Comments Encouraged

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

John Fernald
Federal Reserve Bank of San Francisco

This draft: October 14, 2005

Abstract

Structural vector-autoregressions with long-run restrictions are extraordinarily sensitive to low-frequency correlations. This paper explores this sensitivity analytically and via simulations, focusing on the contentious issue of whether hours worked rise or fall when technology improves. Recent literature finds that when hours per person enter the VAR in levels, hours rise; when they enter in differences, hours fall. However, once we allow for (statistically and economically plausible) trend breaks in productivity, the treatment of hours is relatively unimportant: Hours fall sharply on impact following a technology improvement. The issue is the common high-low-high pattern of hours per capita and productivity growth since World-War II. Such low-frequency correlation almost inevitably implies a positive estimated impulse response. The trend breaks control for this correlation. In addition, the specification with breaks can easily "explain" (or encompass) the positive estimated response when the breaks are omitted; in contrast, the no-breaks specification has more difficulty explaining the negative response when breaks are included. More generally, this example suggests a need for care in applying the long-run-restrictions approach.

I thank Shanthi Ramnath, David Kang, Stephanie Wang, and Andrew McCallum for superb research assistance. I also thank Toni Braun, Susanto Basu, Jeff Campbell, Larry Christiano, Martin Eichenbaum, Jon Faust, Emilio Fernandez-Corugedo, Jonas Fisher, Neville Francis, Jordi Galí, Òscar Jordà, Lutz Killian, Valerie Ramey, John Williams, and Rob Vigfusson for helpful conversations and comments. I thank Valerie Ramey and Rob Vigfusson for providing computer code. 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.

JEL Codes E24, E32, O47. Keywords: Technology, business cycles, structural change.

1. Introduction

How does the economy react to fundamental shocks? Since Blanchard and Quah (1989), a growing body of work addresses this question using structural vector autoregressions (SVAR) with restrictions on the long-run effects of various shocks.[1] A prominent recent example is the literature sparked by Galí (1998), who used the long-run restriction that only technology shocks permanently affect the level of labor productivity. He finds that hours fall for a time after technology improves. This response is consistent with popular explanations for the early 2000s, when U.S. productivity growth was exceptional and hours worked fell.

Christiano, Eichenbaum, and Vigfusson (CEV, 2004) challenge Galí's results. They document the intriguing puzzle that the estimated response of hours changes sign when hours worked per capita enters the VAR in levels rather than differences: hours worked appear to *rise* after a technology improvement. Thus, a small and seemingly reasonable alternative specification completely reverses results. Unfortunately, little clear intuition is available about what drives results with long-run restrictions.

In this paper, I document analytically and via simulations the sensitivity of results to low-frequency correlations. In empirically relevant cases, low-frequency correlations—which need not be causal—completely drive the implied high-frequency impulse responses.

For concreteness, I focus on whether hours worked rise or fall following a technology improvement. CEV provide statistical evidence that rejects a unit root. In addition, they propose an encompassing methodology to choose from alternative specifications. They argue that data generated from the levels specification "explains" the difference results easily, but not the converse.[2] Thus, CEV conclude that hours probably rise after a technology improvement.

However, I find that once one allows for (statistically and economically plausible) trend breaks in labor productivity, competing empirical specifications yield a consistent answer. In particular, the levels specification robustly implies that hours worked fall on impact following a technology improvement. This conclusion holds in bivariate and larger systems, and when estimated over sub-periods that correspond to

---

[1] In addition to Blanchard and Quah (1989), Shapiro and Watson (1988) and King et al (1991) are early developers and promoters of the long-run-restrictions method. The method has been applied to a wide range of questions, such as the international propagation of business cycles (Ahmed, et al, 1993) and the properties of technology shocks.

[2] Francis and Ramey (2003) and Galí and Rabanal (2004) argue that even if hours per person does not literally have a unit root, assuming it does might help capture the persistence of the data. Section 6 discusses alternative approaches.

break dates. I discuss analytically and with simulations why the levels specification is sensitive to these breaks. I also apply the intuitive CEV encompassing methodology to whether or not to include breaks.

The source of the sensitivity to breaks is the low frequency correlation between labor productivity growth and the level of hours per capita. Figure 1 (panels A and B) shows the two series. Although highly variable from quarter-to-quarter or even year-to-year, average productivity growth was faster before the early 1970s and after the mid-1990s. Broadly speaking, hours 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. The bottom panel of Figure 1 shows the common low-frequency patterns by plotting HP-filtered trends.

It turns out that one needs to know little about the data other than this common high-low-high pattern to know that the estimated impulse response of hours to a technology shock is positive. I present a simplified case analytically in which it is clear that the low-frequency correlation dominates the relevant covariances of the VAR: One almost cannot help but find a positive impulse response.

Simulations illustrate the empirical relevance of the analytics for SVARs. First, I use actual hours per person but replace productivity growth with a dummy series with only low-frequency movement—equaling one before the early 1970s and after the mid-1990s, and zero in between. Hours appear to rise significantly when technology improves. Second, I use actual productivity but change the high and medium frequency components of hours. Using a bandpass filter, I estimate frequencies of 2 to 120 quarters in hours. I remove this filtered component from actual hours to identify the low-frequency trend; I then add back the filtered component with the sign *reversed*. The new series has the same low-frequency properties as actual hours, but is the mirror image at high- and medium-frequencies. Again, hours rise. Third, I generate random, independent series using the estimated univariate processes for productivity growth and hours. By chance, these series sometimes have an apparently significant high-low-high pattern over the sample. For series with a common high-low-high pattern, the impulse responses are strongly positive.

These examples make clear that even before running the regression, one expect the levels specification to imply a positive response of hours to technology. Nevertheless, if one believes that in the finite sample at hand, the low-frequency correlations in fact reflect causal links in the DGP, then those responses might be accurate (if inevitable). In this case, one would expect that the no-breaks specification should also explain the results from the break specification. To examine this issue, Section 6 applies the CEV encompassing methodology to distinguish between the breaks and no-breaks specification. Using synthetic datasets generated from the estimated breaks-specification, we can easily explain the observations that (i) the no-breaks specification yields a substantially positive response; (ii) the breaks specification yields a substantially negative response; and (iii) the F-test for two breaks is relatively large. The no-breaks specification has considerable difficulty explaining (ii) and (iii). Hence, the encompassing approach supports the breaks-specification over the no-breaks specification.

Any procedure that reduces the low-frequency comovement, whether it operates on productivity growth or on hours, could change estimated responses. Hence, this paper's analysis helps explain previous results. For example, Francis and Ramey (2005) find that the levels specification is sensitive to adjusting the population available to work for schooling, government employment, and retirement. Business-cycle models typically abstract from these factors, but they substantially change the low-frequency properties of hours per available person. One probably does not want these low-frequency factors to drive business-cycle facts about the high-frequency response to technology. Similarly, Francis and Ramey (2003) and Galí and Rabanal (2004) find that results are sensitive to removing a quadratic trend in hours. Section 6 considers a wide range of specifications. Those that control in one way or another for low-frequency trends yield similar technology series and pass a Granger-causality test from monetary innovations; the CEV levels specification, in contrast, yields technology innovations that show evidence of Granger-causality from monetary innovations.

Other related results include the original Blanchard and Quah (1989) paper, where the effects of supply shocks on unemployment reverse sign when they remove a post-1973 slowdown in trend output growth. More recently, Fisher (2005), in a model with "investment specific" technical change, finds that results are sensitive to using sub-samples that correspond to breaks in his series.

In addition to the empirical literature using long-run restrictions, this paper is related to a range of papers. Faust and Leeper (1997) highlight the theoretical limitations of long-run identification in finite data. Cooley and Dwyer (1998) and Erceg, Guerrieri, and Gust (EGG, 2005) simulate various dynamic stochastic general equilibrium (DSGE) models to assess the sensitivity of long-run-restriction results to particular economic environments.[3] Basu, Fernald, and Kimball (2004) use an augmented-growth accounting approach to identify technology shocks. They seek more directly to distinguish technology change from myriad short-run non-technological effects that affect the Solow residual (e.g., variations in factor utilization). As in this paper, they find that technology improvements reduce hours on impact.

2. Evidence for Trend Breaks

Productivity growth slowed down after about 1973 and sped up again after the mid-1990s. This section discusses univariate statistical evidence that the slowdown and speed-up represented structural change. Changes in mean growth rates could reflect unusual historical 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. But the central argument—that SVARS are sensitive to low-frequency correlations—applies even with no true structural change if, ex post, realized technology innovations simply had a higher mean before the early 1970s and after the late 1990s.

Empirically, we can model persistent growth shocks several ways. First, we can view them as regime shifts. Kahn and Rich (2004), for example, find evidence of "distinct switches in the early 1970s and late 1990s", with the U.S. economy shifting from a high-growth to a low-growth regime in the early 1970s and with a shift back in the late 1990s. Second, we can explicitly model a stochastic trend for productivity growth, implying that productivity itself is I(2). Roberts (2001), for example, estimates a time-varying trend with the Kalman filter. He argues for the statistical significance of the estimated changes in trend growth.

Third, we can model structural change in mean productivity growth. The post-war period appears to have few regime switches or low-frequency swings in growth. Hence, this trend-break approach tells a similar story. In addition, it is easier to apply in the SVAR context and is relatively transparent.

---

[3] As in this paper, the general lesson from the Monte Carlo exercises is that long-run-restrictions must be used with care. EGG emphasize the sensitivity of long-run identification schemes to small samples and to persistent (but not permanent) non-technological shocks. They generate artificial data from a range of calibrated DGE models and then estimate SVARs. They conclude that, used cautiously, long-run-restrictions might help discriminate between alternative models. For example, they find that the estimated impulse response for hours, though biased, is unlikely to yield an incorrect sign. More recent work by Chari, Kehoe, and McGrattan (2005) undertakes an exercise similar to EGG and also argue that responses are probably biased. But Christiano, Eichenbaum, and Vigfusson (2005) argue that, properly used, SVARs have little bias and are informative.

Bai and Perron (BP, 1998 and 2003) provide straightforward statistical tests for multiple structural change. Table 1 summarizes results for the mean growth rate of private-business labor productivity from 1950:2 to 2004:2. (Data are described in an appendix.) The so-called "double-maximum" tests, UD-max and WD-max—which test the null of zero breaks against the alternative of an unknown number of breaks—reject the null of zero breaks at better than the 5 percent level. Thus, structural change is likely.

The statistical evidence suggests two breaks, with a slowdown after 1973:1 and a speedup after 1997:1. Notably, we can reject the null of zero versus two breaks: The SupF test (the maximum F test) of 10.19 is statistically significant at the 5 percent level.[4] In addition, conditional on one break, the second break (in 1997:2) is easy to find: SupF (2|1) is highly significant. Conditional on two breaks, there is no evidence of further breaks. Note that it is hard to detect the first break alone, since mean growth is similar before 1973 and after 1997; as a result, we cannot reject the null of zero versus one break. Bai and Perron (2003) argue that double-maximum tests are particularly informative in cases such as these. There is, of course, considerable uncertainty about the exact break dates, but results that follow appear robust to alternative dates.

Break tests thus yield results qualitatively similar to the regime-shift and stochastic- trend results. All three approaches argue for relaxing the restriction of a constant mean productivity growth rate.

Finally, hours worked per capita (defined as hours per person aged 16 and older) is similar in being high early in the sample, low in the middle, and high again at the end. Figure 1.E shows HP-filtered data on hours-per-capita as well as private-business labor productivity growth; to accentuate low-frequency swings, I use a smoothing parameter of 14,400. The correlation between the HP trends in private business hours and productivity growth is 0.76; for non-farm business, the correlation is 0.47.

