Extremely Preliminary and Incomplete
Comments Encouraged

Trend Breaks, Long Run Restrictions, and the Contractionary Effects of Technology Shocks*

John Fernald
Federal Reserve Bank of Chicago

This draft: March 29, 2004

Abstract

Do hours worked rise or fall following a technology improvement? Recent empirical work using structural VARs with long-run restrictions reaches divergent answers, depending on whether hours worked enters the VAR in log-levels or growth rates. In contrast, I find that once one allows for (statistically and economically plausible) trend breaks in labor productivity, it is unimportant whether hours enters the VAR in levels or growth rates: Hours worked falls by a statistically significant amount on impact following a technology improvement. That technology improvements are contractionary on impact is consistent with other work that tries to measure technology shocks more directly using augmented growth accounting techniques. More generally, the paper shows via several simulations with actual and artificial data that impulse responses estimated via long-run restrictions are sensitive to low frequency correlations in the data. This suggests the need for caution in interpreting the results of estimated responses with long-run restrictions, since results may be driven by effects other than those you think you are identifying.

I thank Shanthi Ramnath for superb research assistance. I also thank Larry Christiano, Neville Francis, Jon Faust, Valerie Ramey, and Rob Vigfusson for helpful conversations. John Williams helped me think about break tests. I thank Valerie Ramey and Rob Vigfusson for providing computer code. JEL Codes E24, E32, O47, keywords Technology shocks, business cycles, structural change. The views expressed in this paper are my own and do not necessarily reflect the views of anyone else affiliated with the Federal Reserve System.

1. Introduction

Do hours worked rise or fall following a technology improvement? There is a growing professional consensus that initially, hours worked fall. This consensus reflects the results of both structural vector autoregressions as well as careful growth accounting exercises. The two approaches identify technology shocks very differently. The SVARs identify technology shocks via long-run restrictions, assuming that technology shocks alone have a long-run impact on the level of labor productivity. "Augmented" growth accounting exercises take a very different approach, seeking more directly to distinguish technology change from myriad short-run non-technological effects that contaminate the Solow residual (e.g., variations in factor utilization).[1] In addition, the experience of the U.S. economy in the early 2000s—with exceptional productivity performance, respectable output performance, but *declining* employment—is frequently attributed to the impact effect of productivity growth on employment.[2]

Recent work challenges this growing consensus, however, since it turns out that the results of the structural VAR estimates is sensitive to whether hours worked enters the VAR in log-levels or in growth rates. In this paper, I find that once one allows for (statistically and economically plausible) trend breaks in labor productivity competing empirical specifications yield a consistent answer: Hours worked falls by a statistically significant amount on impact following a technology improvement.

Galí (1999) and Francis and Ramey (2003) argue that hours worked are best modeled as difference stationary.[3] They find that in the post-World-War-II period, this specification implies that hours worked falls on impact following a technology improvement. In contrast, Christiano, Eichenbaum, and Vigfusson (2003) argue for a different specification of the VAR in which the log-level of hours worked per capita is modeled as being stationary. In their preferred specification, hours worked rises on impact following a technology improvement.

This literature based on long-run identification builds on earlier work by Blanchard and Quah (1989), who estimate a VAR with output and unemployment (with the restriction that only supply shocks affect output in the

---

[1] For SVARs, see, for example, Blanchard and Quah (1989), Shapiro and Watson (1988), Galí (1999), Galí, López Salido, and Vallés (2002), and Francis and Ramey (2003a, b). For "augmented" growth accounting, see Basu, Fernald, and Kimball (2003).

[2] See, for example, Bernanke (2004).

[3] They also consider a range of other specifications, such as stationary around a quadratic trend.

long run). Blanchard and Quah find that their results are sensitive to how they model the post-1973 productivity slowdown. In their preferred specification, they allow for a post-1973 slowdown in trend growth.

The more recent literature, however, ignores the productivity slowdown. But considerable ink has been spilled analyzing and estimating changes in mean productivity growth. In particular, most people believe that productivity growth slowed down after 1973 or so, and many believe that productivity growth picked up again sometime after 1995. Structural change tests confirm that trend growth appeared to slow around 1973 and accelerated again in the 1990s (point estimate is 1997:Q2)—with the two breaks significant at the level of 10 percent or better.

Once one allows for the reasonable hypothesis of structural change in the labor productivity series, the Christino-Eichenbaum-Vigfusson (CEV) levels specification gives results that are qualitatively similar to the Galí/Francis-Ramey (GFR) difference specification: Technology improvements reduce hours worked on impact. These findings apply in both bivariate and larger VAR systems and are robust to reasonable alternative specifications of the break dates.

The source of the sensitivity in the levels specification appears to be the low frequency correlation between labor productivity growth and the level of hours worked. Figure 1 shows the two series. Although highly variable from quarter-to-quarter or even year-to-year, average labor productivity growth was faster before the early 1970s and after the mid-1990s. Broadly speaking, hours worked (per person 16 or older) shows a similar pattern—high (and falling) before the early 1970s, and somewhat higher again towards the end of the sample. For example, from 1948:1 to 2003:4, the correlation between hours per capita and a post-1973 dummy variable is 0.65. As is well-known from Faust and Leeper (1997), estimates using long-run restrictions in finite data can be highly sensitive to low frequency correlations. Hence, one wants to make sure that results are robust to reasonable specifications of the low frequency properties of the data. By focusing on (high frequency) differences, the GFR specification appears robust to this source of misspecification; by contrast, regardless of any other virtues,[4] the CEV levels specification is not robust to low frequency correlations. More generally, whatever

---

[4] For example, CEV argue that the levels specification does much better at "encompassing", or explaining, the difference results, but not vice versa. I discuss encompassing in Section IV. The bottom line is that their tests may be perfectly correct—in that various test statistics are much more easily matched by the levels specification—yet nevertheless irrelevant for the issues at hand.

one concludes about whether the level of hours is stationary or difference-stationary, the level-of-hours specification appears particularly sensitive to low frequency correlations in the finite sample data.[5]

I demonstrate that the source of the sensitivity here reflects low-frequency phenomena in several ways, focusing for simplicity on the bivariate model. First, when estimated over the (sometimes short) sub-periods that correspond to break dates, the levels and difference specifications *always* suggest that by hours worked fall in response to a technology shock; indeed, the declines are more likely to be statistically significant in the levels specification. Second, for the full sample period, I estimate the structural VAR after replacing actual labor productivity growth with a dummy variable for the productivity slowdown and acceleration—that is, equal to one before 1973:2 and after 1997:1 but equal to zero from 1973:2 through 1997:1. When the structural VAR is estimated with this dummy-variable productivity growth along with the actual log-level of hours per capita, the impulse response of hours to the "identified" technology shocks look like the CEV impulse responses: Hours worked rises by a statistically significant amount on impact. Clearly, this positive impact effect is not identifying the response to the kind of high frequency technology shocks that proponents of real-business cycle models have in mind.

Third, I present Monte Carlo results consistent with the empirical findings of this paper that low frequency phenomena can obscure, or even reverse, the "true" high frequency impact effect of a technology shock on hours worked. I present results from two experiments. In one case, I generate artificial quarterly data with a roughly contemporaneous break in both series; even if there is no relationship (or a negative relationship) between technology shocks and hours worked, the estimated impulse responses in the levels specification show a large but spurious positive response. By contrast, specifying hours in growth rates appears to be much more robust to this source of misspecification, even though—apart from the break—the levels specification is correctly specified.

In the other Monte Carlo experiment [NOT REPORTED IN THIS DRAFT], there is no true break in the data generating process for hours. But I select artificially generated series where, simply by chance, there is an

---

[5] Interesting work by Erceg, Gust, and Guierieri (2003) also points out the sensitivity of long-run identification schemes to small samples and to persistent (but not permanent) non-technological shocks. In particular, they simulate calibrated RBC models to generate artificial data and then estimate an SVAR. They find that the impulse responses at impact tend to pick up the effects of persistent but not-permanent non-technology shocks, such as labor supply shocks; for example, if temporary but persistent labor supply shocks are important, then the estimated hours response to a technology shock can be substantially

apparent downward break in levels around the middle of the sample (consistent with the picture in Figure 1 for hours); in those cases, the impulse responses again look like the CEV impulse responses: Hours appear to rise on impact. This second case makes clear that the key question is not whether the "true" data-generating process has a break, but whether the actual small sample has a low frequency correlation between the two series.

That technology improvements are contractionary on impact—and only lead to an expansion with a lag— is consistent with other work (see Basu, Fernald, and Kimball 2003) that tries to measure technology shocks more directly using augmented growth accounting techniques.

2. Evidence for Trend Breaks

Many economists believe that the data-generating process for labor productivity underwent at least one and possibly two breaks in the post-war period. In particular, many economists believe there was structural change sometime around 1973, when trend labor productivity growth slowed, and another change in the late 1990s, when trend labor productivity growth rose again. As we now discuss, the statistical evidence is consistent with such changes. One reasonable interpretation of such structural change is a regime shift (as emphasized by Kahn and Rich 2003). An alternative interpretation is that history often has unusual influences—steam power, electricity, the interstate highway system, information technology, and so forth—that have a persistent, but perhaps not permanent, effect on the economy's potential growth rate. [6]

Of course, even if the data-generating process for labor productivity were well described as a random walk with constant drift (so that labor productivity growth has a constant mean over time), one might nevertheless by chance have seen a prolonged period with either high or low mean growth. It would, therefore, be inappropriate to simply test the apparent break dates (for example, 1973:1 and, say,1995:4). Formal statistical tests for structural change (e.g., Christiano, 1992; Andrews, 1993; Andrews and Ploberger, 1994; Bai and Perron, 1998) take this into account in constructing critical values for their tests. Following the recommendation of Bai and

---

biased upwards. The trend breaks considered in this paper—whether interpreted as true breaks or simply as convenient ways to summarize the data set at hand—are an example of an extremely persistent shock that the SVAR has trouble getting right.

[6] See the extensive literature on general purpose technologies. Fernald (1999) argues that the Interstate Highway System played such a role in the decades before the early 1970s.

Perron (1998), I begin by testing for a single break; conditional on identifying the first break, I then test the two sub-samples on either side of the break for further evidence of breaks.

