:435–59.

Stock, James H., and Mark W. Watson. 1999. Forecasting inflation. *Journal of Monetary Economics* 44, no. 2:293–335.

———. 2005. Has inflation become harder to forecast? Princeton University, Department of Economics, unpublished manuscript.

Watson, Mark W. 1986. Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18, no. 1:49–75.