Table 2 shows results from regressing hours worked per capita (using population aged 16 and older) and labor productivity growth on the 1973:1 and 1997:1 dummies. The subsample dummies have large t-statistics for hours as well as for productivity. Bootstrapped significance levels (from the estimated, persistent AR(2) process) suggest that the break in hours is probably not statistically significant: the bootstrapped synthetic data

---

[4] I downloaded Bai and Perron's (1998, 2003) Gauss code from http://qed.econ.queensu.ca/jae/2003-v18.1/bai-perron/ (August 16, 2004). Tests are robust to heteroskedasticity and autocorrelation, allow the variance-covariance matrix to differ across regimes, and implement AR prewhitening. I checked significance levels by bootstrapping the four-variable VAR (setting constant terms and initial values to zero) discussed later, under the null of no breaks. UDMAX rejects the null of no breaks in favor of the alternative of an unknown number of breaks at the 5 percent level.

sets often have large low-frequency swings in the series. In addition, sharp breaks in the level of hours per capita would imply economically implausible growth rates in hours at a small number of dates.

Francis and Ramey (2005) argue that the low-frequency movements in hours per person aged 16 and up reflect trends in school enrollment, government employment (since the numerator is private hours), and demographics—those 65 and above generally don't work. After accounting for these factors, their alternative measure of the "working population" flattens the low frequency trends in hours per capita. Their work suggests that the forces that drive business-cycle fluctuations in hours are very different from those that cause the U-shaped pattern apparent in Figure 1.

I do not need to take a stand here on which is the "proper" denominator for hours-worked per person. I follow CEV (2004) and use population aged 16 and older; but Section 6 shows that results with trend-breaks in productivity are similar to the no-break results with the Francis-Ramey measure. This paper makes clear why results are sensitive to adjustments such as Francis and Ramey's that affect low-frequency trends.

3. Long-Run Restrictions and the Sensitivity to Trend Breaks

3.1 Structural VARs with long-run restrictions

Suppose that X_t is a vector of variables in the system, and ε_t is a vector of innovations. The moving average representation of the system is $X_t = C(L)\varepsilon_t$, where $C(L)$ is a matrix of lag polynomials. For concreteness, consider the bivariate system where X_t comprises productivity growth, Δp_t, and the level of hours per capita (or some other stationary transformation of hours), n_t. The key identification assumption is that only technology shocks have a permanent effect on the level of labor productivity. Other shocks (such as labor supply shocks, monetary shocks, or transitory technology shocks) might affect labor productivity in the short run, as would typically be the case in models with variable factor utilization. But by assumption, non-technology shocks have no permanent effect on labor productivity.

Shapiro and Watson (1988) discuss an easy way to impose long-run restrictions.[5] In the present context, their approach involves estimating the following set of regressions:

$$\Delta p_t = \sum_{i=1}^{q} a_{P,i}\Delta p_{t-i} + \sum_{j=0}^{q-1} a_{N,j}\Delta n_{t-i} + \varepsilon_t^Z \qquad (1)$$

[5] See CEV (2004) for a more complete description of how to implement the long-run restriction.

$$n_t = c + \sum_{i=1}^{q} b_{P,i} \Delta p_{t-i} + \sum_{i=1}^{q} b_{N,i} n_{t-i} + d \cdot \varepsilon_t^Z + \varepsilon_t^N \quad (2)$$

Hours, n_t , enters equation (1) in differences, which imposes that non-technology shocks do not affect the long-run level of labor productivity. The VAR has q lags, so the summation on hours-growth runs from 0 to q-1. Technology shocks might affect current hours growth, so we estimate (1) with instrumental variables. The standard instruments (which for the just-identified case yield results identical to the Blanchard-Quah matrix methods) are a constant, lags of productivity growth, Δp_{t-s}, and lags of the level of hours n_{t-s}, where s=1 to q. The residuals from (1) are the estimated technology shocks. Equation (2) adds the technology shock to the standard VAR equation. We estimate this equation with OLS using the estimated technology shocks.

With two variables, we identify two shocks. One ($\widehat{\varepsilon_t^Z}$) is the identified permanent shock to technology. The other shock, ε_t^N , in principle, captures all shocks (especially demand shocks) with at most a transitory effect on labor productivity. The specification generalizes easily. A larger VAR system is straightforward: Other variables are treated symmetrically to hours. We then identify additional "non-technological" shocks, but without further assumptions these are not structural. In the difference specification, we can re-interpret n_t as the growth rate of hours worked, so that Δn_t is the second difference of (log) hours. An appendix considers the case where there are two permanent technology shocks, with different dynamic properties.

In the empirical work, I follow CEV, Galí and Rabanal (2004), and Francis and Ramey (2003) and set q=4. I generate impulse responses by simulating the dynamic responses of the estimated system to a technology shock. For each lag in this response function, I report centered 90 percent confidence intervals.[6]

The Shapiro-Watson IV representation makes clear that long-run identification imposes that lagged hours (the instrument) are orthogonal to true technology shocks. This assumption of course seems reasonable; but in any given sample, it is unlikely to hold exactly. Given the common high-low-high pattern in productivity growth and hours in the actual post-war U.S. data, orthogonality imposes restrictions on the properties of estimated technology. In particular, it constrains the ability of technology shocks to explain the high-low-high pattern in labor productivity. This turns out to substantially affect the estimated shocks and responses.

---

[6] I thank Rob Vigfusson for sending me his code to calculate confidence intervals.

Blanchard and Quah (1989), who estimate a bivariate system with output and unemployment, remove pre-1974/post-1973 means from the rate of output growth to control for low-frequency movements. The next subsection compares results from estimating the long-run restriction with the data as is, and after taking out trend breaks from labor productivity before estimation.

3.2 Quarterly VAR results with long-run identification restrictions

3.2.1.BIVARIATE SPECIFICATION

Figure 2 shows the impulse responses from the bivariate SVAR discussed above. The left column uses labor productivity growth and the level of hours per capita (aged 16 and above); the right column of figures removes subsample means from labor productivity growth prior to estimation. (Results are virtually identical if I include sample-period dummy variables in the VAR itself, with or without the constraint that pre-1973 and post-1997 means are the same.)[7]

The left panel reproduces the CEV result that technology shocks appear to raise hours worked on impact. They are not quite statistically different from zero at the 10 percent level; we can, however, reject that the response is substantially negative. In contrast, with trend breaks, technology shocks appear to reduce hours worked quite sharply. The right panel yields qualitatively similar results to the difference and quadratic-detrended specifications in Galí and Rabanal (2004) and Francis and Ramey (2003), as well as in the augmented-growth-accounting of Basu, Fernald, and Kimball (2004).

The estimated impulse response of productivity to a non-technology shock, shown in the bottom row of Figure 2, is also extremely sensitive to the trend breaks. Panel D shows that in the levels specification, productivity falls sharply and significantly in response to a positive non-technology innovation. Although near-term effect could reflect movement down a stable labor demand curve, it is inconsistent with evidence that utilization changes raise productivity in response to demand shocks. In addition, as Francis and Ramey (2003) discuss, the decline in labor productivity is extremely persistent, even if not (by assumption) permanent.[8] Given that identification assumes that non-technology shocks have no permanent effect on

---

[7] An 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. In contrast, the existing literature has focused on how best to model the hours process.

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

productivity, this impulse response suggests misspecification. As Faust and Leeper (1997) suggest, the zero-long-run impact could be too weak a restriction here to achieve reliable identification.

With trend breaks removed, the estimated response is that productivity rises sharply following a non-technological shock, but for a brief period.

3.2.2. LARGER VAR SYSTEMS

CEV also report results from several larger systems. The simulation results in EGG focus on one of these systems, a four-variable system that adds the log of the nominal consumption-output ratio and the nominal investment-output ratio to the VAR. (Following CEV, I include durable consumption in investment; I include government with consumption of non-durables and services.)[9] EGG find that this system has reasonable properties. In the larger system, the only shock that is identified is the technology shock.

Figure 3 shows responses in this four variable system. As in the bivariate case, the no-breaks specification suggests that hours worked (Panel A, upper right) rises strongly; but the breaks specification (Panel B, upper right) suggests that hours worked falls. The breaks specification also suggests that investment doesn't change much on impact.

3.2.3. SUBSAMPLES FROM 4-VARIABLE SYSTEM.

Suppose we estimate the SVAR over subsamples corresponding to the estimated break dates. Figure 4 (panels A, B, and C) shows that with the four-variable VAR, in all subsamples technology shocks reduce hours worked. (This is true even for the very short final sample, though the VAR is explosive.) The decline in the 1973:2-1997:1 period is statistically significant.[10] In essence, the subsample estimation allows different constant terms (as well as other coefficients), so it is not surprising that it is similar to the breaks specification, which essentially allows different subsample constant terms for productivity growth.

Galí, López Salido, and Vallés (2002) also find results that are sensitive to sample period. They focus on presumed differences in the monetary policy reaction function. Panel D shows the 1979-1997 period, which

---

[9] 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 that adds inflation (measured by the GDP deflator) and the fed funds rate. See Section 5.

[10] Erceg, Guerrieri, and Gust (2005) explore the reliability of long-run results in short samples. The reliability of responses uniformly falls off for all variables. But the likelihood of getting the wrong sign on the hours response rises only slowly as sample length declines, so it remains unlikely that one would estimate the wrong sign on the response.

should incorporate the Volcker-Greenspan monetary policy while still excluding the trend breaks. In contrast to Galí et al, the four-variable levels specification still suggest that hours declines.

Finally, Fisher (2005) considers a more general model than the one in this paper, in which he explicitly seeks to identify two types of technology shocks, neutral and investment-specific.[11] Fisher argues for splitting his sample to reflect trends in relative investment prices. When he does so, he finds that hours tend to fall in response to either shock. More interestingly, the responses for both subperiods he considers lies below the full-sample responses. The low-frequency correlations in his data makes that sensitivity unsurprising.

4. How Do Low-Frequency Correlations Affect Results? Analytics and Simulations

4.1 Analytical Discussion

What drives the sensitivity of results to the treatment of low-frequency trends? This section argues that low-frequency comovement dominates the key covariances of the estimation. For analytic tractability, I focus on bivariate systems. The empirical results are qualitatively similar in bivariate and larger systems and the insights here apply to larger systems as well (though it is harder to sign the effects *a priori*).

One comes quite close in estimating the technology residuals as well as the impact effect of technology shocks on hours with the following simplified system. Only the current growth rate Δn_t appears on the right-hand side of the first, IV equation; the instrument is the lagged level n_{t-1}. I suppress constants:

$$\Delta p_t = a^S \cdot \Delta n_t + \varepsilon_t^{Z,S} \tag{3}$$

$$n_t = d^S \cdot \varepsilon_t^{Z,S} + \varepsilon_t^{N,S} \tag{4}$$

The "simplified" residuals $\widehat{\varepsilon_t^{Z,S}}$ have a correlation of 0.94 with the technology residuals $\widehat{\varepsilon_t^{Z}}$ from the "full" equation (1) (with q=4). $\hat{d}^S$ gives the estimated impact effect of technology on hours; $\hat{a}^S$ gives the estimated impact effect of a non-technology shock on productivity. Obviously, there are no dynamics.

The IV estimate of $\hat{a}$ is $n_{t-1}'\Delta p_t \big/ n_{t-1}'\Delta n_t$. In the data, the denominator is negative—when lagged hours are high, current *growth* in hours tends to be low. Hence, the positive covariance between lagged hours and current productivity growth implies a negative $\hat{a}$. Note that a negative coefficient for $\hat{a}$ implies that a

---

[11] The approach taken in this paper in principle seeks to identify a weighted sum of the two types of shocks. As Fisher discusses, correct identification with only a single composite shock imposes restrictions on dynamics.

positive non-technology shock (which pushes Δn_t up) also pushes productivity down, consistent with the empirical impulse responses.

