Available online at www.sciencedirect.com
ScienceDirect
Journal of Monetary Economics 54 (2007) 2467–2485
Journal of MONETARY ECONOMICS
www.elsevier.com/locate/jme

Trend breaks, long-run restrictions, and contractionary technology improvements ☆

John G. Fernald*

Federal Reserve Bank of San Francisco, USA

Received 3 August 2006; received in revised form 27 June 2007; accepted 28 June 2007
Available online 6 July 2007

Abstract

Structural vector autoregressions with long-run restrictions are extraordinarily sensitive to low-frequency correlations. Recent literature finds that the estimated effects of technology shocks are sensitive to how one treats hours per capita. However, after allowing for (statistically and economically significant) trend breaks in productivity, results are much less sensitive: hours fall when technology improves. The issue is that the common high–low–high pattern of productivity growth and hours (i.e., the low-frequency correlation) inevitably leads to a positive estimated response. The trend breaks control for this correlation. This example suggests a practical need for care in using long-run restrictions.

JEL classification: E24; E32; O47.

Keywords: Technology; Business cycles; Structural change; Long-run restrictions

☆I thank Shanthi Ramnath, David Kang, Stephanie Wang, Andrew McCallum, and David Thipphavong 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, Rob Vigfusson, John Williams, and several anonymous referees 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.

*Fax: +1 415 977 4099.
E-mail address: john.fernald@sf.frb.org

0304-3932/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmoneco.2007.06.031

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 (1999), 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 US productivity growth was exceptional and hours worked fell.

Francis and Ramey (2005a) confirm the robustness of Galí's results to several alternative specifications and suggest alternative interpretations. However, Christiano et al. (2003) challenge the Galí–Francis–Ramey findings. 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 seemingly reasonable alternative specification completely reverses results. Unfortunately, little clear intuition is available about what drives results with long-run restrictions.

This paper provides simple analytics and simulations that highlight 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. One thus needs to be sure that the low-frequency movements in the data reflect the economic phenomena that one seeks to identify.

Much of the literature focuses on whether hours worked rise or fall following a technology improvement. Christiano et al. argue for using hours per capita in levels and conclude that hours probably rise after a technology improvement. However, 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 break dates.[2]

The source of the sensitivity to breaks is the low-frequency correlation between labor productivity growth and the level of hours worked per person. Fig. 1 (panels A and B) shows the two series. Average productivity growth was notably faster before the early 1970s and after the mid-1990s. Hours show a similar high–low–high pattern.

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. A simple, analytically tractable example makes clear that the low-frequency correlation dominates the relevant covariances of the VAR: one almost cannot help but find a positive impulse response.

[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.

[2] If there is structural change in the data generating process, one generally needs to account for it econometrically; see, for example, Perron (1989). Section 2 discusses the economic phenomena that might drive the low-frequency variation in productivity growth and in hours. Section 5 discusses when, by removing breaks in productivity growth, one properly identifies the response to a shock to the *level* of technology.