I used BLS data on business sector labor productivity from 1947:1 to 2003:4.[7] The top panel of Figure 2 shows the F-statistic testing for a break at each date in the post-war period, excluding the first and last 9 percent of the sample. That is, the F statistic shows the significance of a dummy variable in a regression of labor productivity growth on a constant and a dummy that equals one up to and including the break-test date and equals zero afterwards. The maximum F statistic of 7.61 is reached in 1973:1. This is at approximately the 10 percent critical value from Andrews (2003). Andrews-Ploberger (1994), however, suggest that an exponential F test (which incorporates more information than looking solely at the maximum F over the sample) has better properties. The exponential F statistic for structural change is 1.89, which has a p-value of under 0.06 (using p-values estimated from Hansen's 1997 approximation formula).[8] To check these asymptotic p-values, I followed Christiano (1992) and simulated artificial series by drawing from the actual labor productivity residuals (under the null of constant mean) and estimated the maximum F and exponential F statistics. The bootstrapped p-values are roughly consistent with the asymptotic values.[9] This evidence is thus fairly strongly supportive of the notion that there is at least one trend break—particularly given the relatively weak power of structural change tests.[10]

Conditional on this first break, I test for a second break in the subperiods before and after the first break. In the pre-1973:2 period, there is no evidence of a second break (Hansen's approximate p-value from the exponential F is only 0.5). In the post-1973:1 period, however, the exponential-F test has a p-value of 0.011—strongly indicating a second break. From Figure 2, panel B, the maximum F statistic is reached for a break in 1997:2,

---

[7] Data were downloaded, via Haver Analytics, on February 7, 2004.

[8] P-values were calculated using parameters obtained from the Gauss program PV_EXP.PRC from Bruce Hansen's web site (downloaded from http://www.ssc.wisc.edu/~bhansen/progs/jbes_97.html on Feb 11, 2004).

[9] In my initial runs, the F-statistic has a bootstrapped p-value of 0.06, whereas the exponential F has a bootstrapped p-value of under 0.04. These are even tighter than the asymptotic values. However, I want to double check the code to be certain that I've implemented the simulations properly.

[10] Note that adding additional data makes it harder to identify the first break in 1973:1. For example, using data from 1947:1 to 2002:4 (i.e., dropping the final year of data) would have yielded an approximate p value of 0.03. This appears to reflect the fact that the sequential approach taken here can sometimes work poorly in practice if the value of the mean returns close to its original value after the second break, as noted by Bai and Perron (2003). Bai and Perron (2003) suggest the alternative procedure of first testing whether *at least* one break is present and then searching for them. I have not yet implemented this approach.

although several other periods are reasonably close.[11] In sum, if we focus on just the post-1973:1 period, the data strongly want a trend break, but the actual date remains a bit uncertain. Indeed, the standard dating of 1995:4 for the productivity acceleration is the first reasonable date one could pick. In what follows, I date the break at 1997:2, but none of the results that follow are affected if I instead chose 1995:4.

There are at least two alternative approaches that also suggest that mean growth in labor productivity is not constant over time. One approach models productivity growth with an explicitly stochastic trend. For example, Roberts (2001) estimates a time-varying trend productivity growth rate; he interprets his results as "suggestive that that the changes in trend productivity growth over this sample have been statistically significant." He estimates an annual standard deviation of the shock to trend productivity growth of 0.2 percentage points; based on Monte Carlo simulations, he finds that this standard deviation is significantly different from zero at the 5 percent confidence level.

Roberts' point estimates appear economically as well as statistically significant. For example, he finds that the upper bound of the (90 percent) confidence interval for the early-1980s trend estimate is well below the point estimate for trend productivity growth in the early 1960s. Similarly, the lower bound of the confidence interval at the end of the 1990s is above the point estimate for the early 1990s. These findings appear in line with the evidence from the trend breaks considered here, although the changes in trend are smoother.

Roberts argues that "because the stochastic-trend model can't be rejected relative to the broken-time-trend model, it is, at a minimum, an equally valid way of interpreting events as the broken-linear-trend model." However, as Roberts notes, the linear-trend-with-break approach has the practical advantage that it is simpler to implement. The important message is that two approaches give a qualitatively similar picture of the changes in the underlying trend growth rate: Both approaches strongly suggest that the assumption of a constant mean growth rate provides an inadequate time series model for the U.S. post-war productivity data.

Another alternative model is the regime-switching model of Kahn and Rich (2003). They find evidence in favor of long-term growth regimes, with "distinct switches in the early 1970s and late 1990s." Their empirical estimation draws on the predictions of growth theory, which helps give them power to identify low-frequency

---

[11] Bai and Perron (1998, Table II, page 61) suggest a maximum-F test for l+1 breaks, conditional on l breaks. That

regime shifts. Kahn and Rich find results similar to those reported below: Once they allow for these regime shifts, the response of hours worked to a technology shock is negative.

Finally, note that hours worked per capita also shows evidence of at least one change in mean. Looking at Figure 1, one's eye tends to focus on the cyclical fluctuations in the series. But the maximum F statistic for a break in level is 298 in 1958:Q1. Following the Bai-Perron sequential approach, the data then want a second break in 1970:2 (F statistic is above 55).[12] In the post 1970 period, the data want a further break in 1994:2 (F statistic of 105).[13]

Table 1 shows regression results from regressing labor productivity growth and hours worked per capita on pre-1973:2 and pre-1997:2 dummies (the latter is chosen because that's the maximum F statistic for a late 1990s break in labor productivity).[14] Both labor productivity growth and the level of hours were higher pre-1973; both series were also higher after 1997 than during the 1973-97 period. There is thus a low-frequency correlation between hours worked and labor productivity.

3. What Do Structural VARs Say?

3.1 Structural VARs

I estimate a range of structural VARs following Galí (1999) who, in turn, builds on Blanchard and Quah (1989) and Shapiro and Watson (1988). The key identification assumption is that only technology shocks affect the level of labor productivity in the long run. Other shocks (such as labor supply shocks or monetary shocks) may affect labor productivity in the short run—as they will, for example, in most dynamic general equilibrium

---

may suggest weaker evidence for the second break than the exponential F.

[12] CEV (2003) argue that in the sample period 1959:Q1 (check), standard unit root tests reject the null of a unit root in favor of trend stationarity. Over this shorter sample period, the maximum F statistic for a break in mean is 51.6. Adding a time trend substantially increases the F test for structural change in mean hours per capita; from 1959:1 on, for example, the maximum F on the change in mean rises to 150 with a statistically significant time trend.

[13] That said, with a highly persistent AR(2) (which is what the series looks like, with estimated AR(1) and AR(2) coefficients from an OLS regression of (about) 1.5 and -0.53, simulations confirm that it is fairly common to find large apparent shifts in mean simply by chance. But this does not necessarily matter for our purposes. That is, suppose the break appears just by statistical chance in the (small) sample, and roughly coincides with the break in trend labor productivity growth. Simulations seem to confirm the intuition that those are exactly the cases where the impulse response tends to look strongly positive, even if the "true" response is flat. So it is perhaps less relevant whether or not there is "true" structural change. I am continuing to think about this point.

[14] Labor productivity, hours worked, and the civilian population aged 16 and older are from the Bureau of Labor Statistics (downloaded via Haver Feb 7, 2004), and refer to the private business sector. In the regressions, hours worked are divided by civilian population 16 and older.

models that have variable factor utilization—but these shocks do not have a permanent effect on labor productivity.

In most cases, I consider a bivariate specification with labor productivity growth as the first variable and either the (log) level or growth rate of hours per capita as the second variable. The exposition focuses on the bivariate log-level case.

Suppose p_t is the log of labor productivity, x_t is the log of hours worked (in larger systems, x_t is a vector of variables other than labor productivity that we include in the VAR). As Shapiro and Watson (1988), show, one can estimate the technology residuals as the residuals from the following regression:

$$\Delta p_t = c + A(L)\Delta p_{t-1} + B(L)\Delta x_t + \varepsilon_t^Z \tag{1}$$

Note that x_t enters the regression in first differences, which imposes the restriction that non-technological shocks do not affect the level of labor productivity in the long run. Since technology shocks might affect the current growth rate of hours worked (or other variables included in x), we estimate this regression with instrumental variables (a constant and Δp_{t-s}, and the levels of x_{t-s}, where s=1 to 4).

Once we have estimated this equation and identified the technology shocks ε_t^Z, we can use OLS to estimate the equation for x_t as:

$$x_t = c + C_1(L)x_{t-1} + C_2(L)\Delta p_{t-1} + \beta\varepsilon_t^Z + \varepsilon_t^D \tag{2}$$

With two variables, we can identify two shocks, one of which (ε_t^Z) represents permanent shocks to technology. The other shock captures all shocks (notably demand shocks) with at most a transitory effect on labor productivity.

In the Galí/Francis-Ramey difference specification, we can re-interpret x_t as the <u>growth</u> rate of hours worked, so that Δx_t represents the second difference of the log-level of hours. The previous discussion is otherwise unchanged. The extension to the case in which x_t represents a vector (i.e., in a larger VAR system) is straightforward.

I generate impulse responses by simulating the dynamic response of x_t to a technology shock ε_t^Z. For each lag in this response function, I follow Christiano, Eichenbaum, and Vigfusson (2003) by generating centered 95 percent Bayesian confidence intervals.[15]

3.2 Quarterly bivariate VAR results with long-run identification restrictions

Faust and Leeper (1997) suggest that in finite data, long-run restrictions can transmit low frequency correlations to higher frequency impulse responses. In part because of this concern, Blanchard and Quah (1989) allow for a post-1973 slowdown in the rate of output growth. However, Galí, Francis and Ramey, and CEV do not make any allowance for potential shifts in mean labor productivity growth. This opens up the possibility that their specifications could inappropriately attribute low-frequency correlations of labor productivity growth and the level of hours to the (short-run) impact of technology on hours. At a minimum, one would not want to rely on results on impact effects of technology shocks that are sensitive to whether one controls for (known) low-frequency correlations in the data.

Figure 3 shows the VAR responses from bivariate VARs with labor productivity and hours per capita. In all cases, labor productivity is entered in growth rates. In the left column of figures, the log of hours worked per capita (for those aged 16 or above) is entered in levels, consistent with the CEV specification. In the right column, hours worked per capita enters in differences, consistent with the GFR specification.

Panel A reproduces the divergent results from the levels and difference specification. In the levels specification, an identified technology shock raises hours worked on impact. By contrast, in the difference specification, an identified technology shock reduces hours worked on impact.

Panel B modifies these two VARs by removing a post-1973:Q1 change in mean productivity growth. For reporting these results, I regressed labor productivity growth on a constant and a pre-1973:Q1 dummy variable, and took the residuals as my measure of labor productivity growth. The levels specification is fairly sensitive to the trend break in labor productivity: An identified technology shock now has essentially zero contemporaneous impact on hours worked rather than a sizeable positive impact. By contrast, results in the second column for the

---

[15] I thank Valerie Ramey and Rob Vigfusson for sending me code that replicated their results and confidence intervals.

difference specification change little from those in panel A—suggesting that the growth-rate specification is less sensitive to removing low frequency components from the labor productivity series.