The estimated value of $\hat{d}$ is $n_t{}' \widehat{\varepsilon_t^{Z,S}} \Big/ \widehat{\varepsilon_t^{Z,S}}{}' \widehat{\varepsilon_t^{Z,S}}$. The denominator is positive. Hence, the sign depends on the numerator, which is the covariance of estimated technology with current hours:

$$
\begin{aligned}
n_t{}' \widehat{\varepsilon_t^{Z,S}} &= n_t{}' \left(\Delta p_t - \hat{a}_t \Delta n_t\right) \\
&= n_t{}' \Delta p_t - \left(\frac{n_{t-1}{}' \Delta p_t}{n_{t-1}{}' \Delta n_t} \right) n_t{}' \Delta n_t
\end{aligned}
\tag{5}
$$

In the data, $n_t{}'\Delta n_t$ is positive, in contrast to the negative $n_{t-1}{}'\Delta n_t$. (These are, of course, the expected signs for a stationary autoregressive time series.) The common high-low-high pattern of hours and productivity growth implies that $n_t{}'\Delta p_t$ and $n_{t-1}{}'\Delta p_t$ are positive. Hence, both terms in equation (5) are positive. Therefore, in this context, the assumed orthogonality between technology growth and hours implies that the impulse response of hours to a technology shock is positive.

I now illustrate the impact of low-frequency correlation in examples. These illustrate that the insights from the simple static example extends to more complicated dynamic estimation.

4.2 Simulation 1: 1-0-1 Productivity Growth Series

Consider an extreme case where productivity growth follows the pattern in the top panel of Figure 5: equal to 1 before 1973:2 and again after 1997:1 and zero between these dates. What does the VAR imply if we run the bivariate VAR with this 1-0-1 productivity regression and the actual level of hours?

The impulse responses in the bottom panels look qualitatively (and, for hours, quantitatively) like the ones reported in Figure 2. A positive technology shock raises hours worked strongly and statistically significantly; and a positive non-technology shock reduces productivity. (Since productivity growth is so persistent in this example, one gets an extremely persistent negative effect of productivity to a non-technology shock.) Both of these responses match the predictions from the simple, non-dynamic analytical framework.

Figure 6A shows the estimated technology series itself along with innovations to an estimated AR(2) process for hours. Strikingly, every wiggle in hours is matched by a corresponding high-frequency wiggle in estimated technology—even though productivity growth itself is constant during subperiods.

Because of the low-frequency correlation between hours and productivity, the regression puts a negative coefficient on hours-growth in the productivity equation, in line with the discussion above. But now suppose we get a high-frequency increase in hours. Because of the negative coefficient on hours, the positive hours blip reduces the fitted value of productivity. But the productivity series itself changes only at two discrete points. So we need a positive innovation to technology to offset that.

The bottom panel of Figure 6 shows that the estimates attribute a lot of high-frequency hours movement to technology innovations. (The low frequency movement largely reflects the constant term). Note also that the estimated technology residuals in this specification are endogenous—anything that causes hours to move (e.g., demand shocks such as monetary shocks) cause estimated technology to move as well.

4.3 Simulation 2: Reversing High- and Medium-Frequency Components of Hours

High (and even medium) frequency movements in hours have little effect on estimated impulse responses. To show this, I modify the hours data at high and medium frequencies to see how responses change. To implement this, I use the Christiano-Fitzgerald (2002) bandpass filter to isolate components of the hours-per-capita series at frequencies of 2 to 120 quarters; in the top panel of Figure 8, this is the long-dashed green line that fluctuates around zero. The dashed line shows the estimated hours trend, defined as actual hours (the thick line) minus the high- and medium-frequency component.

The series I use for estimation, the thin line, *reverses* high and medium-frequency components. In particular, I take the trend line and add back the filtered component with the sign reversed.

The impulse responses in the bottom panels again look qualitatively and quantitatively like those in Figure 3. Hours worked rise strongly in response to a positive technology shock; a positive non-technology shock reduces productivity. Both responses are statistically significant. (The responses somewhat exaggerate what we see in the actual data.) Thus, the estimated responses are largely invariant to reversing the frequency components below 120 quarters. This makes clear that the responses are driven by low frequency movements.

4.4 Simulation 3: Selecting on Series with Apparent Breaks

In the analytics, the true data generating process is not really the issue: Rather, in the realized data sample, hours and productivity growth have a low-frequency correlation. To highlight this point, I simulated 230 quarters of data for two data series using the following univariate DGPs (the coefficients in the hours regression are given by the OLS estimates from 1947:3 to 2004:2):

$$dp_t = \varepsilon_t^Z$$
$$n_t = 1.51 \cdot n_{t-1} - 0.53 \cdot n_{t-2} + \varepsilon_t^N$$

Productivity growth is white noise. The DGP for hours worked is a highly persistent AR(2) process. The disturbance terms are normally distributed (0,1). Figure 8.A shows that the median response from applying the SVAR to 1,000 pairs of simulated data is very close to the true value of zero.

As Section 2 reported, a 1973:2 - 1997:1 dummy has a large t-statistic for both hours per person and productivity growth. Purely by chance, some of the time the simulated univariate processes for hours and labor productivity also show large t-statistics. With the same DGP, I regressed each simulated data series on a 1973:2 to 1997:1 dummy. For both hours and labor productivity, I saved the first 2,500 series where the dummy was negative and had a t-statistic as large in magnitude as 3.5. I then paired up the series.

By construction, the series are not related. But for the pairs of series that appear to have large breaks, Figure 8.B shows that the estimated impulse responses are almost uniformly positive. Econometrically, we have selecting series in which, by chance, true technology shocks are, in fact, correlated with lagged hours.

Figure 8.C shows the effect of taking out apparent 1973 and 1997 trend breaks from labor productivity. The estimates show little bias, and the distribution (though wide) is similar to the unconditional distribution in 8.A. When I difference hours in panel D the specification also has little bias and, for the initial effect, is relatively tightly estimated, despite the fact that the difference VAR is "overdifferenced." (Applying the difference specification to the unconditional simulated data also shows little, if any, bias on impact).

5. Encompassing

Given the low frequency correlation between hours worked and productivity growth, the levels specification inevitably yields a positive impulse response, regardless of the true response. Nevertheless, if the low-frequency correlation reflects causal links in the DGP, then the levels specification might be correctly exploiting those links and yielding the correct impulse response. (One would, of course, expect that the higher frequency variation would yield evidence consistent with those responses, which it does not.)

CEV (2004) argue forcibly that encompassing tests provide an appealing statistical approach to compare alternative specifications. They compare levels versus difference specifications for hours. The levels specification, they find, relatively easily explains the divergent results from the two specifications; the difference specification has more difficulty. In addition, the difference specification implies that the levels specification should suffer from a weak-instruments problem that is not to be apparent in the data. Thus, they argue that the levels specification encompasses the difference results, but not the converse.[12]

Here, I apply the encompassing approach to the issue of whether or not to include trend breaks in the VAR. The specification *with* breaks appears more plausible than the specification without such breaks. In particular, the specification with breaks can much more easily explain the observations that (i) the levels specification on its own yields a substantially positive impulse response; (ii) after removing two breaks (those with the highest level of significance) from labor productivity, the levels specification yields a notably negative impulse response; (iii) the F statistic associated with the breaks test is relatively large.

In particular, I estimated bivariate and four-variable VARs from 1951:2 to 2004:2. The breaks specification includes dummy variables in the VAR for the pre-1973:2 and post 1997:1 periods. I generated 500 bootstrapped synthetic datasets for each estimated DGP. (Each synthetic dataset starts from the actual initial 1950:2-1951:1 data. These initial values along with the (lower) estimated constant term for hours imparts a downward trend to the synthetic hours series and helps the no-breaks specification explain the large F statistic.) For each synthetic dataset, I estimate two break dates endogenously for labor productivity growth by regressing the series on every possible pair of break dates.. I exclude 10 percent at each tail and impose that there are at least 24 quarters between estimated "breaks".

I follow CEV and test on the average response during the first six quarters. For example, the bivariate levels specification implies that the average response of hours to a technology shock over the first six quarters is 0.70 percent. The "breaks" specification implies an average response of -0.41 percent.

Table 3 summarizes results from the encompassing tests. Rows 1A and 1B use synthetic data generated from the no-breaks specification (two-variable and four-variable, respectively). Rows 2A and 2B incorporate the trend breaks into the DGP. Column (1) shows that regardless of the DGP, when we estimate the VAR ignoring any possible breaks, the average levels specification is almost always positive. In particular, even if

---

[12] CEV argue that a covariates-adjusted Dickey Fuller test has much more power to reject a unit root. A related argument extends to their encompassing tests: The difference specification has a very difficult time explaining the explanatory power of lagged hours per capita as an instrument for the current growth rate of hours per capita.

the DGP has breaks, and if the true response is negative, then—consistent with the earlier simulations in this paper—it is no surprise that when we ignore the breaks, we find a positive response.

More interesting is column 2, where we estimate and impose two break dates, regardless of whether there are true breaks. In the no-breaks case (rows 1A and 1B), we find a negative response about a quarter of the time. When there are breaks (rows 2A and 2B), though, we are much more likely to find the negative response. (That it's only about 80 percent of the time reflects that we are looking at the response over six quarters, so that even initially negative responses might cumulate to a positive 6-quarter response.)

Column 3 shows the likelihood of getting an F statistic as large as 6.83 (the two-break F value with no allowance for heteroskedasticity or autocorrelation). In the no-breaks specification, this occurs about 10 percent of the time.[13] In the breaks specification, this occurs about 90 percent of the time.

Columns 4 and 5 combine the first three results. With the no-breaks DGPs, only about 1/5 of the time do we get the opposite signs. And only 2 to 3 percent of the time can the DGP explain a positive response with no breaks, a negative response with breaks, and a large F. In contrast, the breaks specifications in rows 2A and 2B easily explain all three observations. In the bivariate case, 64 percent of the synthetic datasets are consistent with all three; in the four-variable case, nearly half are.

Hence, the breaks specification can much more easily explain the empirical observations we see in the data. The logic of the CEV encompassing approach argues in favor of the results from the breaks specification when compared with the no-breaks specification.

6. Robustness and Alternative Approaches

So far, I have focused on controlling for sub-sample changes in trend productivity over time. Two issues arise. First, if we interpret these as shocks to the economy's underlying growth rate, then those shocks themselves should affect behavior in ways we have not modeled. Hence, the "breaks" VAR itself could be misspecified. Second, there are other ways to control for low-frequency correlations. Section 6.1 discusses several methods of controlling explicitly for low-frequency movements and growth shocks. Section 6.2

[13] The F statistic is smaller than the Bai-Perron test that allows for heteroskedasticity and autocorrelation. The relatively large share of the time in which the bootstrapped values yield such a large F reflects in part the constant terms, which are far from the initial values. So each bootstrapped simulation builds in considerable low-frequency movement,

compares results from a range of specifications. The important and robust conclusion is that once we control for low-frequency correlations, the specifications identify similar shocks and have similar responses.

6.1 Alternative Approaches with Low Frequency Movement

Take out the apparent break. Removing the estimated break is appropriate if there are no significant transition dynamics in productivity to a growth shock (so we can simply take out the change in mean) and if there are no significant differences in the response of employment to such a shock. If the apparent break is, indeed, "spurious," the approach is also appropriate: If we assume people knew the DGP and were, therefore, surprised by the realizations of productivity, then there is no concern that the behavioral response of employment differs based on the (ex post) persistence of the shock. Finally, the appendix discusses the formal restrictions placed on the dynamics if there are distinct responses to growth shocks. The appendix also discusses an approach that accounts explicitly of growth shocks; results appear qualitatively unaffected.

Estimate over sub-samples. The subsample responses discard the periods where the shocks occur, so the dynamics shouldn't be driven by those shocks. As reported above, the subsample responses suggest hours fall when technology improves. However, the short samples lead to relatively large confidence intervals and potentially exacerbate small-sample biases.