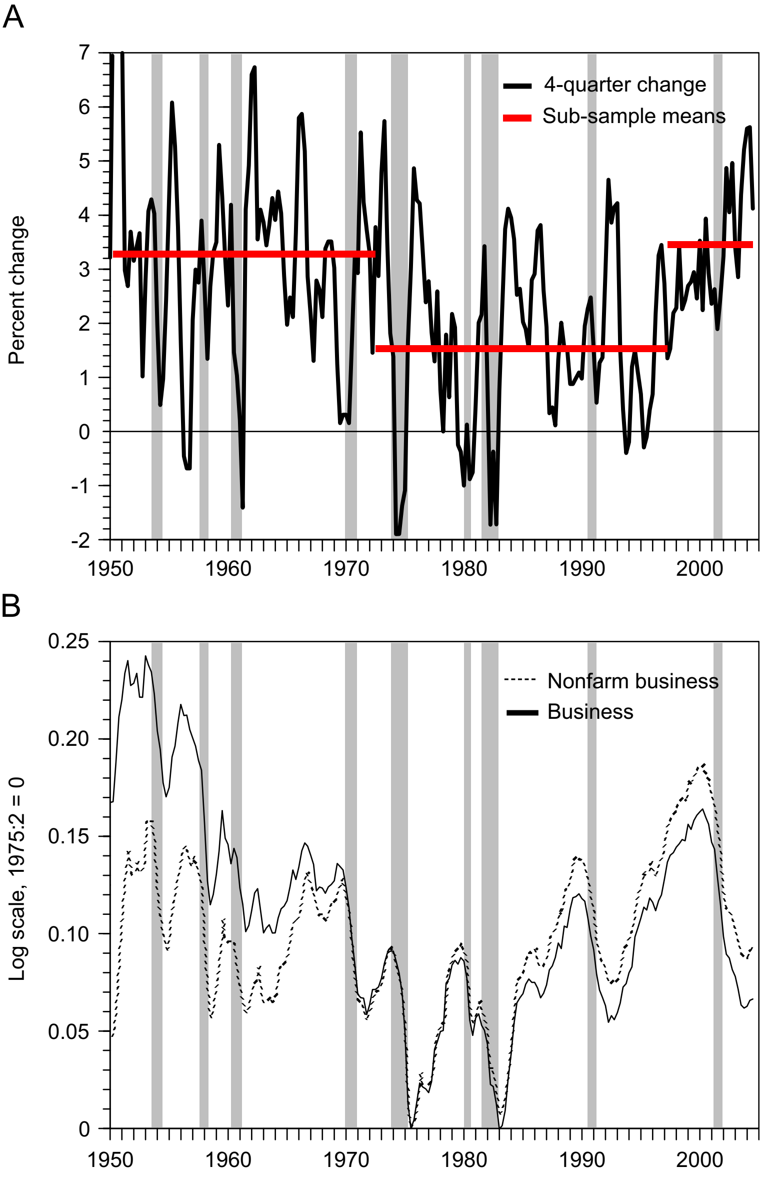

Fig. 1. Productivity and hours. (A) Labor productivity, business sector. (B) Hours worked per capita. *Source*: Bureau of Labor Statistics.

Simulations give further insight into what drives the VAR results and illustrates the empirical relevance of the analytics. First, suppose one uses actual hours per person but replaces 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, suppose one uses actual productivity but changes the high- and medium-frequency components of hours. In particular, using a bandpass filter to estimate and remove frequencies of 2–120 quarters in hours, one can measure the low-frequency trend. If one then adds back the filtered component with the sign *reversed*, one obtains a new series with the same low-frequency properties as actual hours; but which is the mirror image at high- and medium frequencies. This transformation has little effect on the estimated impulse responses, which again imply that hours rise when technology improves. Third, suppose one generates 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 would expect the levels specification to imply a positive response of hours to technology—regardless of the true business-cycle response.[3]

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. These include the original Blanchard and Quah (1989) estimates of how supply shocks affect unemployment, which switch sign when they remove subsample means of output growth before and after 1973. Similarly, Francis and Ramey (2005a) and Galí and Rabanal (2004) find that results are sensitive to removing a quadratic trend in hours. Francis and Ramey (2005b) find that the levels specification is sensitive to demographic adjustments for schooling, government employment, and changing demographics. These adjustments do not affect the business-cycle properties of hours per available person, but they substantially change the low-frequency properties.

Basu et al. (2006) use an augmented-growth accounting approach to identify technology shocks, using controls to distinguish technology change from the myriad short-run non-technological effects that affect the Solow residual (e.g., variations in fact or utilization).[4] As in this paper, they find that technology improvements reduce hours on impact. Section 5 considers a wide range of VAR specifications. Those that control in one way or another for low-frequency trends yield similar technology series to the BFK series, with similar impulse responses.

The technology shocks identified in this paper and in BFK are, conceptually, a weighted average of shocks to all sectors, including those that hit the production of investment goods. Fisher (2006) suggests decomposing aggregate technology shocks into their "neutral" and "investment specific" components. Consistent with the findings of this paper, Fisher finds that his results are sensitive to using subsamples that correspond to breaks in his series. In his preferred subsample results, the dynamic responses of hours to the two types of shocks are very similar—and are, in fact, similar to the results from the model that has a single, composite shock: especially prior to 1979, Fisher finds that hours

[3] If the low-frequency correlations in the finite sample do, in fact, reflect causal links in the (properly specified) DGP, then those responses might be accurate, if inevitable. The working paper version, as well as the appendix (Fernald, 2007, available online) applies the Christiano et al. (2003) encompassing methodology to distinguish between the breaks and no-breaks specification. Those results support the breaks-specification over the no-breaks specification. The online appendix also includes data description and an analysis of restrictions imposed on the dynamics of growth shocks if we simply remove breaks from labor productivity.

[4] Importantly, the BFK approach is robust to failures of the long-run restriction. For example, Barlevy (2004) presents a model and evidence that demand shocks might raise technology permanently. Sarte (1997) also questions the identifying assumption of no permanent effect of non-technological shocks on the level of labor productivity.

fall sharply after a technology improvement. Fisher's results thus suggest that, at least for the dynamics of hours, the composite-shock model is a reasonable approximation.

In addition to the empirical literature using long-run restrictions, Faust and Leeper (1997) highlight the theoretical limitations of long-run identification in finite data. Cooley and Dwyer (1998), Erceg et al. (2005), and Christiano et al. (2006) simulate various dynamic stochastic general equilibrium (DSGE) models to assess the sensitivity of long-run restriction results to particular economic environments.[5] As in this paper, the Monte Carlo exercises highlight that long-run restrictions must be used with care.

2. Evidence for trend breaks

Productivity growth slowed down after about 1973 and sped up again after the mid-1990s. 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.[6]

Is there statistical evidence of structural change? Consider the simple model $\Delta p_t = \mu_T + \varepsilon_t$, where Δp_t is productivity growth, μ_T is the mean during interval T, and ε_t is an innovation. Bai and Perron (1998, 2003) provide straightforward statistical tests for the null that μ_T is the same in all intervals versus the alternative that it changes one or more times. One strategy is to first test for an unknown number of breaks (changes in mean) and then to determine the number of breaks. The first two rows of Table 1 show "double-maximum" tests, UD-max and WD-max, which test the null of zero breaks against the alternative of an unknown number of breaks. These data are private-business labor productivity growth from 1950:2 to 2004:2. (The appendix to this paper—Fernald, 2007, available online—describes the data.) These tests reject the null of zero breaks at better than the 5% level. Thus, structural change is likely.