Finally, panel C shows results allowing for structural change in mean productivity growth in both 1973:2 and 1997:2. In the levels specification (left column), the identified technology shocks now have a statistically significantly negative impact on hours worked. Indeed, the magnitude of that negative effect is even larger than in the difference specification (right column). Qualitatively, though, the responses in the two charts in Panel C are very similar—and both suggest that technology improvements significantly reduce hours worked on impact.

Several things to note. First, I could have put the trend breaks into the VAR itself, which would have the effect of allowing the breaks to affect the hours equation as well. That gives very similar results to those shown here, where the breaks are removed before estimation. The advantage of taking out the breaks before estimation is that it makes clear that allowing for structural change affects the impulse responses by affecting the properties of labor productivity rather than hours. Given that the existing literature has focused (without resolution) on how best to model the hours process, the sensitivity to detrending labor productivity growth (and the resulting robustness to how one models hours) is worth recognizing.

Second, there is another dimension by which the levels specification with trend breaks looks better than the specification without breaks: the responses to other shocks. In particular, Figure 4 shows the response of labor productivity to a positive demand shock. (This demand shock is the shock to the hours variable, and is identified as the shock in the bivariate system that isn't technology.) Without a trend break (panel A), the levels-specification in the left column suggests that labor productivity falls sharply. Although the near-term effect could be interpreted as reflecting the movement down a stable labor demand curve, the sign of the effect is inconsistent with considerable evidence of the importance of utilization changes in response to demand shocks.

Even more troubling is that the impulse response shows a persistent decline in labor productivity following a demand (i.e., anything other than permanent technology) shock. (Francis and Ramey (2003a, b) also discuss this persistent decline.) Given that identification comes from the assumption that demand shocks have no permanent effect on labor productivity, this impulse response suggests misspecification. (Although not shown, I ran the

All impulse responses reported in this paper use the Vigfusson code to calculate Bayesian confidence intervals.

impulse response out to 30 years, and even then the impulse response was notably below zero, suggesting the identification is being imposed at something close to infinity.) To defend the bivariate levels/no-trend-break specification, one must believe that positive demand shocks lead to extremely persistent—almost permanent—declines in labor productivity.[16] Economically, this seems implausible; statistically, the results of Faust and Leeper suggest that in this scheme, the zero-long-run impact may be too weak a restriction to achieve reliable identification. (Note: I'm not sure what these responses looks like in the larger VAR, in case somehow the misspecification reflects the small size of the system). By contrast, once one allows for trend breaks in the bottom panel, an identified demand shock now has the expected positive, but temporary, effect on labor productivity.

Once trend breaks are removed, the levels specification again matches the qualitative features found in the difference specification of a positive, but temporary effect. Once again, Figure 4 suggests that the difference specification is not sensitive to removing low frequency components from the labor productivity series.

In sum, results are qualitatively very similar between the two hours specifications, once one allows for plausible trend breaks in the growth rate of labor productivity. The levels specification is particularly sensitive to this low frequency specification issue in a way that the difference specification is not.

3.3 Estimates by Subsamples

Suppose we estimate the SVAR over subsamples that correspond approximately to the break dates. Figure 5 shows that the bivariate VAR in both differences and levels suggests that in all subsamples, technology shocks reduce hours worked, regardless of whether hours is in growth rates or levels. Indeed, the levels specification is more likely to show a statistically significant decline in hours on impact. Not surprisingly, the short final sample in the bottom panel is poorly estimated; nevertheless, despite wide confidence intervals, the impact effect is statistically negative with a point estimate that is much larger than in any of the preceding periods.

Note that Galí, López Salido, and Vallés (2002) also find results that are sensitive to sample period. They look at subperiods based on presumed differences in the monetary policy reaction function rather than based on presumed changes in trend productivity growth. Nevertheless, it is interesting that in the 1973-1997 period—a

---

[16] Bar-Levy (2003) presents a model and evidence that demand shocks might raise technology permanently by inducing R&D. Sarte (1997) also suggests that the identifying assumption of no permanent effect of non-technological shocks on the level of labor productivity might fail. However, the model estimated here *imposes* the restriction, regardless of whether it is true.

period which incorporates the "bad" monetary policy of the 1970s but also the presumably better monetary policy of the Volcker-Greenspan era—the evidence is particularly strong in the levels VAR that hours fall on impact when technology improves.

3.4 Estimates in Larger Systems

Christiano, Eichenbaum, and Vigfusson (2003) also report results from larger 4- and 6-variable systems with inflation, the federal funds rate, and the ratios of consumption and investment to output. Results from these larger systems generally look qualitatively and quantitatively similar to the bivariate results: In the levels specification, technology improvements always reduce hours worked by a statistically significant amount on impact once one allows a post 73:1 and post-97:2 dummy.

Figures 6 and 7 report results from their 4-variable system, i.e., where we add the consumption-output and investment-output ratios to the VAR, again with the restriction that only shocks to technology affect labor productivity in the long run.[17] Those results look qualitatively and quantitatively similar to the bivariate results: Once one allows for breaks in the productivity series, technology improvements reduce hours worked on impact. Indeed, the evidence for this is even stronger when hours per capita enters in levels rather than in growth rates—not only is the point estimate on the impact effect more negative, but the confidence intervals are tight enough that one can reject zero.

3.5 Evidence Using a Direct Measure of Technology Shocks

Christiano, Eichenbaum, and Vigfusson (CEV 2004) report results from using the Basu, Fernald, Kimball (BFK, 1999) measure of the cleansed Solow residual. They interpret the (annual) BFK residual as a direct but noisy measure of technology. They seek to identify "true" technology shocks under the assumption that only true technology shocks affect the BFK measure of technology in the long run. In bivariate systems, they report the

[17] CEV (2003, page 20) argue that it is important to include at least consumption and investment in the VAR. They find (and we confirm) that the impulse responses from the 4-variable VAR are similar to those arising in their larger 6-variable system. The 6-variable levels VAR with trend breaks also suggests that hours fall after a technology improvement. The corresponding difference specification, however, finds a small positive impact effect once one allows a break—but with an extremely large confidence interval. Given CEV's argument that the difference VAR is misspecified, this result is not necessarily interesting. (The argument in this paper is not that the difference VAR is necessarily 'correct', but that it is generally robust to low-frequency correlation between hours and labor productivity growth.)

same anomaly noted above: When hours is in levels, technology shocks raise hours worked on impact, but when hours is in growth rates, technology shocks reduce hours worked on impact.

The resolution turns out to be the same as discussed above: The divergence is completely driven by the post-1973 slowdown in trend technology growth. When one uses residuals from regressing the BFK technology series on a post-1973 dummy to control for the slowdown in technology growth, the bivariate system again suggests that technology shocks reduce hours worked by a statistically significant amount, regardless of whether hours worked enters in levels or growth rates.

Another way to see this point is to note that CEV (2004) argue that the BFK residuals are a noisy measure of technology on the grounds that hours worked Granger-causes the BFK residuals. They interpret this Granger causality as reflecting measurement error in technology: "The sort of measurement errors we have in mind are the transient, high-frequency discrepancies between true and measured outputs and inputs that occur as a result of the way the economy adjusts to shocks."

But the Granger-causality evidence is not consistent with the CEV interpretation that it reflects high frequency measurement error. Instead, the Granger-causality results appear completely driven by the fact that the log-level of hours worked Granger-causes the 1973 productivity slowdown! In particular, suppose one defines a slowdown variable as a dummy variable equal to one before 1974 and zero afterwards. Running a Granger causality test from 1950-1989 (the BFK 1999 sample period) of the slowdown variable on the "lag" of itself as well as the lagged level of hours worked, the p-value of the coefficient on the lagged level of hours is 0.08. It is thus not surprising that once the slowdown is removed from the BFK residuals, there is no longer evidence of Granger causality from hours to technology.[18]

4. Explaining the Sensitivity to Trend Breaks

In order to understand the sensitivity of the levels VAR to low-frequency correlations, I now discuss several exercises. I return to the bivariate case for simplicity and clarity, given that the larger systems give similar results.

4.1 Putting a Productivity-Slowdown/Acceleration Dummy Variable Into the SVAR

I have argued that the levels VAR appears highly sensitive to low frequency phenomenon. A clear demonstration of this sensitivity comes from estimating the structural VAR after replacing actual labor productivity growth with a dummy variable for the productivity slowdown and acceleration—that is, equal to one before 1973:2 and after 1997:1 but equal to zero from 1973:2 through 1997:1.

Figure 8 shows that the impulse responses to the VAR estimated with this dummy-variable productivity growth along with the actual hours series. In the levels VAR, the apparently "identified" technology shocks have a statistically significant positive impact effect on hours worked! Clearly, this positive impact effect is not identifying the response to the kind of high frequency technology shocks that proponents of real-business cycle models have in mind, given that there are, in fact, no high frequency movements in the labor productivity series, only low-frequency shifts (essentially, low-frequency regime shifts).

4.2 Monte Carlo Evidence

4.2.1. AN ACTUAL BREAK

It turns out to be relatively easy to replicate the empirical findings of this paper that low frequency phenomena can obscure, or even reverse, the "true" high frequency impact effect of a technology shock on hours worked. I begin simulating quarterly bivariate data with length equal to the actual data (1948-2003) with (relatively) correlated structural change in the levels of the series in the two variables in the VAR. The results confirm that the low frequency (non-causal) correlation carries over to the apparent impact effect. The advantage of a Monte Carlo, as opposed to focusing solely on encompassing tests, is that it is much easier to verify what aspects of the data generate the results one finds. (That said, I plan to add encompassing tests later; but I've only just gotten started on those at this point.)

I generate data with the following processes:

$$\text{Labor productivity:} \quad \Delta p = s^P D73 + v_t^P , \tag{3}$$

$$\text{Hours:} \quad x_t = s^X D70 + b_0 v_t^P + b_1 v_{t-1}^P + u_t^X , \tag{4}$$

---

[18] Basu, Fernald, and Kimball (2004) find strong statistical evidence of an apparent productivity slowdown in non-manufacturing technology.

$$\text{Shock to Hours: } u_t^X = \rho_1 u_{t-1}^X + \rho_2 u_{t-2}^X + \nu_t^X \,, \tag{5}$$

ν_t^P and ν_t^X are uncorrelated white-noise processes with variances $\sigma_J^2 = 1$, $J = P, X$.. ν_t^P is the true technology shock here, in that it has a permanent effect on the level of labor productivity.[19] I allow the shock to hours to be an AR(2) process to better match the empirical evidence (this is relevant in the next section).[20] The parameters b_0 and b_1 capture any potential feedback between true technology shocks and employment and, hence, govern the "true" impulse response to a technology shock.