Discard data around the break dates. If we run the break specification where we omit the years around the breaks themselves, then the transitory dynamics associated with the breaks should have little effect on estimates. This makes little qualitative difference to results. For example, running the SVAR over the sample 1951:2-1966:1, 1977:1-1994:1, and 2000:1-2004:2, we still find a substantially negative response.

Differencing. Differencing could help deal with the effects of structural change, even if hours were, in fact, levels stationary.[14] Whether over-differencing biases results is an open issue. EGG (2005) discuss results from simulated DGE models, in some of which over-differencing is a problem; in others it is not.

Low-order deterministic trend. Francis and Ramey (2003) and Galí and Rabanal (2005) suggest removing a quadratic trend from hours. CEV (2004) report that if they detrend *all* variables from 1959:1-

---

helping the test explain a large F. Setting the constants and initial values to zero substantially reduces the proportion of cases with large Fs. Larry Christiano pushed me to include the F test in my encompassing results.

[14] Clements and Hendry (1999 suggest that with structural change, overdifferencing often works well for forecasting: "With structural change, well-specified models do not necessarily outperform badly specified models" (p46).

2001:4, their 6-variable system does not yield a negative. Adding either the 1950s or the 2000s restores the negative relationship, as shown below.

6.2 Similarity of shocks and responses across specifications

So far, we have focused on two- or four-variable specifications for the private business economy, with or without removing sub-sample means from productivity growth. Table 4 compares correlations of estimated technology residuals from various specifications and shows the estimated impact effect of technology on hours. To conserve space, I show correlations with the two- and four-variable VARs with trend breaks, and with the annual technology innovations estimated by Basu, Fernald, and Kimball (BFK, 2004). BFK use a completely different identification scheme—estimating industry production functions and aggregating residuals. Their approach is data intensive, but relaxes some of the implicit assumptions underlying the long-run restriction, such as perfect competition/constant returns (with increasing returns, total factor productivity rises if inputs rise). The BFK innovations are available annually through 1996. For the BFK correlations, I annualize the quarterly VAR residuals for the years 1953-1996.

Table 4 shows that once we control for low-frequency correlations, results are consistent across a range of specifications. The top of table (lines 1 to 14) controls in some way for low-frequency movements: removing sub-sample means from productivity growth; removing an estimated quadratic trend from all variables; using the Francis-Ramey measure of available population to construct hours per person (with or without productivity trend breaks); taking differences of hours. The bottom of the table shows the unmodified CEV specifications with no control for breaks. For both the break and no-break specifications, I also show non-farm business, where hours per person 16+ shows less of a U-shape (though it is still there).

When we control for low-frequency movements, the correlations of estimated technology shocks across methods are very high. In addition, all shocks are statistically significantly correlated with the BFK shocks. All of the specifications imply a negative impact response of hours to a technology innovation, typically in the -0.4 to -0.5 range. Thus, it is clear that these methods, though they differ in how they control for low-frequency correlations, are identifying similar shocks and they have similar responses.

In contrast, the specifications that do not control for low-frequency movements (lines 15-19) look very different. They have lower correlations with the break-specification shocks (and, though not shown, with the other specifications in the table). Correlation with the BFK shocks is generally much lower; the two-and four-variable business-sector residuals are not statistically correlated. In all cases, these approaches yield a positive response. The no-breaks specifications thus yield different shocks and responses.

Section 5 argued that on statistical grounds, the breaks-specification was preferred. Francis and Ramey (2003) suggest that Granger-causality tests might also provide insight into the reliability of different methods. The final column of Table 4 shows p-values from regressing the estimated technology shocks on four lags of a monetary-shock variable; I use the quarterly monetary-shock residuals estimated by BFK (2004) from an identified VAR. There is little evidence that these shocks have explanatory power for the series that control for low-frequency correlations; the most significant value from the 14 regressions is 0.10. In contrast, the no-breaks specifications have much more significant p-values; in several cases, better than 0.05.

Thus, Granger-causality tests suggest that the breaks specifications pick up the endogenous response of the economy to other shocks. The simulations in this paper help explain not only why the shocks look different but why monetary shocks affect them. Consider, for example, the extreme 1-0-1 productivity-growth simulation. Every wiggle in hours is matched by a wiggle in estimated technology. Thus, any shock that affects hours, including monetary innovation, would cause the estimated technology shock to move, as well.

Finally, I note one qualification to the choice of specifications. The existing literature, including this paper, has focused exhaustively on the properties of productivity growth and hours. But in the larger specifications, other variables show trends, as well. Although results are relatively robust to adding these variables, these trends could be affecting the estimates in unknown ways. Only the quadratic-trend specifications have any controls for these trends. Thus, the quadratic-trend specifications might be the most reliable of the four- and six-variable specifications.

7. Conclusions

The goal of this paper is to better understand the strengths and weaknesses of VARs identified with long-run restrictions. This widely used approach is extremely sensitive to low-frequency correlations in the data. I document this sensitivity analytically and via simulations, focusing on the example of technology and hours. In the so-called levels specification, the apparent positive impulse response is driven by the common high-low-high pattern of productivity growth and hours per person in the post-war period. After allowing for the statistically significant and economically plausible productivity slowdown and speedup, the impulse response reverses sign and turns negative.

The resulting negative response of hours is consistent with other reasonable data choices that affect the low-frequency properties of the data used, such as using low-order trends or alternative demographic adjustments. The argument in this paper about low-frequency correlations explains the sensitivity to these alternatives. In addition, the simulations make clear that the low-frequency correlation in the data need not be causal. Indeed, there is little reason to think that the low-frequency movements in hours per person over 16 are driven by the same factors that drive business-cycle movements. Francis and Ramey (2005) discuss a variety of low-frequency trend factors behind the U-shape in hours.

My main recommendation for practitioners is to check sensitivity to alternative detrending methods. If results are sensitive, that should raise warning flags. Practitioners use several approaches to justify specifications, none of which is perfect. First, consider *a priori* reasoning, e.g., hours per capita cannot have a unit root. But this paper shows that low-frequency movements can lead to problems even if they don't literally reflect a unit root. Second, one can use statistical tests. Again, unit-root tests might reject a unit root for hours per person and one might not be persuaded that the productivity trend breaks are statistically significant; but the common low-frequency correlations in the data nevertheless drive results. Third, one can do Monte Carlo simulations of specific economic models (e.g, Erceg, Guerrieri, and Gust, 2005). Like this paper, that literature also suggests that long-run restrictions need to be used with care. Since the specific model is likely to abstract from low-frequency movements—particularly if they are coincidental, reflecting factors such as demographics that are not modeled—it could miss important features of actual data. Indeed, macroeconomists often, quite reasonably, ignore low-frequency fluctuations entirely—e.g., by using an HP or bandpass filter. In short, all of these alternatives contribute to our understanding but are incomplete.

In terms of the technology and hours debate, the robustness of results to whether hours is modeled as stationary, difference stationary, or otherwise adjusted; and the consistency of results identified with long-run restrictions with those from augmented growth accounting (as in Basu, Fernald, and Kimball 2004) suggest that in the post-war period, increases in the level of technology reduce hours worked on impact. Research continues on what model features best explain this contractionary effect.

APPENDIX 1: DATA

I follow Christiano, Eichenbaum, and Vigfusson (2004) and use BLS data on hours and labor productivity for the business economy. The BLS population data have occasional discrete jumps: When the BLS obtains new information on the population (e.g., at decennial censuses), it does not revise previous data. I use a series adjusted at the Federal Reserve Bank of Chicago that smoothes through these jumps; this alternative series appears to have little or no effect on results. Other national accounts data for larger VAR systems were downloaded via Haver Analytics. For the fed funds rate, I use the average effective fed funds rate from Haver back to 1954:3 (when the series begins); prior to that date, I use the NY Fed discount rate. (Using the 3-month treasury rate or using the NY Fed discount rate into the 1960s—when the fed funds market became more active—give similar results.) Data were downloaded October 2004.

In most cases, I start my sample in 1950:1, which, with four lags of growth rates, allows me to run regressions beginning 1951:2. Although several years of earlier data BLS data are available, there are at least two reasons for discarding those data. First, in the aftermath of World War II, there was an enormous amount of economic adjustment that affected labor productivity, employment, and other variables in important ways; the dynamics of these variables are not necessarily well captured by the same model that captures the dynamic response to more "normal" technology and non-technology shocks. Conceptually, one can think of the war and its end in 1945 as a large shock that we are not explicitly accounting for in the estimation; the dynamic effects were almost certainly still in play in the late 1940s. Second, the earlier quarterly data is likely to be of lower quality than the later data. For example, Young (1974) cites several studies suggesting (both from statistical analysis and from "expert opinion") that the early quarterly data from the post-war period are less accurate than later data. Nevertheless, results appear only slightly affected qualitatively by experimenting with different choices of starting dates from 1948:2 to 1959:1.

APPENDIX 2: RESTRICTIONS IMPOSED BY IGNORING GROWTH SHOCKS IN ESTIMATION

Suppose that there are two types of technology shocks—unit-root shocks to the level of technology, and shocks to the growth rate of technology. In the moving average representation:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ g_t \\ \varepsilon_t^N \end{pmatrix} \tag{6}$$

μ_t incorporates the time-varying constant term in the productivity equation, so it directly captures changes in trend productivity growth; $g_t \equiv \mu_t - \mu_{t-1}$ captures the transition dynamics associated with such changes. I omit the constant term for hours for simplicity. Faust and Leeper (1997) and Blanchard and Quah, (1989) discuss conditions under which we can represent this system in terms of a composite aggregate supply shock (defined as a linear function of the two underlying supply shocks) and a demand shock. In essence, we need the following conditions on the matrix *C(L)*:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ g_t \\ \varepsilon_t^N \end{pmatrix} = \begin{bmatrix} G_{11}(L) & G_{12}(L) \\ G_{21}(L) & G_{22}(L) \end{bmatrix} \begin{bmatrix} D_{11}(L) & D_{12}(L) & 0 \\ 0 & 0 & D_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ g_t \\ \varepsilon_t^N \end{pmatrix}$$

For the estimated supply shock to represent contemporaneous shocks alone, we need $D_{ij}(L)=D_{ij}$, all i,j,:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} G_{11}(L) & G_{12}(L) \\ G_{21}(L) & G_{22}(L) \end{bmatrix} \begin{bmatrix} D_{11}\varepsilon_t^Z + D_{12}g_t \\ D_{23}\varepsilon_t^N \end{bmatrix} \tag{7}$$

This representation implies that, up to a scalar, the two sources of supply shocks have identical impulse responses on either Δp_t or n_t. Theory suggests that the sign of the effect of the two growth shocks on hours worked might differ (e.g., Campbell (1994), Pakko (2001), and Edge, Laubach, and Williams (2004)). If the employment response does diverge for level versus growth shocks, then the coefficients D_{11} and D_{12} must have opposite signs. If so, then the system in equation (7) implies that the response of productivity to the two sources of supply shocks must also have different signs. Since it is reasonable to impose that a shock to the level of technology has a positive effect on the level of productivity, then the transitory dynamics of a growth shock must be to push productivity growth down temporarily relative to its new mean. This is not a priori implausible, since the mean itself captures the direct effect of the change in mean growth. Indeed, given transitional dynamics associated with capital accumulation, such a path might be reasonable. In addition, the dynamics are constrained by the *G(L)* matrix to be proportional apart from scaling/sign.