The statistical evidence then suggests two breaks, with a slowdown after 1973:1 and a speedup after 1997:1. Notably, one rejects the null of zero versus two breaks: the $\mathrm{SupF_T}$ (2) test (the maximum F test) of 10.19 is statistically significant at the 5% level.[7] And 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. Because mean growth is similar before 1973 and after 1997, one cannot statistically 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. Results that follow appear robust to alternative dates around the selected (maximum F) dates.

[5]EGG generate artificial data from a range of calibrated DSGE models and then estimate SVARs. They conclude that long-run restrictions might help discriminate between alternative models—e.g., the estimated impulse response for hours, though biased, is unlikely to yield an incorrect sign. More recent work by Chari et al. (2005) undertakes an exercise similar to EGG and emphasizes that that results might be biased. Christiano et al. (2006), however, argue that in empirically relevant models, SVARs have little bias and are informative.

[6]See, for example, David (1990) and Fernald (1999).

[7]Bai and Perron (1998, 2003) Gauss code was downloaded from http://qed.econ.queensu.ca/jae/2003-v18.1/bai-perron/ on August 16, 2004. Tests are robust to heteroskedasticity and autocorrelation, allow the variance–covariance matrix to differ across regimes, and implement AR prewhitening. To check significance levels, the four-variable VAR discussed later (setting constant terms and initial values to zero) was bootstrapped 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% level.

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^{**}	
$\text{SupF}_{\text{T}}(1)$	7.44	1973:2
$\text{SupF}_{\text{T}}(2)$	10.19^{**}	1973:2, 1997:2
$\text{SupF}_{\text{T}}(3)$	6.99	1973:2, 1982:4, 1997:2
$\text{SupF}_{\text{T}}(4)$	6.90^{*}	1955:2, 1961:1, 1966:2, 1997:2
$\text{SupF}_{\text{T}}(2\|1)$	17.01^{***}	
$\text{SupF}_{\text{T}}(3\|2)$	2.06	
$\text{SupF}_{\text{T}}(4\|3)$	1.60	

Note: All tests generated by Bai and Perron (1998)'s Gauss code allowing for heteroskedasticity- and autocorrelation–robust covariance matrix, AR pre-whitening, and heterogenous variance–covariance matrices across subsamples. 10% of sample trimmed. Break dates shown correspond to first date of new subsample. Sample 1950:2–2004:2.
***, ** and * significant at 1%, 5% and 10%, respectively.

Empirically, allowing for occasional trend breaks is one of several ways to model persistent growth shocks. First, one can view them as regime shifts. Kahn and Rich (2006), for example, find a switch from a high- to a low-productivity growth regime in the early 1970s, with a switch back in the late 1990s. Second, one can model productivity *growth* as having a stochastic trend, so productivity itself is $I(2)$. Roberts (2001), for example, finds economically significant variation in trend productivity growth. Overall, the post-war period appears to have few regime switches or low-frequency swings in growth, so the trend-break approach tells a similar story to these two alternatives. In addition, it is easier to apply in the SVAR context and is relatively transparent. All three approaches argue for relaxing the restriction of constant mean productivity growth.

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 (see Fig. 1B). Francis and Ramey (2005b) argue that the low-frequency movements reflect trends in school enrollment, government employment (since the numerator is private hours), and demographics—those over 65 are much less likely to work. These forces that cause the U-shaped pattern in hours are likely to be very different from those that cause business-cycle fluctuations. Since standard RBC models aim to explain fluctuations at business-cycle frequencies—rather than explaining school decisions, the size of government, or demographics—Francis and Ramey argue for removing these low-frequency influences from hours (leaving the business-cycle properties unaffected) prior to estimation.

In addition, there is no obvious reason to presume that the common high–low–high pattern of productivity and hours reflects causal links rather than the chance correlation of three data points. For example, Fernald (1999) argues that building the interstate highway system raised productivity growth in the 1950s and 1960s. It is not clear that one should link the factors driving the low-frequency movements in hours (e.g., schooling or demographics) to interstate highways.

The results that follow use population aged 16 and older with no demographic adjustment. This allows comparability with Christiano et al. (2003) and highlights the analytical and practical issues. Section 5 shows that the Francis–Ramey measure yields results similar to removing trend breaks from labor productivity and then using the CEV hours. Thus, the analysis in this paper makes clear why results are sensitive to adjustments such as Francis and Ramey's that affect low-frequency trends.

3. Sensitivity to trend breaks

This section now discusses the sensitivity of empirical results of VARs identified with long-run restrictions to controlling for breaks in labor productivity.

3.1. Structural VARs with long-run restrictions

Let X_t be a vector of variables with moving average representation $X_t = C(L)\varepsilon_t$. $C(L)$ is a matrix of lag polynomials and ε_t is a vector of innovations. For concreteness, consider the bivariate system where X_t comprises productivity growth, Δp_t, and the level of hours (or some other stationary transformation of hours), n_t. The identification assumption is that only technology shocks permanently affect of the level of labor productivity. Other shocks (such as monetary policy shocks or transitory technology shocks) can have only a short-run effect on labor productivity.