D73 is a dummy equal to one prior to 1974 and zero after; D70 is a dummy equal to one prior to 1970 and zero after. One apparent break in the labor data appears to be around 1970. The actual dating turns out not to be overly important for these purposes; as long as the two breaks are any time close to one another, the results shown below hold. By putting the break at slightly different times, it makes it clearer that it not simply an artifact of creating breaks at exactly the same date. The parameters s^X and s^P capture the size of the break. When I allow a break, I set the size for labor productivity to half the standard deviation of the true technology shocks, consistent with the fact that the actual break is estimated to be about 1.7 percent whereas the actual standard deviation of labor productivity is about 3.5 percent. I set the size of the hours break to 3 times the standard deviation of hours shocks, somewhere between the unconditional size of the break (about 10 times) and the size if one estimates it along with an AR(2) (closer to 1 time—check). By construction, shocks other than the levels-shift D70 have only a temporary effect on the level of hours worked. If D70 equals zero, the DGP for hours matches the CEV model exactly. I.e., there is no unit root in hours worked.

Figure 9 reports results from estimating the levels and difference VARs for three cases, with the following parameterizations:

---

[19] In actual data, labor productivity appears well modeled as a random walk, in that labor productivity growth is pretty close to white noise.

[20] Except that I didn't actually do it properly yet. This is an AR(2) for the disturbance to hours, not an AR(2) for the log of hours. For this set of results, it may not matter. For simulations with an apparent but not actual break, though, this specification makes it harder to find a spurious break since you don't get the slow swings that are visible in Figure 1.

	b_0	B_1	s_P	s_X
(1) Feedback but no breaks	-0.5	+0.5	0	0
(2) Breaks but No Feedback	0	0	0.5	3
(3) Breaks AND Feedback	-0.5	+0.5	0.5	3

First, in the top panel of Figure 9, the data are generated with feedback but no breaks. In this case, both the level and difference VARs do a reasonable job of capturing the qualitative and quantitative effect of a technology shock: Hours fall on impact and rise with one lag, with estimated magnitudes that are very close to the actual b_0 and b_1. The simulations from the hours VAR are a bit more tightly constrained around the true/median response, which is consistent with the fact that the hours VAR is properly specified here and the difference VAR is not.

Second, the middle panel reports the case of breaks but no true feedback. This simulation makes the point most starkly that low frequency correlations (the fact that the artificial series for both hours growth and for labor productivity growth are higher in the first half than the second half of the sample) show up in the impulse responses. Since there is no feedback from technology shocks to hours, the true impulse responses are zero, but the levels VAR suggests a strong positive impact effect in every single simulation. By contrast, the difference VAR gets the impulse responses about right. (Note that the difference simulation bands seem wide but they really aren't—note the scale.)

Third, the bottom panel shows breaks with feedback. This panel most closely captures what I am arguing is the case in the actual data.[21] In particular, the levels specification implies an apparent positive and highly persistent impact effect; the difference specification, in contrast, yields the truth that the impact effect is negative with the (once) lagged effect being negative.

[21] I still need to make sure the parameterization matches the data reasonably well, which it does other than for the AR(2) process for hours.

4.2.2. APPARENT BUT NOT REAL STRUCTURAL CHANGE--INCOMPLETE

Much of the debate over the levels versus growth rate specification has concerned what is the “true” data generating process for hours—does it have a unit root or not? Does it have a quadratic trend? One might, thus, be tempted to argue that the simulations in the preceding section are uninteresting because there is no true break in the level of hours per capita (or, perhaps, even in labor productivity growth).

But a reasonable interpretation of the results from the preceding section is that this may not matter: In the data, the sample means of both labor productivity growth and hours are, in fact, higher before the early 1970s than after; and the sample means both recover somewhat after the mid-1990s. Even if this pattern occurred just by chance—with a stable, stationary data-generating process that does not undergo structural change or regime shifts—in the small sample we have (roughly 55 years of actual data), the low frequency correlations that are evident in the data would still drive the results.

To confirm this, I have experimented with generating data using the model from the preceding section. Now, we focus on the case in which there is no true break in hours; instead, labor productivity is a very persistent AR(2) process (as the actual data appear to be) with $\rho_1 = 1.5$ and $\rho_2 = -0.53$. For simplicity, I continue to assume there is a break in labor productivity but no feedback from technology shocks to hours. For each simulation, I run a simple Chow test to see if there is an apparent break in hours at 1973:1. As expected, roughly 2-1/2 percent of the time, there is an apparent levels-decline in the hours series with a t-statistic of 2 or greater.

Results TO FOLLOW plot the mean impulse response from these simulations in which there is an apparent break (plus/minus 1.64 standard deviations in the actual responses.) NEEDS TO BE COMPLETED.

4.3 Explaining Encompassing Tests [INCOMPLETE THOUGHTS, NO RESULTS]

Much of the focus of CEV (2003) is the notion that carefully constructed “encompassing tests” can help distinguish whether the levels or difference specification gives more reliable results. The idea is intuitive that the properly specified VAR should predict the results from the mis-specified VAR, but not vice versa. For the purposes of this paper, of course, the CEV encompassing tests as specified are not the interesting ones—after all, both the level and the growth rate VAR yield the same result, that hours worked fall following a technology improvement. (Indeed, if the encompassing tests prefer the levels specification, that's fine, since the estimates are, in fact, distributed much more tightly.)

Nevertheless, the results in this paper suggest that one can make sense of the encompassing tests CEV do report. In particular, if the apparent positive response in the levels VAR is being driven by the low frequency correlations, it is not surprising that the difference VAR fails to encompass the levels VAR, since the difference VAR has removed the low-frequency effects before simulation.

The much more interesting question, though, is whether one can use encompassing tests to see whether the data "prefer" the levels VAR with the trend breaks to the levels VAR without the trend breaks. That is, one could conjecture that, under the assumption that there is no true trend break, the levels VAR would not generally find that incorrectly imposing such a trend break reverses the results. In contrast, if there is a true trend break, then the Monte Carlo results suggest that it should be easy to find such a reversal.

In initial simulations, not shown, I simulated artificial datasets from the levels VAR without breaks. I then followed my break procedure by taking the residuals from regressing the simulated labor productivity on a pre-73:1 and post-97:2 dummy. The estimated impulse responses look similar to the actual levels VAR responses, confirming the intuition that simply taking breaks out that aren't there won't change results much. Of course, that doesn't allow properly for the fact that, whether by chance or in fact, these breaks actually appear in the data. So I should perhaps focus only on simulations where there is, in fact, an apparent break.

5. Conjectural Causes and Implications of the Sensitivity

Why are results sensitive to low frequency phenomenon? Statistically, the basic finding is that the estimated impact effect of shocks can be affected (even reversed) by low-frequency changes in the productivity trend; the simulations show that this reversal can reflect the fact that the levels specification is sensitive to low-frequency correlation, even if this correlation isn't causal.

Are there economic stories that could give rise to the same phenomenon? That is, suppose that the low-frequency correlation reflects shifts in trend growth and the corresponding response of hours. CEV (2003, page 1) argue that "…we find that a permanent shock to technology has qualitative consequences that a student of real business cycles would anticipate." But of course, RBC models react very differently to permanent shocks to the *level* of technology (the usual RBC specification) versus persistent *growth rate* shocks. Campbell (1994), Pakko (2001), and Edge, Laubach, and Williams (2004) document that in frictionless DGE models, an unanticipated increase in the technology growth trend reduces hours worked and investment on impact. Intuitively, there is a wealth effect that, on impact, reduces labor supply; and the equilibrium interest rate also rises, reducing the desired capital-output ratio (which suppresses investment).[22]

Is it possible that the CEV specification is, in fact, picking up that the economy reacts very differently to one-time shocks to the productivity level versus persistent growth rate shocks? When one removes the low frequency ‘growth shocks’, the economy responds to one-time technology shocks with a reduction in hours worked. By contrast, it could be that a persistent reduction in technology growth (as in the 1970s) *also* leads to a persistent reduction in hours worked on impact. One interpretation potentially consistent with the results in this paper is that productivity-growth-rate improvements are, indeed, expansionary (which is why taking them out of the data removes the expansionary effect).

But as already noted, RBC models with shocks to both the level and growth rate of productivity tend to predict that levels-shocks *raise* hours worked while growth-rate shocks *reduce* hours worked, for a while.[23] In the difference specification, however, we have removed this low frequency information before estimation. If this interpretation were correct, then the results from structural VARs with long-run restrictions consistently give rise to effects that are *opposite* the usual RBC results—for growth shocks as well as levels shocks. Under this interpretation of the results in the paper, the data are thus inconsistent with either prediction of the RBC model.

Nevertheless, the results from this paper so far do not support that interpretation. The Monte Carlo simulations make clear that there need not be a causal link for the impulse responses to show an apparent strong but completely spurious positive impact effect. Further work might suggest, however, that one-time non-causal (but correlated) breaks are not the only mechanism that would lead to the empirical results we observe.

---

[22] Edge, Laubach, and Williams (2004) argue, however, that it takes time to learn about the improvement in trend productivity; this restores something close the standard RBC responses, since agents think that most of the shock is a one-time shock to the productivity level.

[23] To avoid having completely offsetting income and substitution effects, one presumably would want to model the growth shocks as persistent but not permanent. E.g., if one looked at data by decade, one might find the RBC model does pretty well, under the assumption that some decades are better than others (and people know it).

6. Conclusions

The existing literature using structural VARs with long-run identifying restrictions find that estimates of the impulse responses of hours worked to a technology shock are often sensitive to how one models hours per capita. As a result, this literature has debated extensively whether hours should be modeled as stationary, a unit root process, having a quadratic trend, and so forth.

Yet largely unnoticed in this debate is that most economists believe that productivity growth slowed down after the early 1970s and accelerated again after the mid-1980s, views consistent with the statistical evidence. Allowing for these changes in mean labor productivity growth resolves the apparently divergent estimates of the impact effect of a technology shock on hours worked. Whether hours enter the VAR in levels or differences, the results support the conclusion from augmented growth accounting exercises that over the post-war period, technology improvements reduce hours worked on impact.