In any case, one can relax these restrictions under the assumption that the break dates are known. Suppose we rewrite equation (6) as follows:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ g_t \\ \varepsilon_t^N \end{pmatrix} = \begin{bmatrix} C_{11}(L) & C_{13}(L) \\ C_{21}(L) & C_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t, or \tag{8}$$

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \tilde{C}(L) \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t$$

We can now express this in VAR form:

$$\tilde{C}(L)^{-1} \begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \tilde{C}(L)^{-1} \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t \tag{9}$$

or,

$$\begin{aligned} \begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} &= B(L) \begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} + \tilde{C}_0 \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \tilde{C}_0 \tilde{C}(L)^{-1} \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t \\ \begin{bmatrix} \Delta p_t \\ n_t \end{bmatrix} &= \left(I - B(L)\right) \begin{bmatrix} \mu_t \\ 0_t \end{bmatrix} + B(L) \begin{bmatrix} \Delta p_t \\ n_t \end{bmatrix} + \tilde{C}_0 \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \left(I - B(L)\right) \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t \end{aligned} \tag{10}$$

where $[I - B(L)] \equiv \tilde{C}_0 \tilde{C}(L)^{-1}$.

If we assume the break dates are known with certainty, at the time they occurred, then we can estimate equation (10). In particular, we impose the long-run restriction and augment the SVAR with a dummy variable (including lags) for the break dates. Conceptually, this is similar to the approach taken to government spending dummies by Ramey and Shapiro (1998) and, in subsequent work, by Edelberg, Eichenbaum and Fisher (1999).

In real time, there is surely considerable uncertainty about the timing of breaks. But agents are likely to have better ideas about their own permanent income prospects than an aggregate analyst, so they may respond to changes in trend growth before an econometrician can detect the change in aggregate productivity data. In addition, lags that arise from slow learning may be incorporated into the lag coefficients, assuming that these learning effects are reasonably symmetric between the early 1970s and late 1990s experience.

To implement this, one needs dates for the growth shocks. As a benchmark, I used the estimated break dates of 1973:2 and 1997:2. I impose a priori that the post-1973:1 experience was the mirror image of the post-1997:1 experience; so in the regression specification, I set g_t to be the following:

$$g_t = \begin{cases} -1 & t = 1973:2 \\ 1 & t = 1997:2 \\ 0 & \text{otherwise} \end{cases}$$

I do not show these results, since they are not noticeably different from those in the text for the response of hours to a shock to the level of technology: These shocks reduce hours worked on impact. This result is robust to a wide range of alternative dates for the two breaks. .

The estimates imply that shocks to the *growth rate* of technology raise hours worked quite substantially, with a peak effect about three years out. Nevertheless, these results are at best suggestive (and I do not report the results), since they essentially reflect only two observations: First, the productivity slowdown in 1973:1 was followed by a prolonged reduction in hours worked; second, the productivity acceleration in the late 1990s was followed by a prolonged increase in hours worked. The results are consistent with a view that growth shocks raise hours worked, but with two observations (and with uncertainty about the exact timing), these results are not dispositive. Instead, I interpret these estimates as suggesting that the main results in the text are robust to allowing growth shocks to have distinct effects from the response to levels shocks.

References

Ahmed, Shaghil, Barry Ickes, P. Wang, and B.S. Yoo (1993). "International Business Cycles." *American Economic Review,* 83: 335-359.

Andrews, Donald (1993). "Tests for Parameter Instability and Structural Change With Unknown Change Point." *Econometrica*, Vol. 61, No. 4, pp. 821-856.

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

Bai Jushan and Pierre Perron (1998). "Estimating and testing linear models with multiple structural changes." *Econometrica* 66 47-78.

Bai, Jushan and Pierre Perron (2003). "Computation and analysis of multiple structural change models." *Journal of Applied Econometrics*, Volume 18, Issue 1, Pages 1 – 22.

Basu, Susanto, John Fernald, and Miles Kimball (2004). "Are Technology Improvements Contractionary?" NBER Working Paper 10592.

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).

Campbell, John Y. (1994). "Inspecting the Mechanism: An Analytical Approach to the Stochastic Growth Model." *Journal of Monetary Economics* 33:463–506, June 1994.

Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan (2005). "Are Structural VARs Useful Guides for Developing Business Cycle Theories?" Working Paper 631, Federal Reserve Bank of Minneapolis.

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.

Christiano, Lawrence, Martin Eichenbaum, and Robert Vigfusson (2005). "Assessing Structural VARs." Downloaded June 3, 2005 from http://www.faculty.econ.northwestern.edu/faculty/ christiano/research/VAR/manuscript.pdf (draft dated May 8, 2005).

Christiano, Lawrence and Fitzgerald (2002). "The Bank Pass Filter." *International Economic Review.*

Clements, Michael and David Hendry (1999). *Forecasting Non-stationary Economic Time Series.* MIT Press.

Edge, Laubach, and Williams (2004). "Learning and Shifts in Long-Run Productivity Growth." WP 2004-04, Federal Reserve Bank of San Francisco.

Edelberg, Wendy, Martin Eichenbaum, and Jonas D.M. Fisher (1999). "Understanding the Effects of a Shock to Government Purchases," *Review of Economic Dynamics*, Academic Press for the Society for Economic Dynamics, vol. 2(1), pages 166-206.

Erceg, Christopher, Luca Guerrieri, and Christopher Gust (2005). "Can Long-Run Restrictions Identify Technology Shocks?" Forthcoming, *Journal of the European Economic Association.*

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.

Fisher, Jonas (2005). Working Paper, Federal Reserve Bank of Chicago. "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks." (Draft dated May 16, 2005).

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

Francis, Neville and Valerie Ramey (2004). “The Source of Historical Economic Fluctuations: An Analysis Using Long-Run Restrictions.” Forthcoming, *NBER International Seminar on Macroeconomics 2004*, eds. Richard Clarida, Jeffrey Frankel, and Francesco Giavazzi.

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.

Galí, Jordi and Pau Rabanal (2004). "Technology Shocks and Aggregate Fluctuations: How Well Does the Real Business Cycle Model Fit Postwar U.S. Data?," *NBER Macroeconomics Annual, 2004.*

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

Kahn, James A. and Robert Rich (2004). “Tracking the New Economy: Using Growth Theory to Detect Changes in Trend Productivity.” Working Paper, Federal Reserve Bank of New York.

King, Robert G., Charles I. Plosser, James H. Stock, and Mark W. Watson (1991). “Stochastic Trends and Economic Fluctuations.” *American Economic Review,* 81: 819-840.

Pakko, Michael R. (2002). "What Happens When the Technology Growth Trend Changes?: Transition Dynamics, Capital Growth and the 'New Economy'." *Review of Economic Dynamics*, vol. 5(2), pages 376-407.

Ramey, Valerie and Matthew D. Shapiro (1998). “Costly Capital Reallocation and the Effects of Government Spending.” *Carnegie-Rochester Conference Series on Public Policy,* vol. 48, pp. 145-194.

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.

Table 1: Bai-Perron tests for structural change in business-sector labor productivity growth

Test	**Statistic**	**Break Date(s)**
UDmax	10.19**	
WDmax	11.71**	
$SupF_T(1)$	7.44	1973:2
$SupF_T(2)$	10.19**	1973:2, 1997:2
$SupF_T(3)$	6.99	1973:2, 1982:4, 1997:2
$SupF_T(4)$	6.90*	1955:2, 1961:1, 1966:2, 1997:2
$SupF_T(2\|1)$	17.01***	
$SupF_T(3\|2)$	2.06	
$SupF_T(4\|3)$	1.60	

****, **, *, significant at 1%, 5%, and 10% respectively.*

Note: All tests generated by Bai and Perron's (1998) Gauss code allowing for heteroskedasticity- and autocorrelation-robust covariance matrix, AR pre-whitening, and heterogenous variance-covariance matrices across subsamples. 10 percent of sample trimmed. Break dates shown correspond to first date of new subsample. Sample 1950:2-2004:2.

Table 2: Regressions with Changing Means: Business Economy

	Labor Productivity Growth (Percent, annual rate)			Log Hours Per Capita		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.48 (0.23)	3.17 (0.34)	3.22 (0.28)	0.11 (0.01)	0.15 (.014)	0.14 (.012)
Dummy: 1973:2-1997:1		-1.70 (0.47)	-1.74 (0.43)		-.075 (.017)	-0.070 (0.015)
Dummy: Post-1997:1		0.19 (0.49)			-.024 (.024)	
R^2		0.06	0.06	--	0.41	0.39

Note: Standard errors are heteroskedastic and autocorrelation-robust (estimated via GMM in TSP with NMA=8). Estimated with growth in private business labor productivity from 1947:2 to 2004:2.

Table 3: Encompassing Results

	Data Generating Process	% Positive (6 quarter sum), No Breaks	% Negative (6 quarter sum), Using Estimated Breaks	% with maximum F-statistic > 6.83	% satisfying (1) and (2)	% satisfying (1), (2), and (3)
		(1)	**(2)**	**(3)**	**(4)**	**(5)**
1.A	**No Breaks, Bivariate**	97	25	8	22	2
1.B	**No Breaks, 4-Variable**	85	28	11	21	3
2.A	**Breaks, Bivariate**	90	82	88	73	64
2.B	**Breaks, 4-Variable**	69	77	90	50	45

Rows 1.A and 1.B use synthetic data generated under the null that there are no breaks in labor productivity. Rows 2.A and 2.B are generated under the null that there are labor-productivity breaks in 1973:2 and 1997:2, with break magnitudes estimated in the VAR equation for labor productivity. For each synthetic dataset, I (i) calculated the maximum F statistic (with associated dates) for two breaks; (ii) estimated the SVAR without making allowances for any breaks; and (iii) estimated the SVAR with the estimated break dates in the VAR. Columns (1) and (2) use the average estimated response over the first six quarters for each synthetic dataset. For each DGP, column (1) shows the fraction for the datasets in which the response is positive when no breaks are allowed; column (2) show the fraction that are negative when one uses the estimated break dates. Column (3) shows the fraction of the datasets in which the F statistic for two breaks exceeds 6.83. Column (4) shows the fraction of the datasets that have a positive response when there are no breaks allowed <u>and</u> have a negative response when two estimated breaks are used. Column (5) shows the fraction that satisfy the restrictions in Column (4) (positive response when there are no breaks allowed and negative response when breaks are used) <u>and</u> have an F statistic greater than 6.83. Sample period is 1950:2-2004:2.

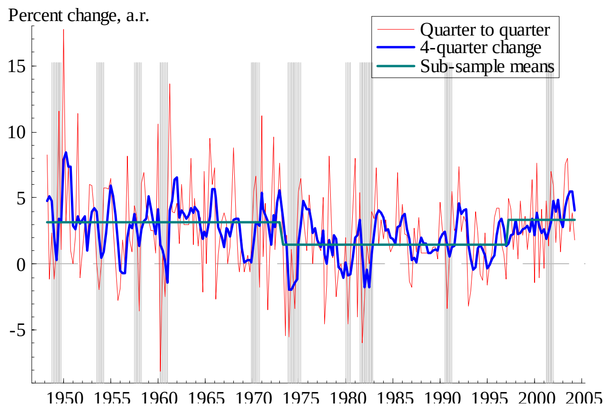
Table 4: Robustness Across Specifications:
Correlations, Impact Effect, and Granger Causality Tests