Following Shapiro and Watson (1988), one can impose the long-run restrictions by estimating the following regressions (constant terms not shown):

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

$$n_t = \sum_{i=1}^{q} b_{P,i}\Delta p_{t-i} + \sum_{i=1}^{q} b_{N,i}n_{t-i} + \mathrm{d}\varepsilon_t^Z + \varepsilon_t^N \tag{2}$$

Hours, n_t, enters (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 $q-1$. Since technology shocks might affect current hours growth, one estimates (1) with the standard instruments: lags of productivity growth, Δp_{t-s}, and of the level of hours n_{t-s}, where $s = 1$ to q. For the just-identified case, these instruments yield results identical to the Blanchard–Quah matrix methods. The residuals from (1) are the estimated technology shocks. One then estimates Eq. (2) with OLS by adding the contemporaneous estimated technology shock to the standard VAR equation.

With two variables, one identifies two shocks. $\widehat{\varepsilon_t^Z}$ is the identified technology shock. ε_t^N, in principle, captures all shocks (especially demand shocks) with at most a transitory effect on labor productivity. (Blanchard and Quah (1989) and Faust and Leeper (1997), discuss when bivariate systems can adequately capture the dynamics when there are more than two underlying shocks.) The specification generalizes easily. Larger VAR systems are straightforward: other variables are treated symmetrically with hours. There are then additional "non-technological" shocks, but without further assumptions these are not separately identified. In the difference specification, one can re-interpret n_t and Δn_t as the first and second difference of log-hours.

The empirical work below follows Christiano et al. (2003), Galí and Rabanal (2004), and Francis and Ramey (2005a) and sets $q = 4$. Impulse responses come from simulating the dynamic responses of the estimated system. The figure shows centered 90% confidence intervals.[8]

The Shapiro–Watson IV representation makes clear that long-run identification assumes that lagged hours (the instrument) are orthogonal to technology shocks. This assumption seems reasonable but will not hold exactly in any given sample. Given the common high–low–high pattern in productivity growth and hours in the actual post-war US data, orthogonality 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.

Following Christiano et al. (2003), estimation uses labor productivity and hours worked per capita in the private business sector, from the Bureau of Labor Statistics. All regressions are run from 1951:2 to 2004:2. (Fernald, 2007, describes the data). To control for trend breaks, subsample means are removed from labor productivity before estimation (using the estimated 1973:2 and 1997:2 break dates). This approach follows Blanchard and Quah (1989). In general, one must control for structural change or else the VAR is misspecified. Section 5 and the appendix (Fernald, 2007) consider when it is appropriate to simply remove subsample means.

3.2. Bivariate VAR results with long-run identification restrictions

Fig. 2 shows impulse responses from the bivariate SVAR above. The left column uses actual labor productivity growth and the level of hours per capita (aged 16 and above); the right column shows the "breaks specification," which removes subsample means from labor productivity growth prior to estimation. (Results are very similar with sample-period dummies in the VAR itself.)

Panel A reproduces the CEV result that technology improvements appear to raise hours worked on impact. In contrast, with trend breaks in Panel B, technology improvements sharply reduce hours. The break results are qualitatively similar to the difference and quadratic-detrended specifications in Galí and Rabanal (2004) and Francis and Ramey (2005a), and to the Basu et al. (2006) growth accounting.

The estimated response of productivity to a non-technology shock, in the bottom row, is also extremely sensitive to the trend breaks. Panel C shows that in the no-breaks specification, productivity falls sharply and significantly in response to a positive non-technology innovation. In contrast, with breaks removed, in Panel D, productivity rises sharply following a non-technological shock. The latter effect is consistent with utilization rising in response to demand shocks.

3.3. Larger VAR systems and subsample results

Christiano et al. (2003) report results from several larger systems, including a four-variable system that adds the log of the nominal consumption-output ratio and the nominal investment-output ratio to the VAR (they combine durable consumption with

[8]Rob Vigfusson graciously provided his code to calculate confidence intervals. They are calculated by estimating the VAR on bootstrapped simulated data and calculating standard deviations of the resulting impulse responses. Christiano et al. (2006) find that in a range of plausible, simulated DSGE models, this procedure has relatively good coverage rates.