In particular, this paper finds that the apparent positive effect of technology shocks on hours worked in Christiano, Eichenbaum, and Vigfusson (2003) is driven by the post-1973 productivity slowdown and the late-1990s productivity acceleration. Their level specification is very sensitive to low frequency correlations in the data which may have no causal element. Indeed, suppose one runs their bivariate VAR with actual hours worked but, instead of actual labor productivity, one uses a dummy variable equal to 1 before 1973:1 and after 1997:2 (the two dates with the maximum F statistics for a break in labor productivity). The estimated impulse response looks much like the one they report of a statistically significant positive response.

More generally, the simulations in this paper provide a clear example of the results of Faust and Leeper (1997) that one needs to be careful in interpreting the results of long-run restrictions: If you aren't careful, they can give highly misleading results.[24]. Results can easily be driven by effects other than those you think you are identifying. Nevertheless, the robustness of results to whether hours is modeled as stationary or difference stationary; and the consistency of results identified with long run restrictions with those from augmented growth accounting (as in Basu, Fernald, and Kimball 1999) suggest that in the post-war period, technology shocks reduce hours worked on impact.

---

[24] Erceg et al (2003) make a similar point, albeit in the context of an explicit DGE model.

References

Andrews, Donald (1993). *Econometrica*.

Andrews, Donald and W. Ploberger (1994). "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative." *Econometrica*, 62: 1383-1414.

Bernanke, Ben S. (2004). "Monetary Policy and the Economic Outlook: 2004." Remarks at the Meetings of the American Economic Association, San Diego, California, January 4, 2004. Available at http://www.federalreserve.gov/boarddocs/speeches/2004/20040104/default.htm (downloaded March 10, 2004).

Christiano, Lawrence (1992).

Christiano, Lawrence, Martin Eichenbaum, and Robert Vigfusson (2003). "What Happens After a Technology Shock?" NBER Working Paper No. 9819.

Christiano, Lawrence, Martin Eichenbaum, and Robert Vigfusson (2004). "The Response of Hours to a Technology Shock: Evidence Based on a Direct Measure of Technology." NBER Working Paper No. 10254.

Edge, Laubach, and Williams (2004).

Erceg, Christopher, Christopher Gust, and Luca Guierieri (2003). "

Faust, Jon and Eric Leeper (1997). "When Do Long Run Restrictions Give Reliable Results?" *Journal of Business and Economic Statistics*, 15:3, pages 345-53.

Francis, Neville and Valerie Ramey (2003a). "Is the technology-driven real business cycle model dead? Shocks and aggregate fluctuations revisited." Manuscript.

Francis, Neville and Valerie Ramey (2003b). "The Source of Historical Economic Fluctuations: An Analysis Using Long-Run Restrictions." Manuscript.

Galí, Jordi (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review.*

Galí, Jordi, D. López Salido and Javier Vallés (2002) "Technology Shocks and Monetary Policy: Assessing the Fed's Performance." *Journal of Monetary Economics*, vol 50, 723-743.

Hansen, Bruce (1997). "Approximate Asymptotic *P* values for Structural-Change Tests." *Journal of Business and Economic Statistics*, 15:1 (January), pp60-67.

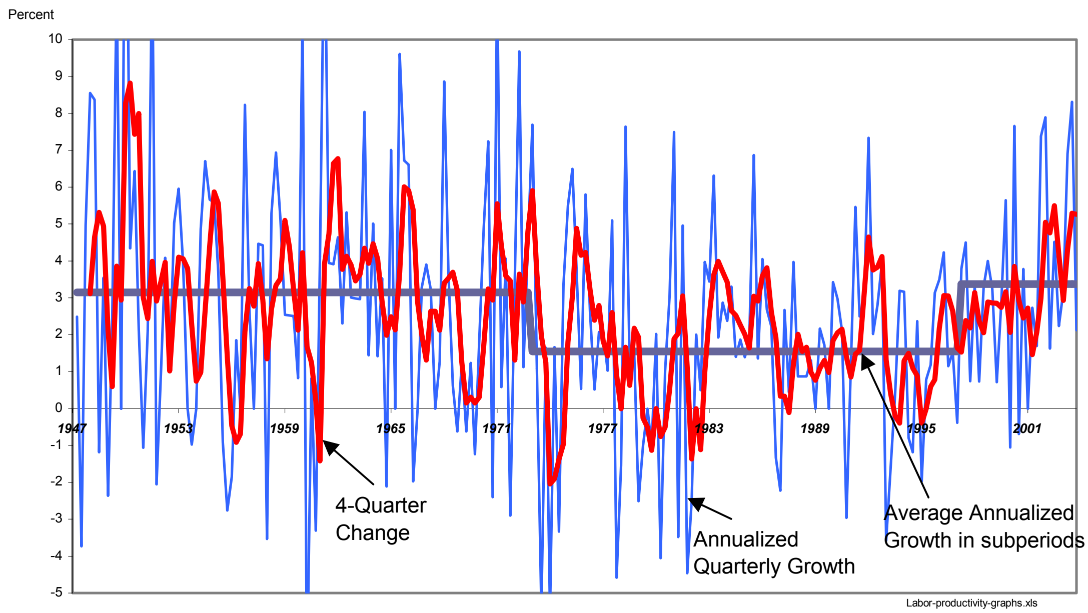
Kahn, James and Frank Rich (2003).

Roberts, John M. (2001). "Estimates of the Productivity Trend Using Time-Varying Parameter Techniques." *Contributions to Macroeconomics*, Volume 1, Issue 1 2001 Article 3.

Shapiro, Matthew and Mark Watson (1988). "Sources of Business Cycle Fluctuations." In Stanley Fischer (ed), *NBER Macroeconomics Annual 1988*, pages 111-148.

Figure 1

A. Labor Productivity Growth



B. Log-Level of Hours Worked Per Capita

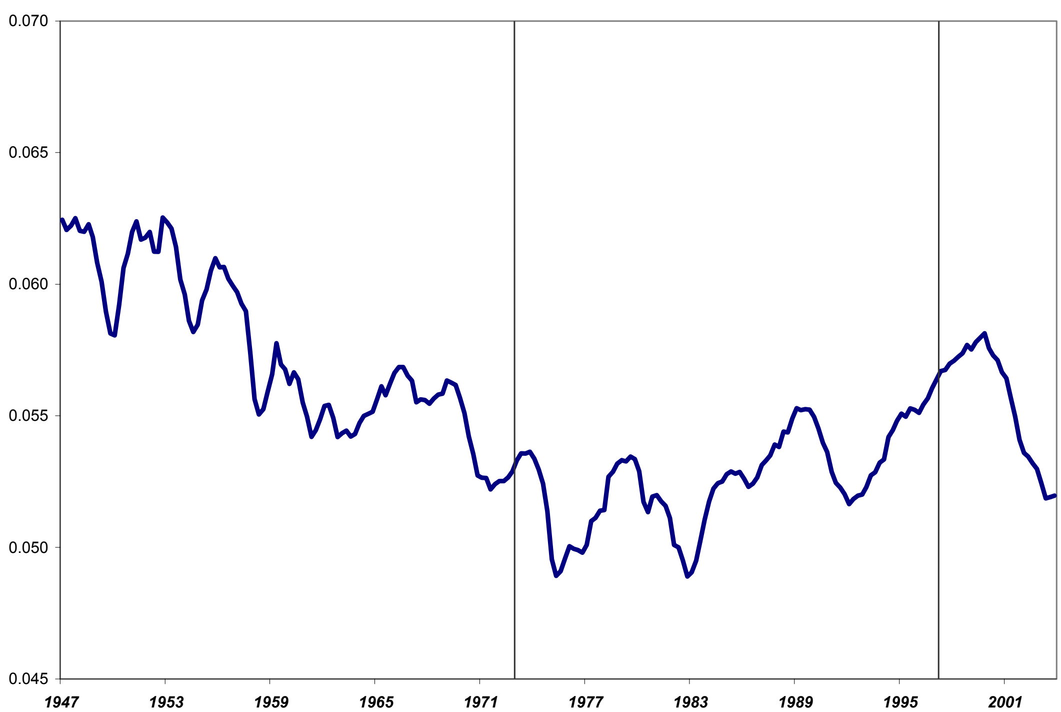
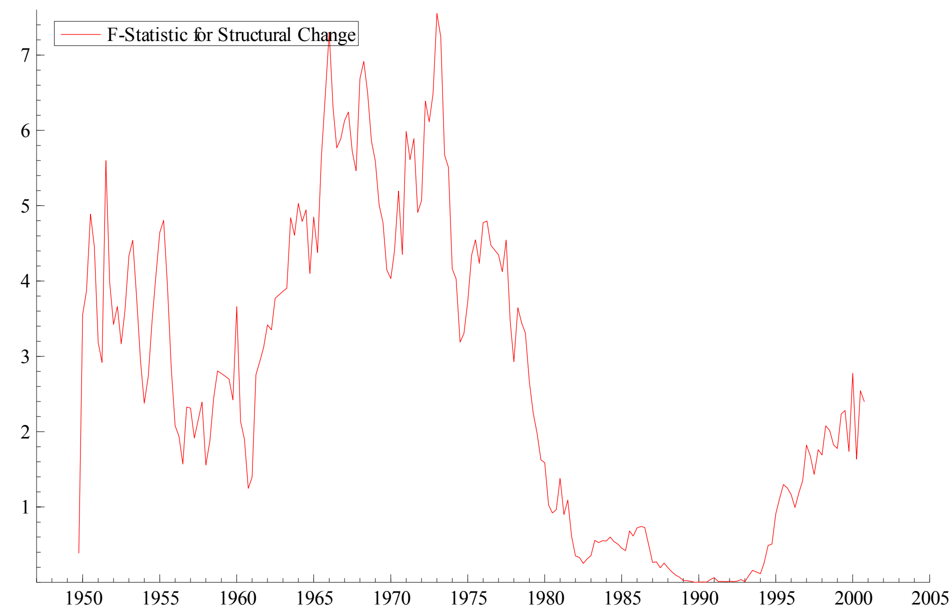
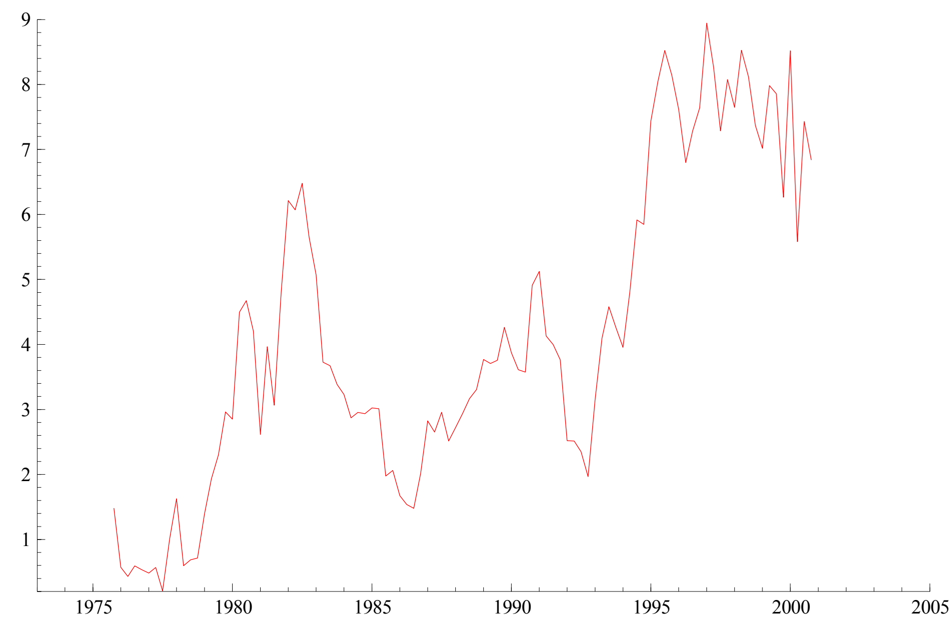