	Specification:	Correlation of technology residuals: 2 variable, level, breaks	4 variable, level, breaks	BFK Annual Residuals	Impact Effect on Hours	Granger Causality of Money Shocks (p values)
1.	2-variable, level, breaks	1	0.85	0.50	-0.43	0.87
2.	4-variable, level, breaks	0.85	1	0.48	-0.45	0.36
3.	6-variable, level, breaks	0.84	0.87	0.30	-0.40	0.10
4.	2-variable, quadratic trend, no breaks	0.97	0.81	0.48	-0.40	0.96
5.	4-variable, quadratic trend, no breaks	0.90	0.90	0.45	-0.35	0.73
6.	6-variable, quadratic trend, no breaks	0.87	0.83	0.41	-0.34	0.35
7.	2-variable, non-farm business, level, breaks	0.85	0.74	0.47	-0.45	0.40
8.	4-variable, non-farm business, level, breaks	0.49	0.79	0.32	-0.29	0.22
9.	2-variable, Francis-Ramey level, no breaks	0.90	0.72	0.53	-0.21	0.82
10.	4-variable, Francis-Ramey level, no breaks	0.48	0.25	0.33	-0.06	0.90
11.	2-variable, Francis-Ramey level, breaks	0.99	0.86	0.49	-0.48	0.84
12.	4-variable, Francis-Ramey level, breaks	0.61	0.90	0.34	-0.36	0.51
13.	2-variable, difference, no breaks	0.92	0.75	0.54	-0.29	0.91
14.	2-variable, difference, breaks	0.93	0.73	0.38	-0.18	0.47
15.	2-variable, level, no breaks	0.48	0.28	0.18	0.28	0.02
16.	4-variable, level, no breaks	0.33	0.56	0.20	0.17	0.11
17.	6-variable, level, no breaks	0.24	0.30	0.38	0.10	0.00
18.	2-variable, non-farm business, level, no breaks	0.60	0.42	0.34	0.11	0.15
19.	4-variable, non-farm business, level, no breaks	0.25	0.59	0.22	0.04	0.10

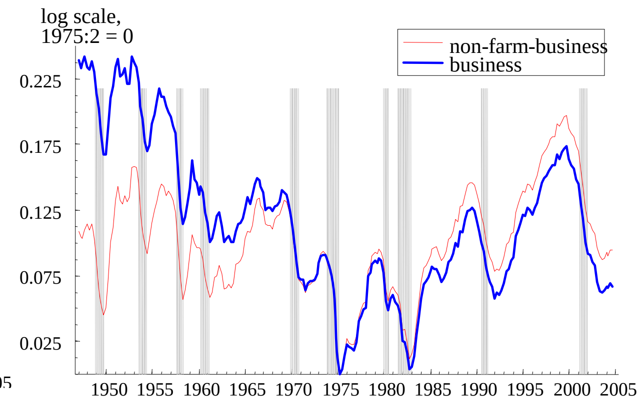
Notes: 2-variables are productivity growth and hours; 4-variables add the log of the consumption-output and investment-output ratios; 6-variables add the fed funds rate and the growth in the GDP deflator. Levels versus differences refers to whether hours per person enter in log-levels or log-differences. Breaks refers to whether subsample means are removed from productivity growth prior to estimation. "Quadratic trend" removes quadratic trend from all variables prior to estimation. Francis-Ramey uses their (2005) demographically adjusted measure of the population available to work. For annual correlations (44 observations), 0.30 is significant at the 5 percent level and 0.25 is significant at the 10 percent level. Granger-causality test is the p-value from regressing residuals on four lags of the monetary shock variable (from an identified VAR) used by Basu, Fernald, and Kimball (2004).

Figure 1. Basic Data.

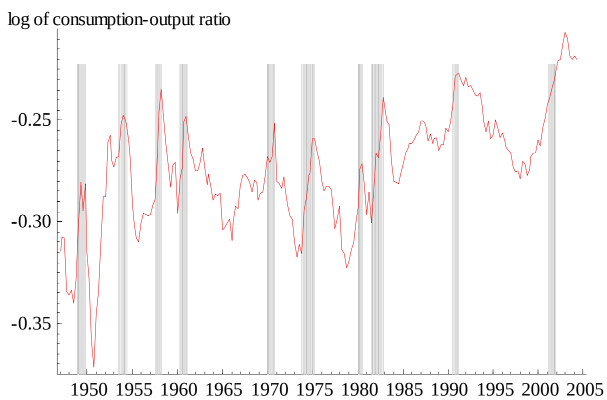
A. Labor productivity, business sector



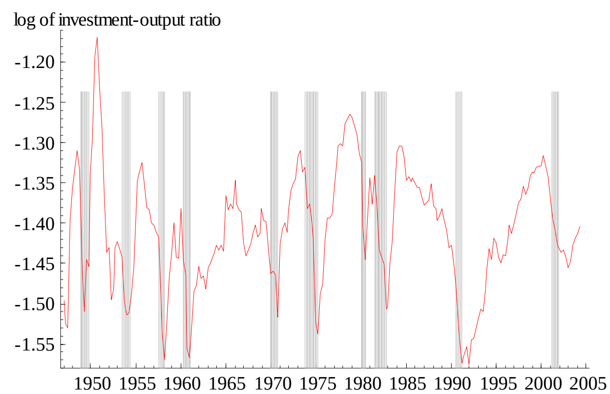
B. Hours worked per capita



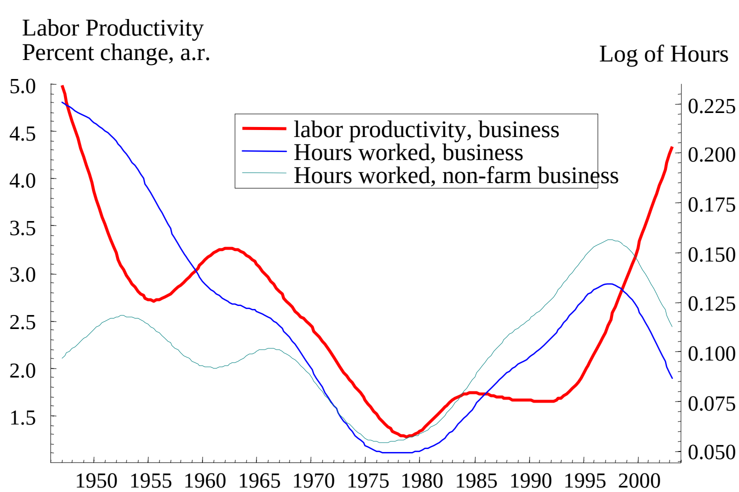
C. Consumption-Output Ratio



D. Investment-output ratio



E. Trends from HP-filter in labor productivity and hours per capita (lambda = 14,400, smoother than "usual" quarterly value of 1,600).

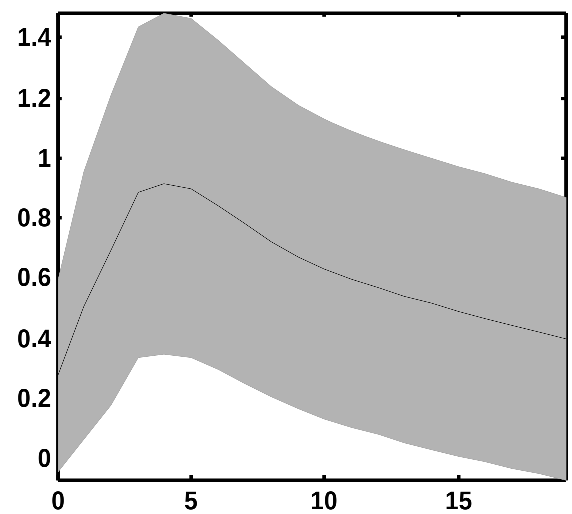


Sources: Bureau of Economic Analysis and Bureau of Labor Statistics.

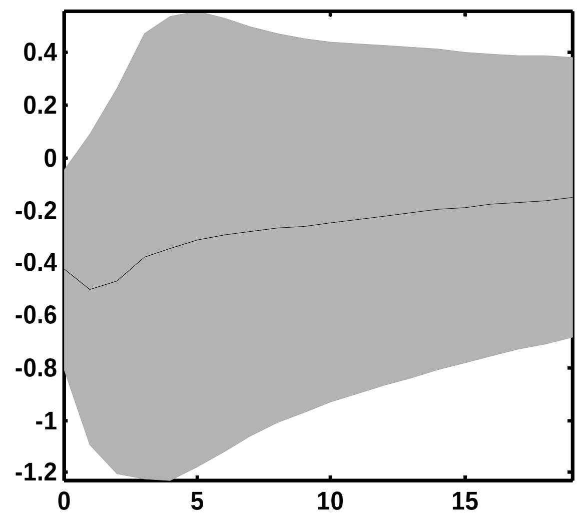
Figure 2. Impulse Responses from Bivariate Specification
(Log difference of labor productivity; log-level of hours per person 16 and older)