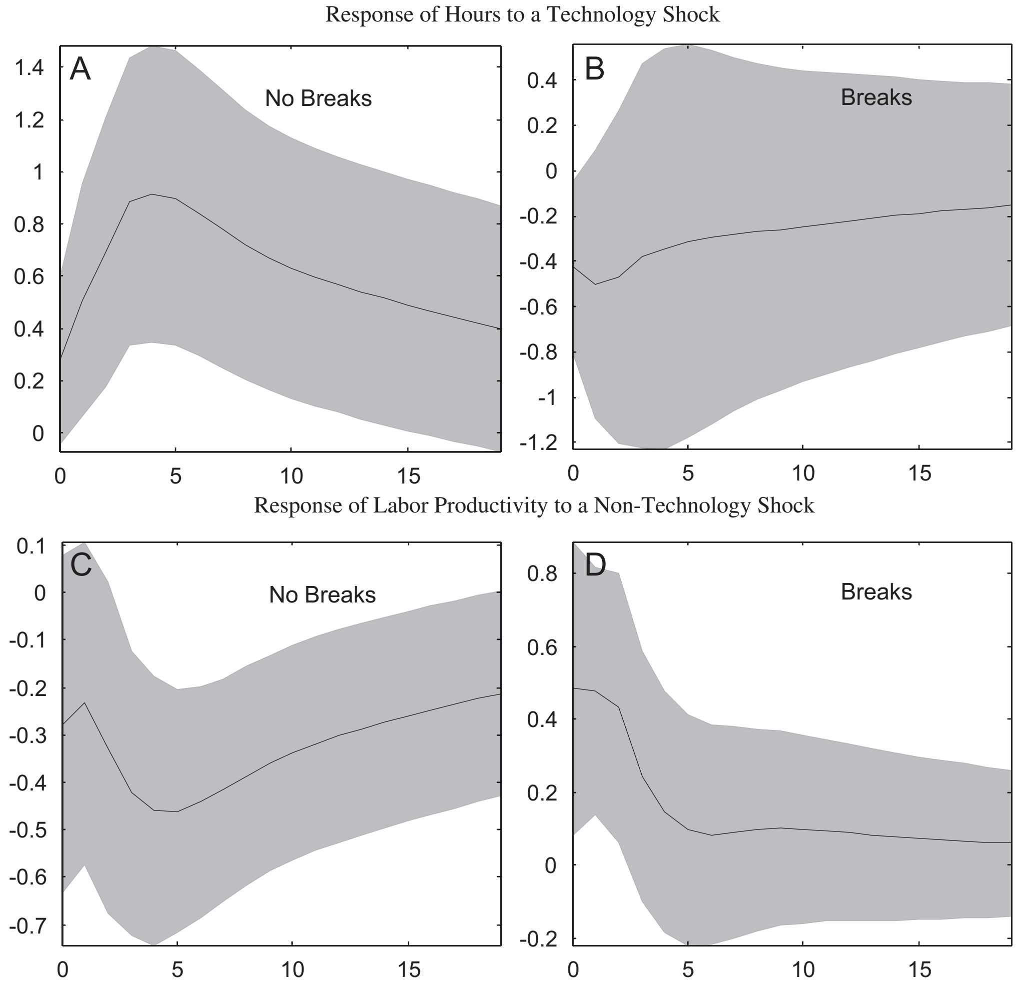

Fig. 2. Impulse responses from bivariate specification response of hours to a technology shock. (A) No trend breaks. (B) Pre-1973:2 and post-1997:1 break. (C) No trend breaks. (D) Pre-1973:2 and post-1997:1 break. *Notes*: Bivariate system includes log difference of labor productivity and log-level of hours per person aged 16 and older. Responses shown are to a one percent innovation. 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.

investment, and combine government with consumption of non-durables and services). Erceg et al. (2005) find that this system has reasonable properties. Christiano et al. (2003) also report results from a larger six-variable system that adds inflation (measured by the GDP deflator) and the fed funds rate. In the larger systems, the only shock that is identified is the technology shock.

As in the bivariate case, the four- and six-variable systems without breaks in labor productivity suggests that hours rise; but the breaks specification suggests that hours fall. Thus, the issues identified in the bivariate case carry over to larger systems.

An alternative way to control for structural change is to use subsamples—pre-1973:2, 1973:2–1997:1, post-1997:1. With the two- or four-variable systems, in all subsamples technology shocks reduce hours.[9] In the six-variable system, hours fall sharply pre-1973:2, but they are roughly flat in the 1973:2–1997:1 period. Adding inflation and the interest rate thus makes some difference for the post-1973:2 results, but the key point is that results are systematically different from the full-sample results without breaks. In particular, they never suggest substantial increases in hours.

Galí et al. (2003) emphasize subsample differences, which they link to presumed differences in monetary policy. The 1982:3–1997:1 period should incorporate the Volcker–Greenspan monetary policy while still excluding the productivity acceleration. Our four-variable system still suggests that hours decline but the six-variable results resemble those of Galí et al., in that hours change little on impact in this sample. Galí et al. include the interest rate and inflation in their system, explaining the similarity in results.

Fisher (2006) also considers a larger system that adds the real investment price to the Galí et al. regression, in order to decompose aggregate technology into a two components: neutral shocks and investment-specific ones. (Note that the approach in this paper and most of the literature, including Basu et al. (2006), seeks to identify a weighted sum of these two shocks.) In part because of time-varying trends in the data, Fisher focuses on subsample results. He also includes the interest rate and inflation in the VAR, and obtains results consistent with the composite-shock six-variable results in this paper as well as with Galí et al: prior to 1979, both shocks sharply reduce hours worked on impact; after 1982, both shocks have little impact effect on hours but, after a few quarters, the effect is noticeably positive. Hence, adding the investment price appears to make little difference to results. But Fisher's results support the view that results can be highly sensitive to low-frequency trends.[10]

4. How do low-frequency correlations affect results?

What drives the sensitivity of results to the treatment of low-frequency trends? This section discusses analytics and simulations that provide insight into results with long-run restrictions.