Figure 2
F-Statistic for Structural Change in Labor Productivity (Business Sector)

Panel A—1947:1 – 2003:4



Panel B—Post-1973:2-2003:4 period



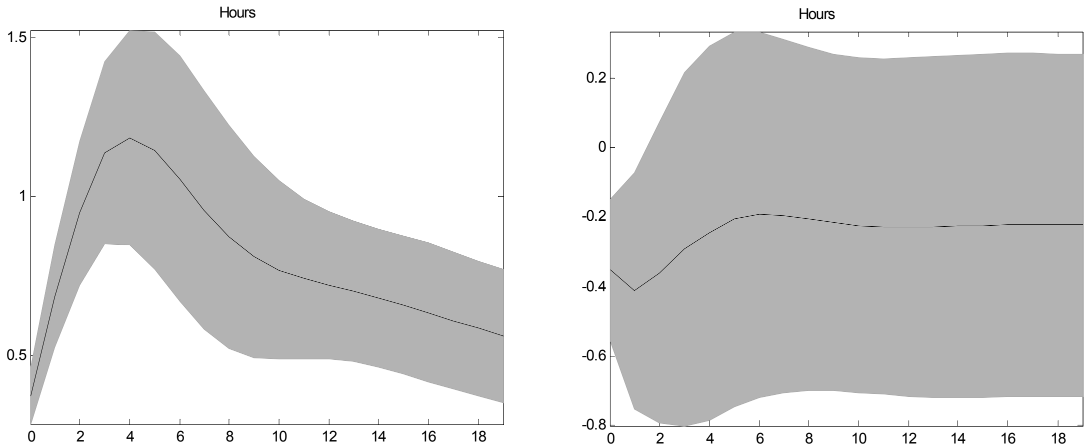
Notes: Figures show F tests for structural change in the constant term of labor productivity growth at each date. We exclude break dates in the first and last 10 percent of each sample. The Andrews-Ploberger (1994) exponential F test is a function of these F statistics.

Figure 3

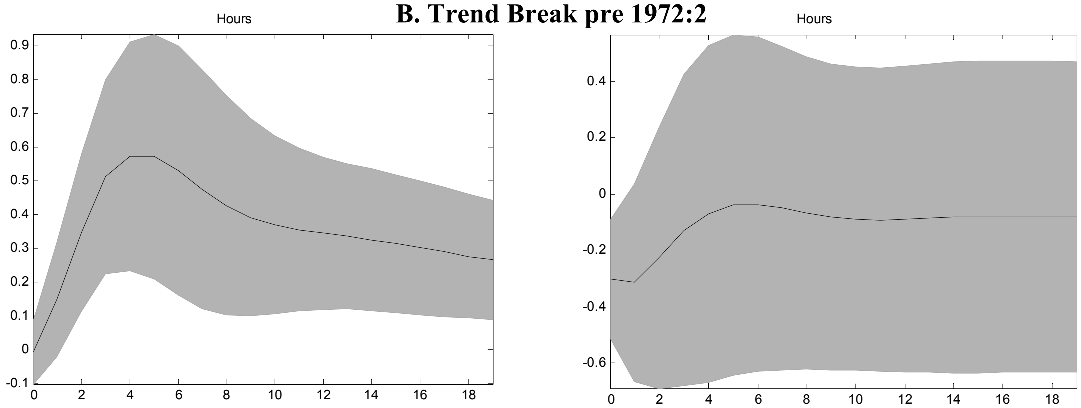
Impulse Responses of Hours Worked to a Technology Improvement

Hours in Levels **Hours in First Differences**

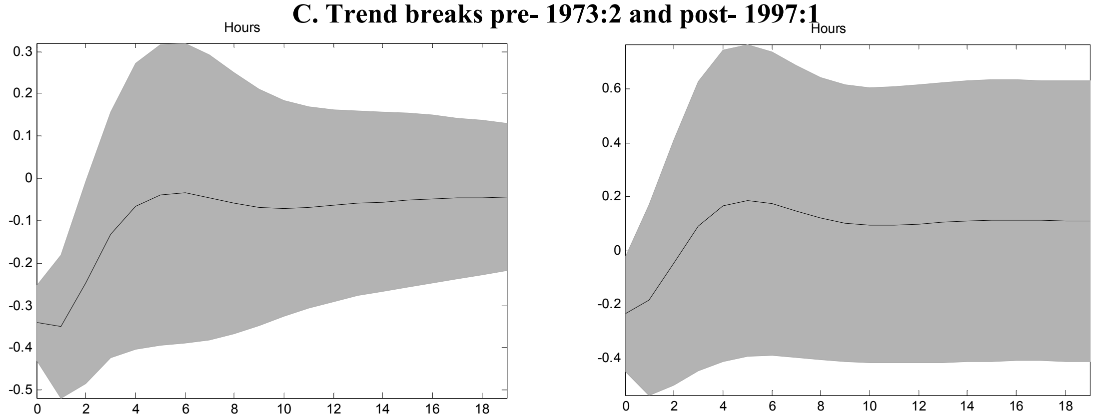
A. No Breaks



B. Trend Break pre 1972:2



C. Trend breaks pre- 1973:2 and post- 1997:1

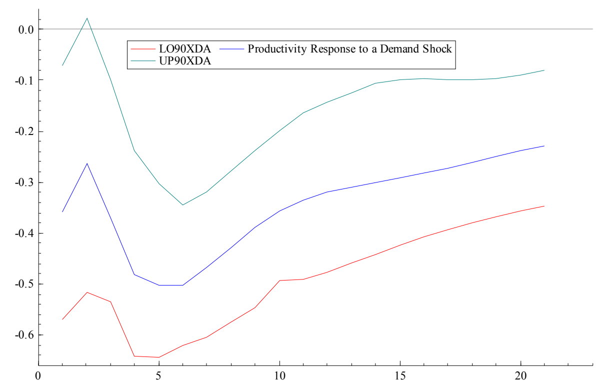


Notes: Impulse responses are from bivariate VARs estimated over 1948:2 to 2003:4. The first variable in the VAR is labor productivity growth (with identifying assumption that only technology shocks have a permanent effect on labor productivity); the second variable is labor hours per capita. In the left column, hours enters in (log) levels; in the right column, hours enters in growth rates. In panels B and C, labor productivity growth is a residual from a regression on time-period dummies shown.

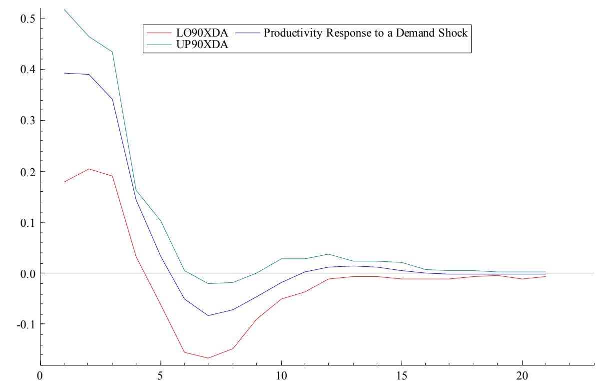
Figure 4
Responses of Labor Productivity to a Demand Shock

(no trend breaks, full sample)