Response of Hours to a Technology Shock

A. No Trend Breaks

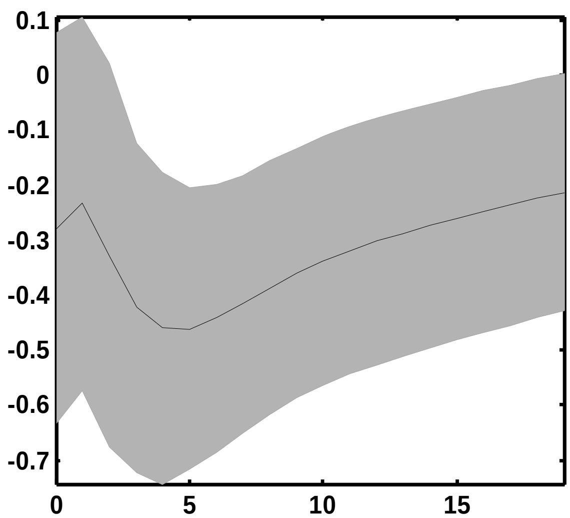


B. Pre-1973:2 and Post-1997:1 Break

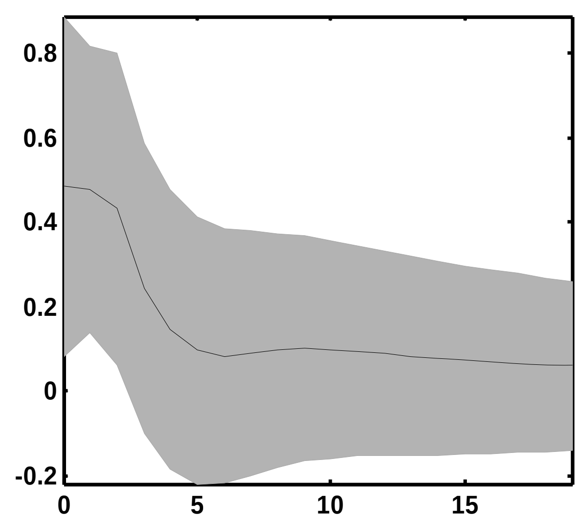


Response of Labor Productivity to a Non-Technology Shock

A. No Trend Breaks



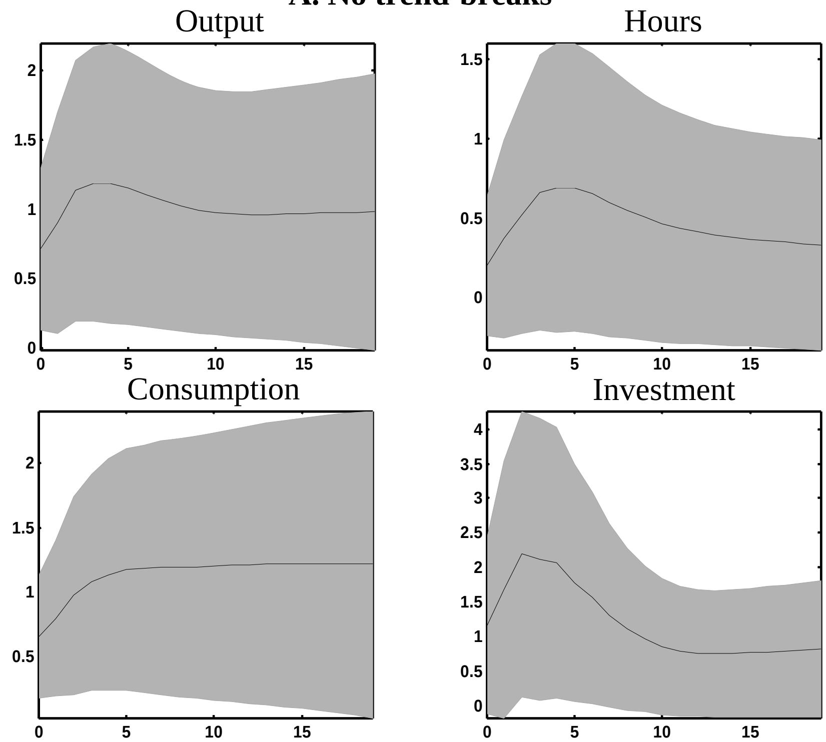
B. Pre-1973:2 and Post-1997:1 Break



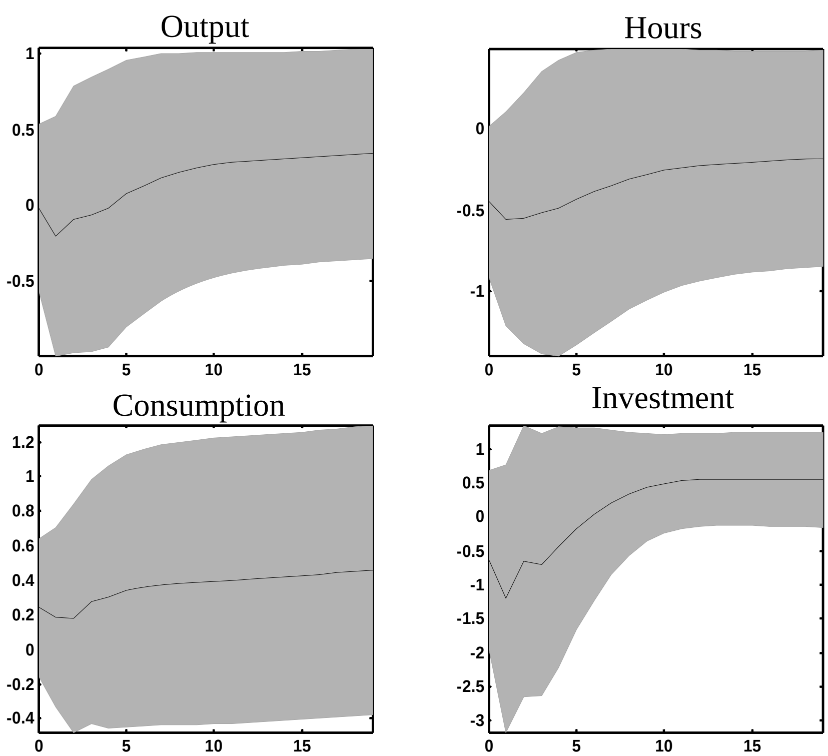
90 percent confidence interval in shaded region. "Break" estimates in right column remove subsample means from productivity growth before estimating SVAR. Sample period is 1951:2 – 2004:2.

Figure 3: Impulse responses from four variable system
(Labor productivity growth, log-level of hours, consumption-output, investment-output)

A. No trend breaks



B. Controlling for trend breaks in labor productivity (1973:2, 1997:2)



90 percent confidence interval in shaded region. Sample period is 1951:2 – 2004:2.

Figure 4. Sub-sample responses of hours worked to technology
(Four variable system)

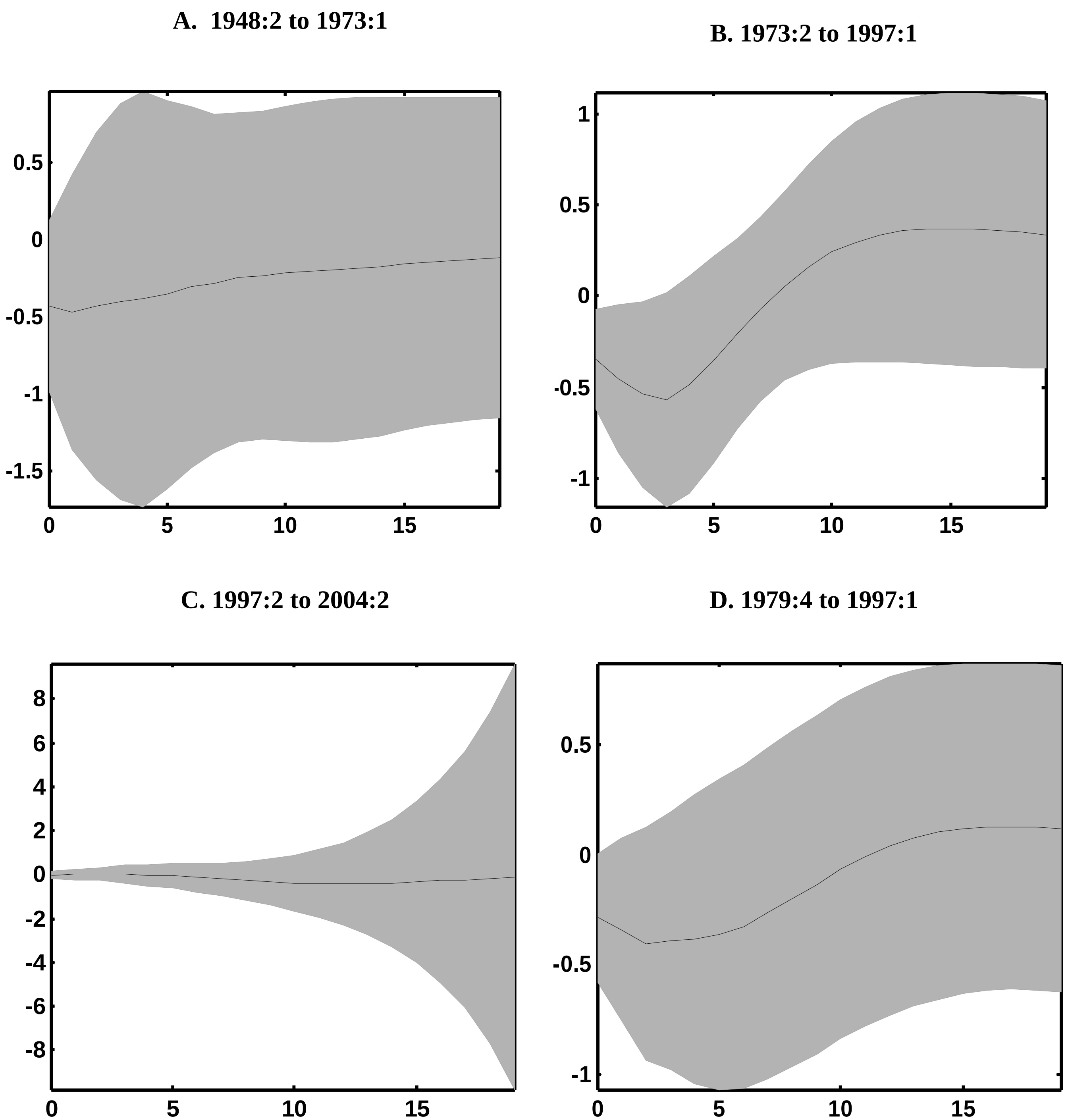
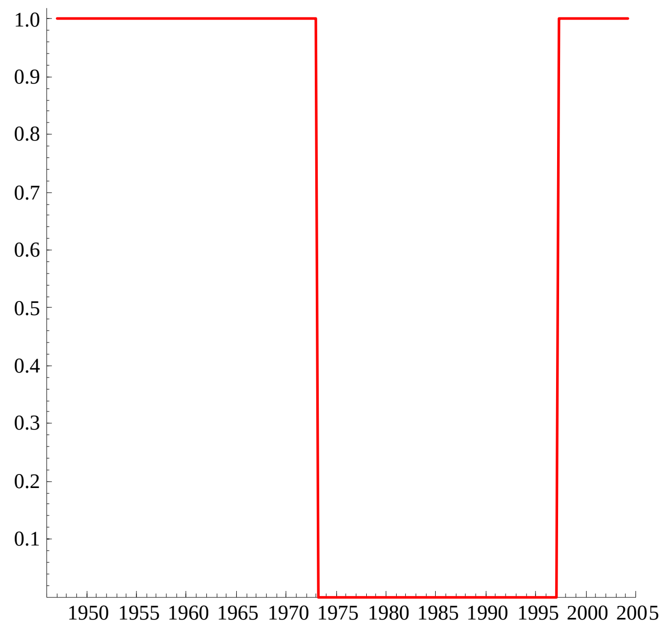


Figure 5. Diagnostic Example: Productivity Growth is a 1-0-1 Series



B. Hours response to a technology innovation

C. Productivity response to a non-technology innovation

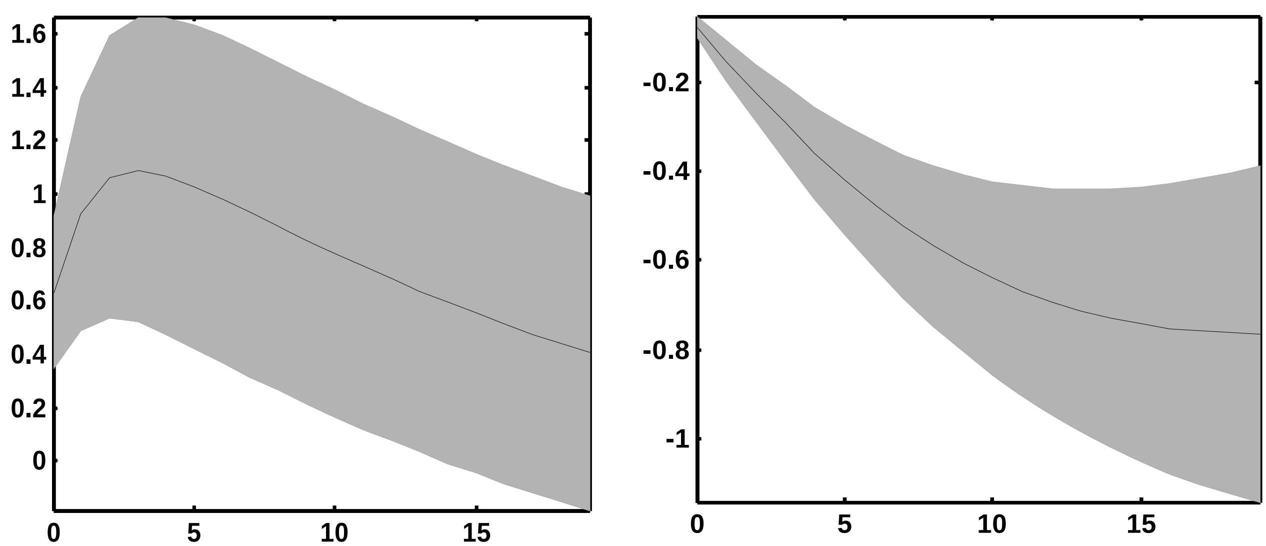
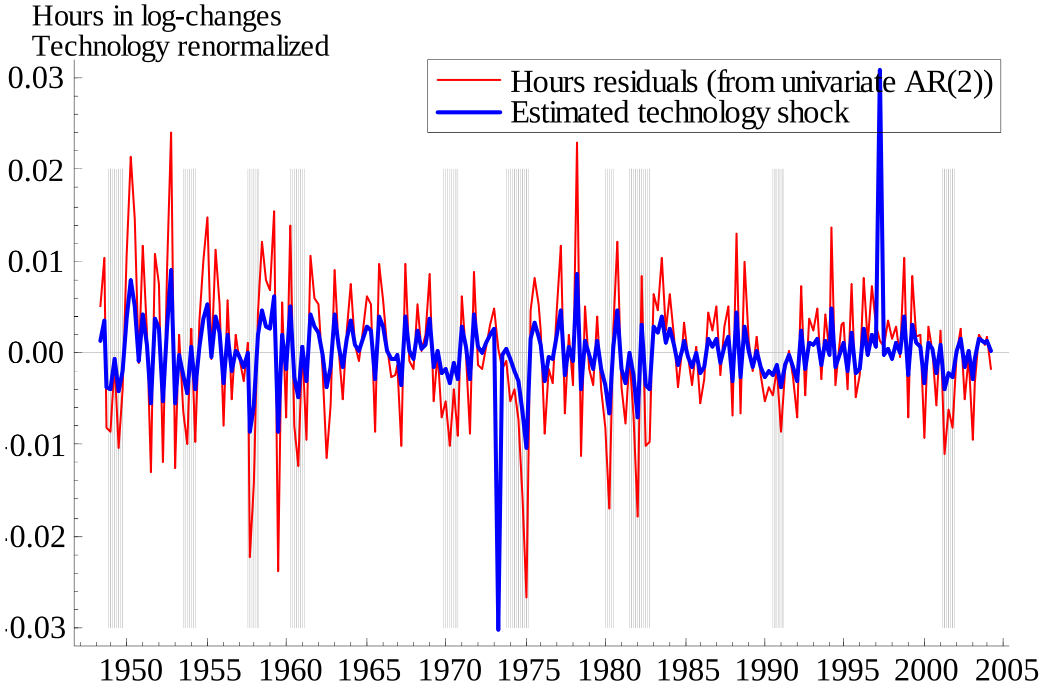