4.1. Analytical discussion

It is well known that the spectral density at frequency zero plays a central role in estimates with long-run restrictions; Christiano et al. (2006) discuss this point extensively, and suggest alternative estimators taking the data as given. Rather than take a frequency-domain approach, this section instead discusses relatively straightforward intuition for why low-frequency comovement dominates the key covariances of the estimation. For analytic tractability, the discussion focuses on bivariate systems. The empirical results are

[9] Erceg et al. (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.

[10] Canova et al. (2007) find different labor-market responses to neutral and investment-specific shocks in a system that accounts for job flows. They discuss sensitivity to low-frequency correlation at length.

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*).

Consider the following simplified, non-dynamic system; the estimated technology residuals and impact effects are similar to the full 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}. Constants are suppressed:

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

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

The IV estimate of $\hat{a}$ is $n'_{-1}\Delta p / n'_{-1}\Delta n$, where the vector notation is obvious. In the data, the denominator is negative: if lagged hours are high, current *growth* in hours tends to be low. At the same time, the common high–low–high pattern of hours and productivity growth pushes the sample covariance $n'_{-1}\Delta p$ (as well as the contemporaneous $n'\Delta p$) to be positive—hours tends to be high when productivity is high, towards the beginning and end of the sample. Hence, the low-frequency correlation is likely to lead to a negative estimate of $\hat{a}$. Note that a negative coefficient for $\hat{a}$ implies that a positive non-technology shock (which pushes Δn_t up) also pushes productivity down, consistent with the empirical impulse responses.

$\hat{d}$ gives the estimated impact effect of technology on hours and equals $n'\widehat{\varepsilon^{Z,S}} \Big/ \widehat{\varepsilon^{Z,S}}'\widehat{\varepsilon^{Z,S}}$. The denominator is positive. Hence, the sign depends on the numerator, the covariance of estimated technology with current hours:

$$n'\widehat{\varepsilon^{Z,S}} = n'(\Delta p - \hat{a}\Delta n) = n'\Delta p - \left(\frac{n'_{-1}\Delta p}{n'_{-1}\Delta n}\right) n'\Delta n \tag{5}$$

In the data, $n'\Delta n$ is positive, in contrast to the negative $n'_{-1}\Delta n$. (These are, of course, the expected signs for a stationary autoregressive time series.). As mentioned, the low-frequency correlation tends to imply that $n'_{-1}\Delta p$ and $n'\Delta p$ will be positive. Hence, both terms in Eq. (5) are positive. Thus, the assumed orthogonality between technology growth and hours implies that the estimated impulse response of hours to a technology shock is likely to be positive.

Several examples illustrate the impact of low-frequency correlation and improve our intuition for what features of the data drive the SVAR results. They show that the insights from the simple static example extend to more complicated dynamic estimation.

4.2. Simulation 1: 1–0–1 productivity growth series

Consider an extreme case where productivity growth equals 1 before 1973:2 and again after 1997:1 and zero between these dates; this is the thick line in Fig. 3C. What does the VAR imply if one runs the bivariate VAR with this 1–0–1 productivity series and the actual level of hours?

The impulse responses in Fig. 3A and B look qualitatively (and, for hours, quantitatively) like the no-breaks responses shown earlier in Fig. 2A and C. A positive