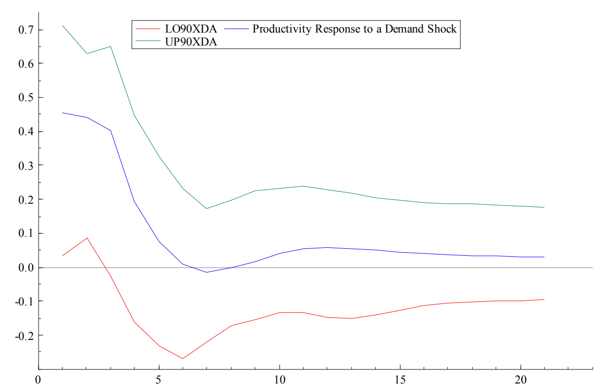
Level Specification

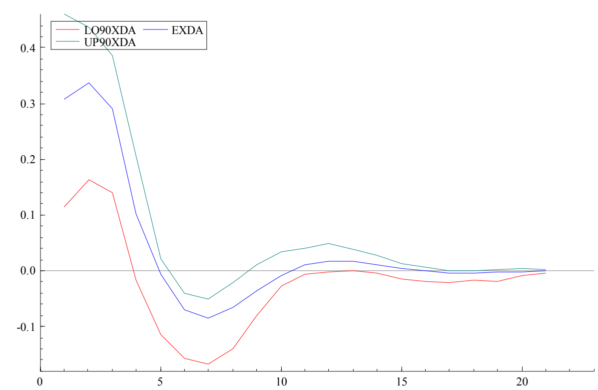


Difference Specification



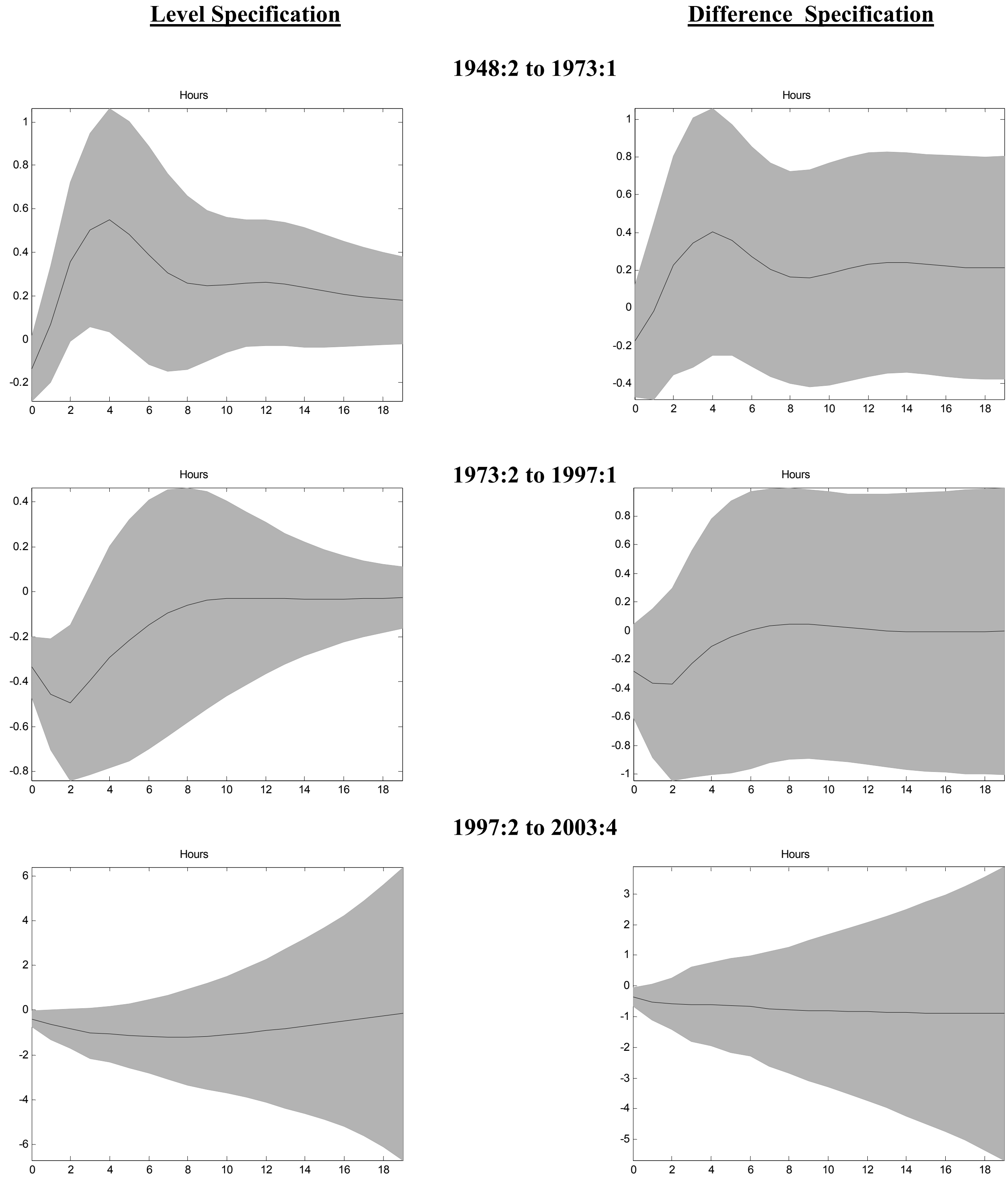
(pre-73 and post-97 trend breaks, full sample)





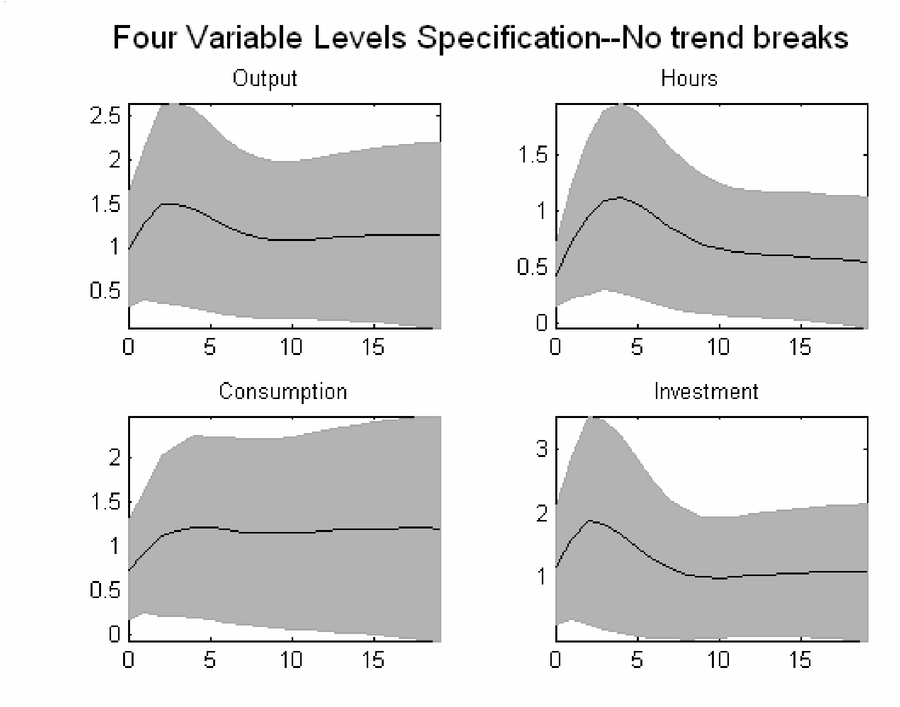
Notes: Estimates are from bivariate VAR and show response of labor productivity to the 'non-technology' shock in the system. Preliminary 90 percent confidence intervals for this figure are from 1000 Monte Carlo simulations. (Unlike the other impulse responses shown in this paper, they are also not centered, which accounts for the lack of statistical significance in the top panel.)

Figure 5
Subsample Estimates of the Response of Hours Worked to a Technology Improvement



Notes:

Figure 6:
Robustness to Larger VAR Systems:
Responses to Technology in a Four Variable Systems with No Breaks
(Sample period 1948:2 to 2003:4)



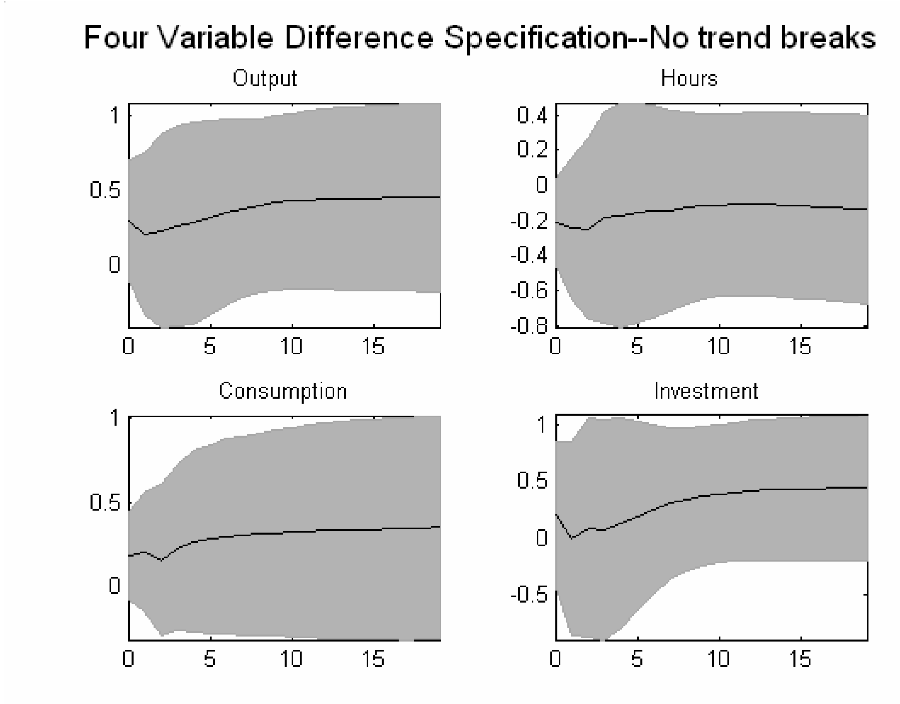


Figure 7
Robustness to Larger VAR Systems:
Responses to Technology in a Four Variable Systems after Removing Breaks
(Pre-1973:1 and Post 1997:1 breaks in labor productivity removed; sample period 1948:2 to 2003:4)