Figure 6. Diagnostic 1-0-1 Productivity Continued:
Estimated technology innovations and their contribution



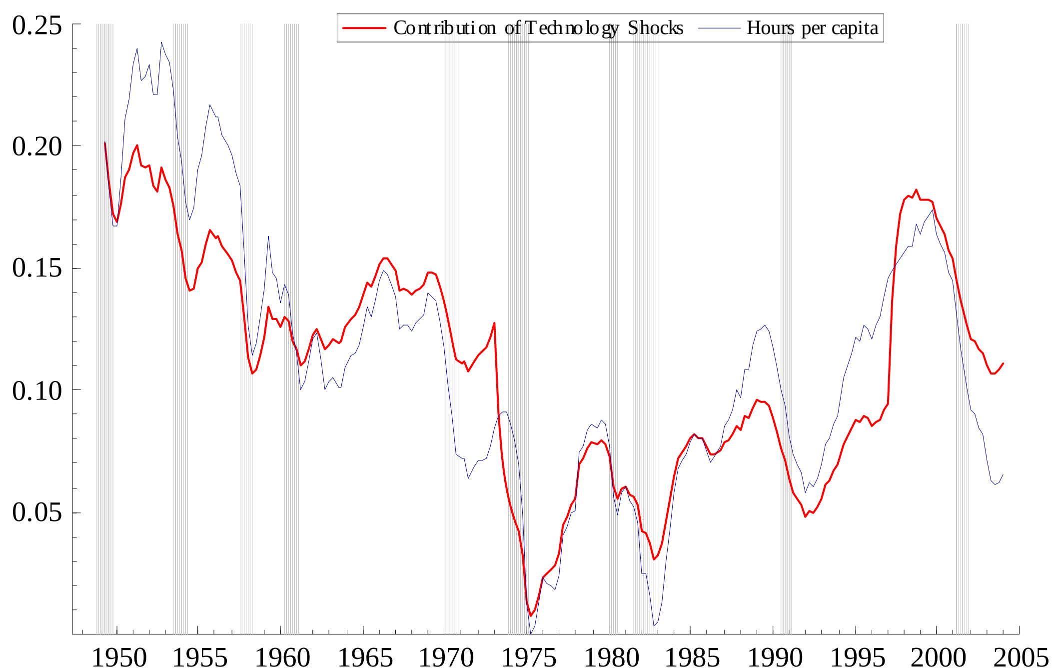
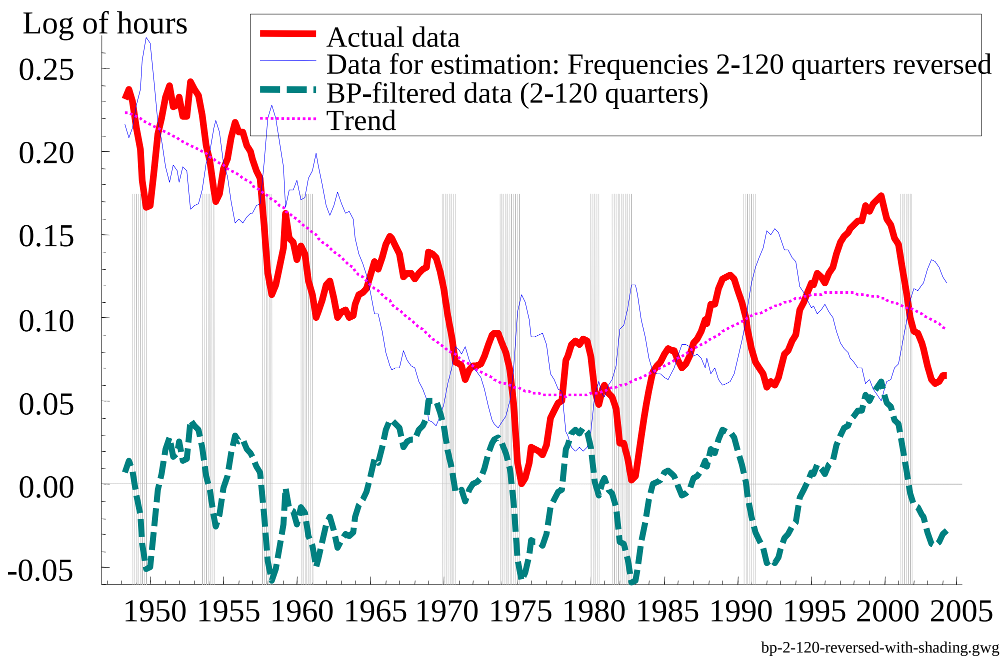
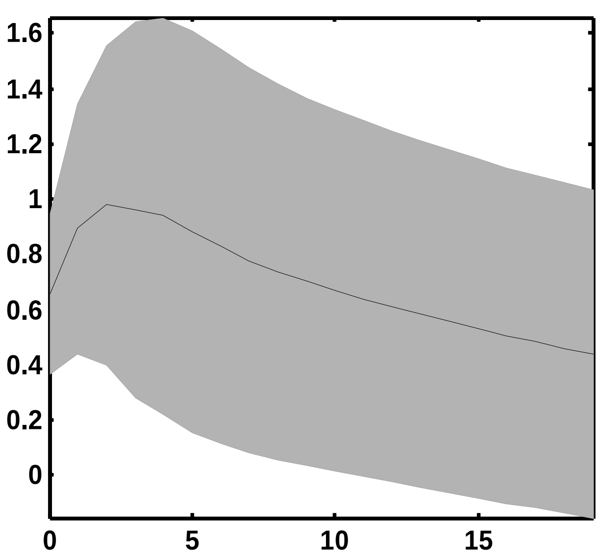


Figure 7. Diagnostic Example: Reversing Frequencies from 2-120 quarters in Hours

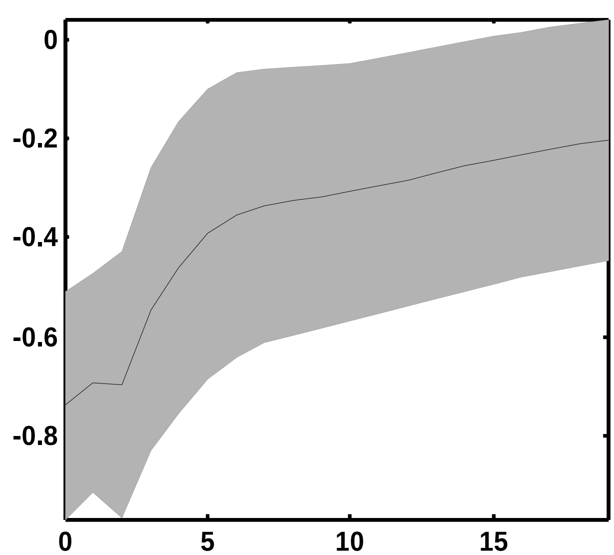
Hours data were bandpass-filtered; then frequencies 2-120 were removed (yielding trend) and then removed again



B. Hours response to a technology innovation



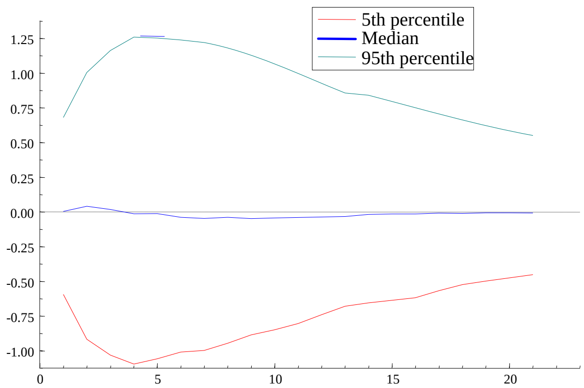
C. Productivity response to a non-technology innovation



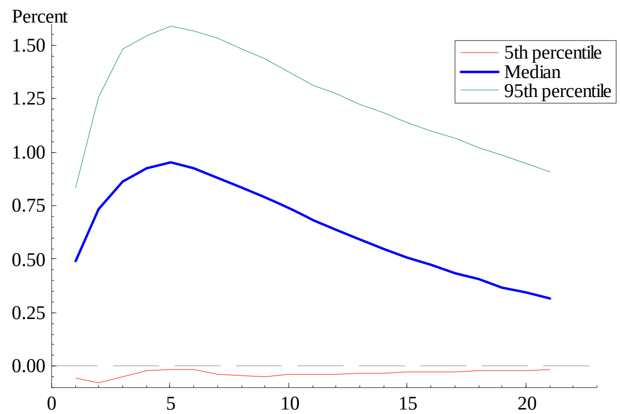
Figures show responses from bivariate SVAR with actual labor productivity; and with hours per capita data where frequencies from 2 to 120 quarters (using Christiano-Fitzgerald filter) have been reversed from actual data, as shown in top figure. Bottom panel show 90 percent confidence intervals.

Figure 8: Selecting on Series with Apparent (but Spurious) Breaks:
Hours responses to a technology improvement

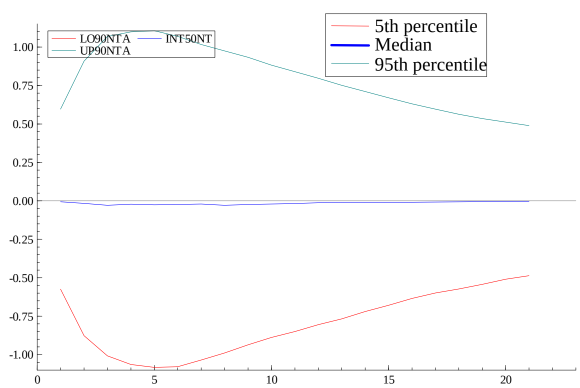
A. Estimates with random draws from two series



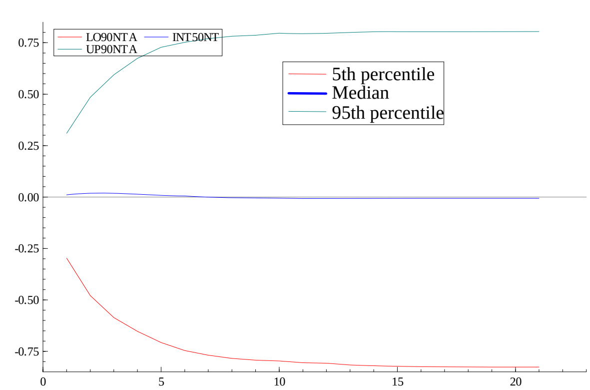
B. Estimates selecting on large t-statistics in both series: Levels Specification



C. Estimates selecting on large t-statistics in both series: Incorporating trend break into estimation



D. Estimates selecting on large t-statistics in both series: Difference specification



Notes: Figures show response of hours to a technology shock, using simulated data for a large number of artificial, independent series for productivity growth and hours. Panels B, C, and D select based on a t-test for apparently significantly higher mean pre-73 and post-97. See text for further details.