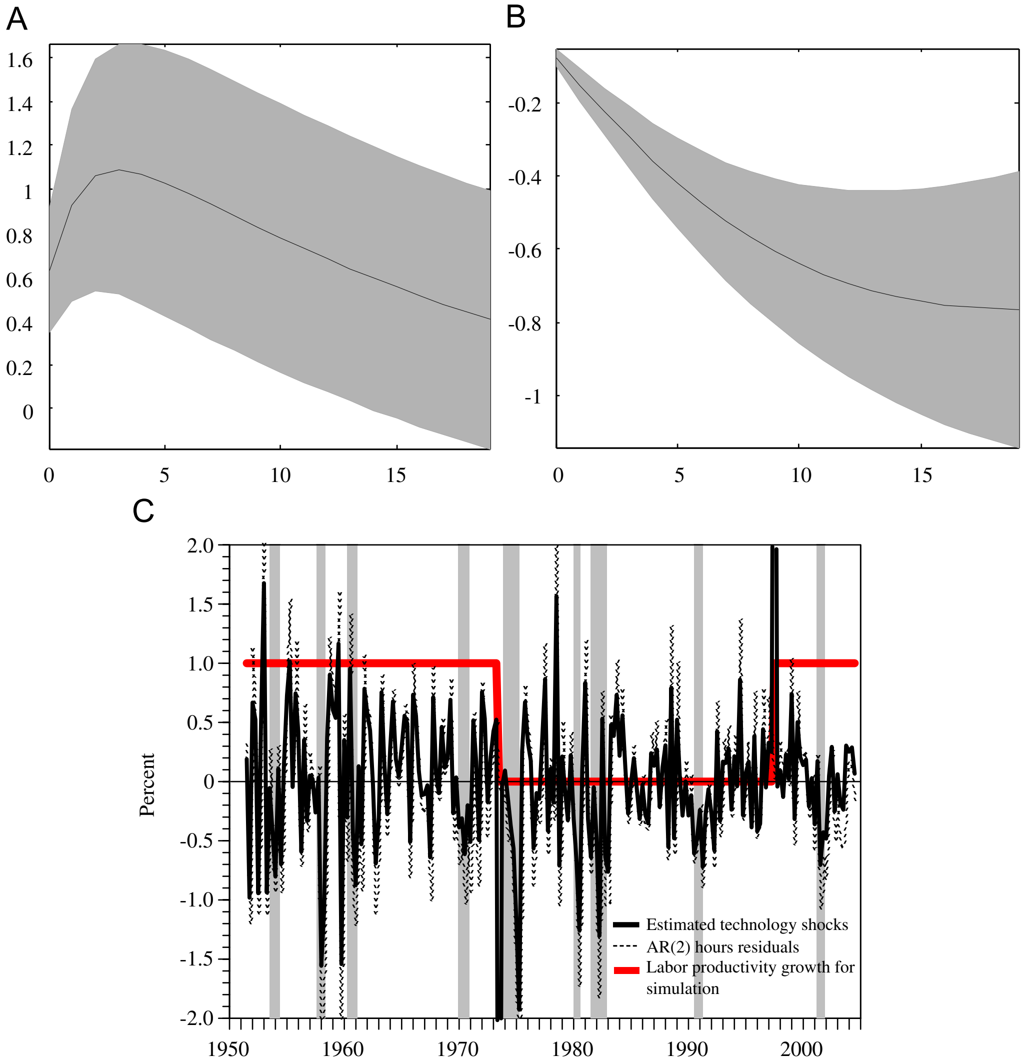

Fig. 3. Diagnostic example: productivity growth is a 1–0–1 series. (A) Hours response to a technology innovation. (B) Productivity response to a non-technology innovation. (C) Estimated technology shocks. *Notes*: Responses in Panels A and B are from a bivariate SVAR with actual hours per capita but where labor productivity growth is a 1–0–1 series as shown by the solid line in Panel C. Responses are to an estimated one percent innovation. 90 percent confidence intervals in shaded region. In Panel C, estimated technology shocks have been rescaled to correspond to the scale of innovations to hours. Sample period is 1951:2–2004:2.

technology shock raises hours strongly and statistically significantly; and a positive non-technology shock reduces productivity. (Productivity growth is highly persistent here, explaining the persistence of this latter response.) Both responses match the predictions from the simple, static analytical framework.

The bottom panel 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 sub-periods. Mechanically, because of the low-frequency correlation between hours and productivity, the regression puts a negative coefficient on hours-growth in the productivity equation, as discussed above. But now suppose hours increase temporarily for any reason (e.g., because of a labor-supply or demand shock). The negative coefficient on hours means that the positive hours blip reduces the fitted value of productivity. Since productivity itself changes at only two discrete points, one needs a positive innovation to technology to offset that. Thus, anything that causes hours to move causes 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. One can see this by modifying the hours data at high and medium frequencies and looking at how responses change. In particular, the long-dashed line that fluctuates around zero in Fig. 4A shows the components of hours at frequencies of 2–120 quarters (estimated with the Christiano and Fitzgerald, 2003, bandpass filter). The dashed line shows the estimated hours trend, defined as actual hours (the thick line) minus the high- and medium-frequency component. The thin line *reverses* high- and medium-frequency components. That is, it takes the trend line and adds back the filtered component with the sign reversed.

The bottom panels show the estimated impulse responses when one uses hours with the high- and medium-frequency components reversed (the thin line). The responses again look qualitatively and quantitatively like those in Fig. 2A and C. Hours worked rise strongly in response to a positive technology shock; a positive non-technology shock reduces productivity. Both responses are statistically significant. Thus, the estimated responses are largely invariant to reversing the frequency components below 120 quarters, making it clear that low-frequency movements drive the responses.

4.4. *Simulation 3: selecting on series with apparent breaks*

In the analytics, the true data generating process is not the issue. Rather, in the realized data sample, there is a low-frequency correlation between the series. To highlight this point, one can simulate 230 quarters of data for two series from the following univariate DGPs: $\mathrm{d}p_t = \varepsilon_t^Z$ and $n_t = 1.51n_{t-1} - 0.53n_{t-2} + \varepsilon_t^N$. Hours are a highly persistent AR(2) process; productivity growth is white noise. The simulated disturbance terms were normally distributed (0,1). As one would hope and expect, the median impulse response from estimating the SVAR on 1000 pairs of simulated data is close to the true value of zero.

In the data, a 1973:2–1997:1 dummy has a negative coefficient and large t-statistic for both hours and productivity growth. Purely by chance, some of the simulated univariate processes for hours and labor productivity show a similar pattern. For both hours and labor productivity, suppose one saves the first 2500 series where the dummy is negative and has a t-statistic as large in magnitude as 3.5, and then pairs up the series.