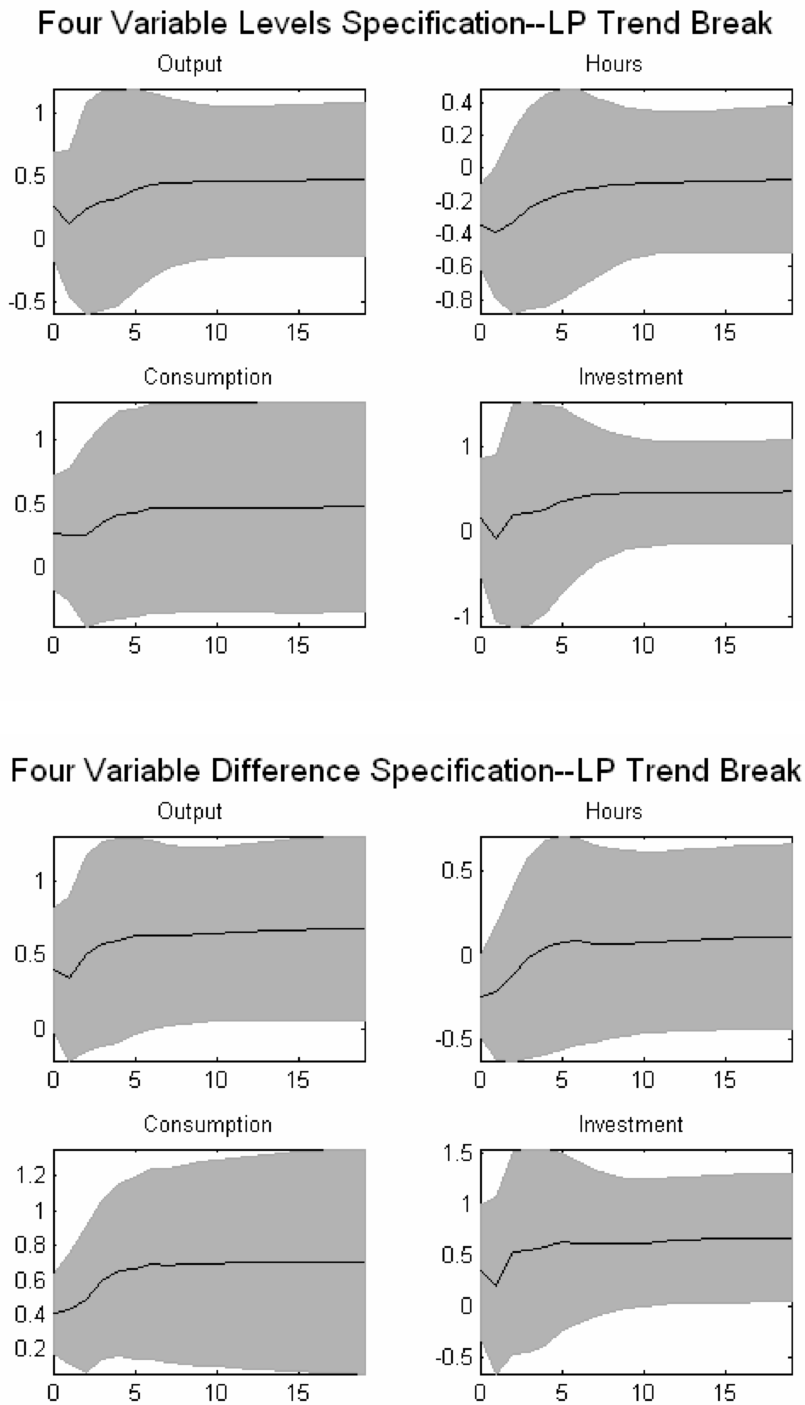
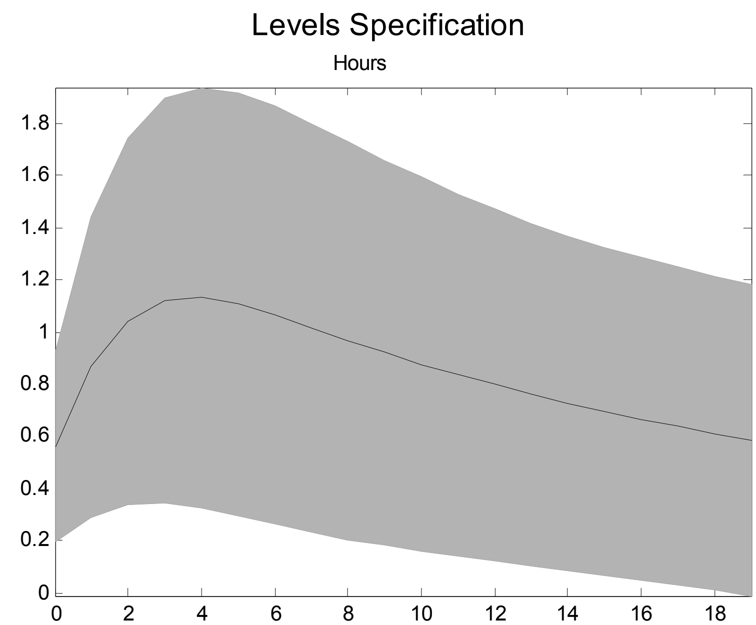
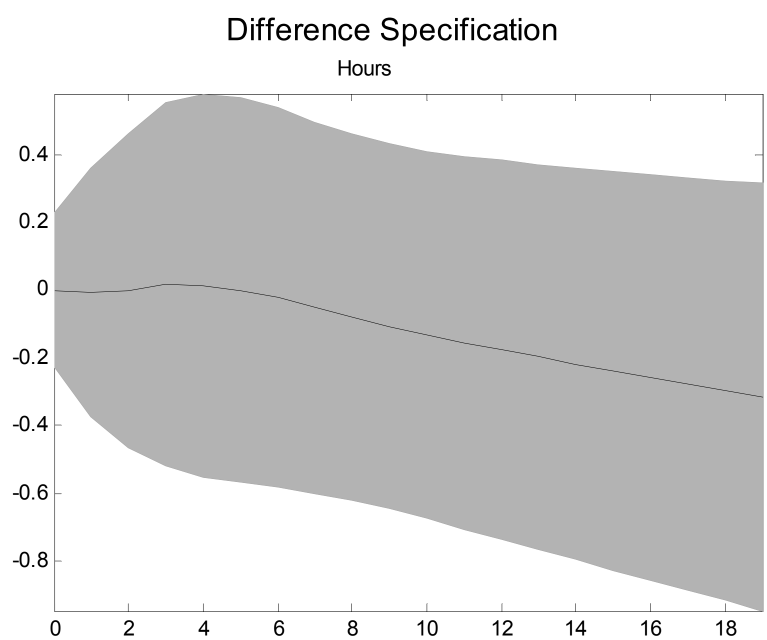


Figure 8
Replacing Actual Labor Productivity Growth with a 1-0-1 Dummy in the VAR
Sample Period 1948:2 to 2003:4, using actual data for hours



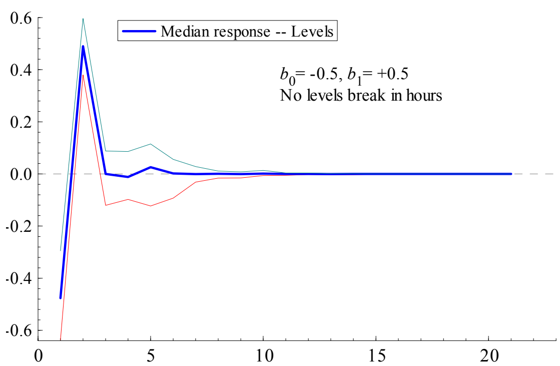


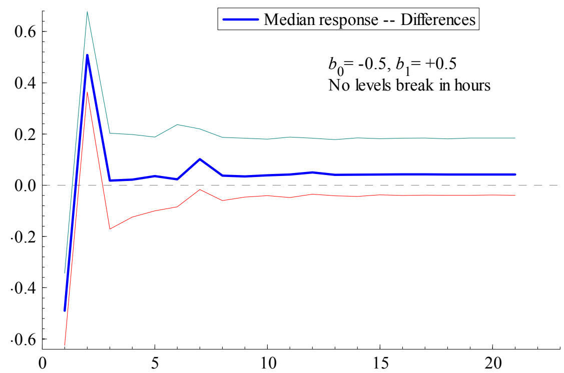
Notes:

Figure 9
Simulation Results

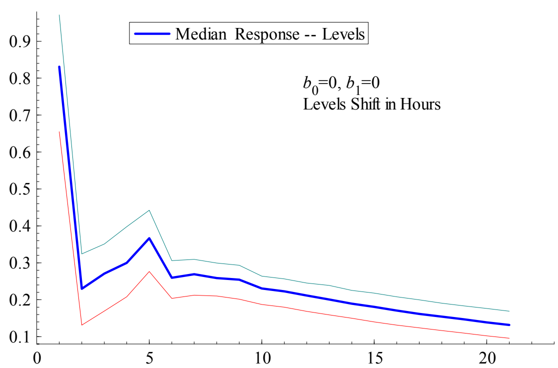
Level Specification **Difference Specification**

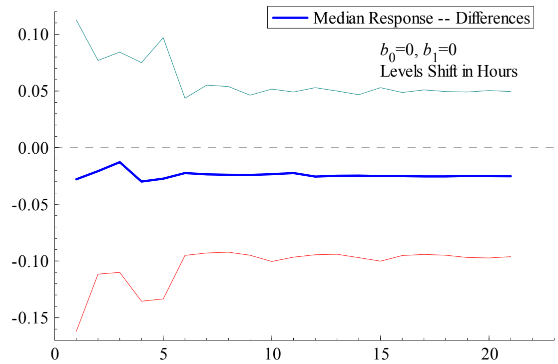
Case 1: No Hours-Levels Shift; Technology Affects Hours



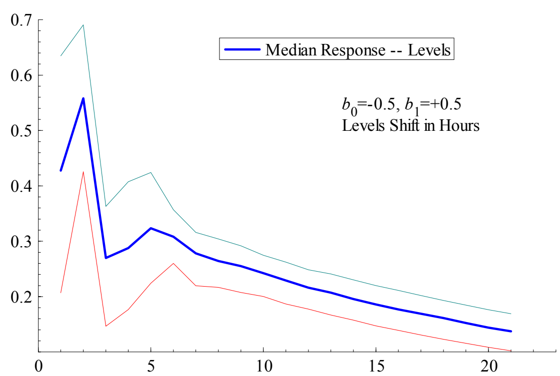


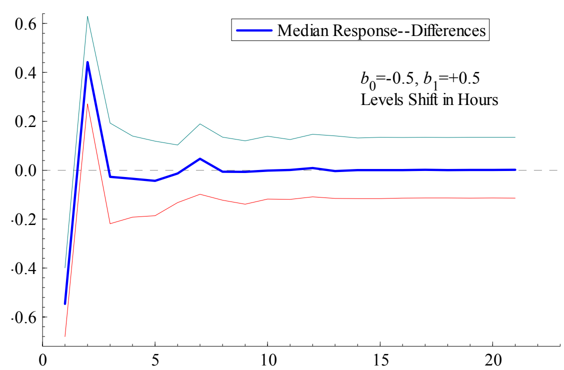
Case 2: Hours-Levels Shift, No True Effect of Technology on Hours





Case 3: Hours-Levels Shift AND Technology Affects Hours





Notes: For each specification, 250 simulated quarterly data sets were created from 1948:1 to 2003:4. Left column shows estimated responses of hours to a technology shock from bivariate VAR with hours in levels; right column shows estimated responses with hours in differences. Each figure shows median, 10%, and 90% impulse response. In all simulations, labor productivity growth slowed in 1973. In Cases 2 and 3, there is also a downward shift in the levels of hours in 1970. In Cases 1 and 3, the true impulse response shows a negative contemporaneous impact effect of -0.5 and a once-lagged impact effect of +0.5 (all other lags are zero).

Table 1

Regressions on Dummy Variables

	Labor Productivity growth		Hours Worked Per Capita (16+)	
Pre-1973:2 dummy	1.61 (0.48)	1.94 (0.51)	8.78 (0.69)	10.3 (0.69)
Post-1997:1 dummy		1.32 (0.75)		6.1 (1.0)
R^2	0.049	0.06	0.42	0.50

Constant terms suppressed. Labor productivity growth is at an annual rate. Hours worked are the log of an index, so the units have no natural interpretation. Standard errors in parentheses.