By construction, the series are not related. But, for the pairs of series that appear to have large common breaks, the estimated impulse responses are almost uniformly positive. Econometrically, one has selected series in which, by chance, true technology

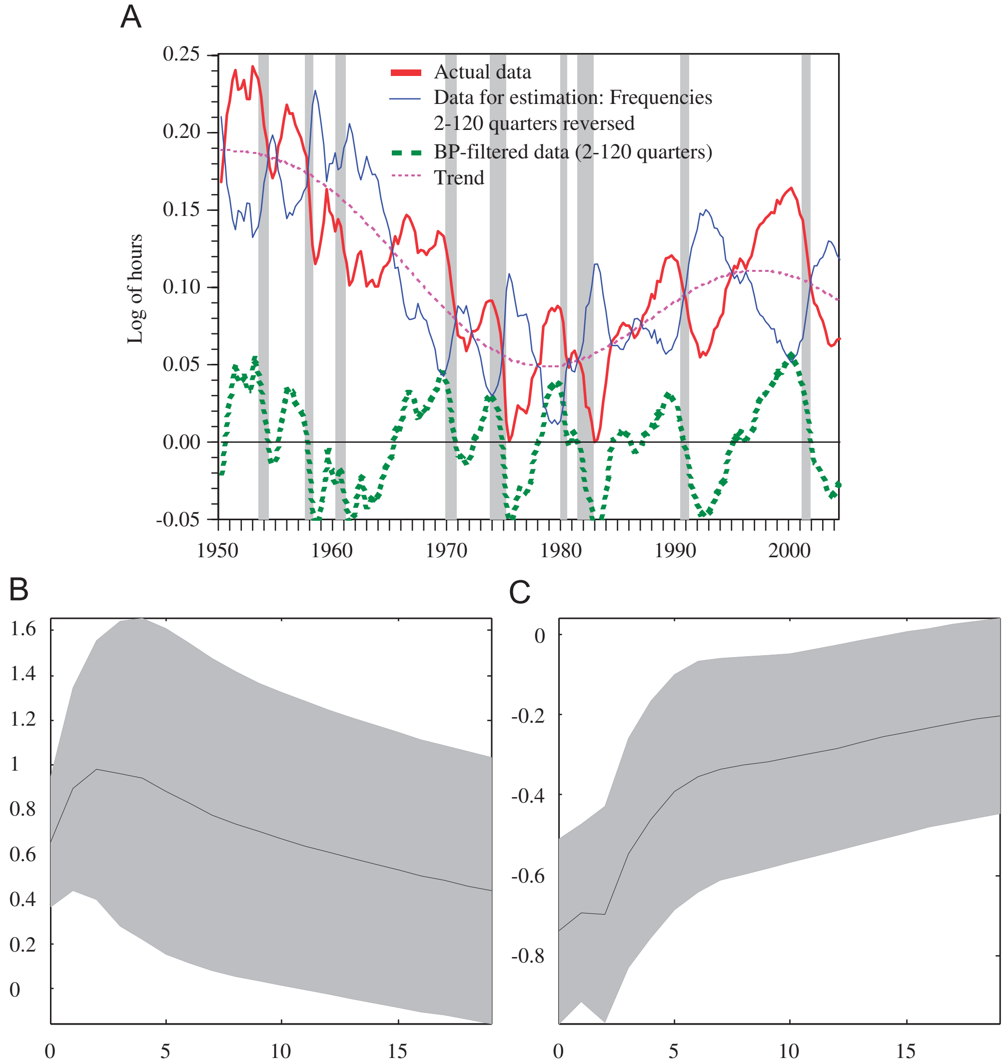

Fig. 4. Diagnostic example: reversing frequencies from 2 to 120 quarters in hours. (A) Hours per capita and the bandpass filter. (B) Hours response to a technology innovation. (C) Productivity response to a non-technology innovation. *Notes*: Figure A shows actual hours per capita, its high- and medium frequency component (2–120 quarters, estimated with Christiano-Fitzgerald filter), and the fitted trend. The data for simulation takes the trend, but then subtracts off (rather than adding back) the high-and medium frequency component. Figures B and C show responses from a bivariate SVAR with this hours series and actual labor productivity. Shaded regions show 90 percent confidence intervals.

shocks are, in fact, correlated with lagged hours. If one takes out the apparent (though spurious) 1973 and 1997 trend breaks from labor productivity, the estimates show little bias.[11]

[11] For these simulations, the difference specification also has little bias, despite the fact that the difference VAR is "overdifferenced." This is true both in the unconditional simulated data and conditional on spurious breaks.

5. Robustness

This section discusses alternative approaches to controlling for low-frequency correlation and discusses sensitivity to alternative choices. The working paper version, as well as the appendix (Fernald, 2007), discusses further robustness checks as well as encompassing arguments.

5.1. Is it appropriate to remove subsample means?

Shocks to the economy's underlying growth rate should themselves affect behavior. Hence, the ''breaks'' VAR, which simply removes subsample means from labor productivity, could itself be misspecified by ignoring the dynamics of growth shocks.

The online appendix considers the case of two permanent technology shocks with potentially different dynamic properties. Removing subsample means from productivity growth is appropriate for identifying the dynamic response to a shock to the level of technology if the transition dynamics in response to a growth shock are sufficiently similar to the dynamics in response a levels shock.[12] The appendix also discusses the formal restrictions placed on the dynamics if there are distinct responses to growth shocks and discusses an approach that accounts explicitly of growth shocks; results on the effects of a shock to the level of technology appear qualitatively unaffected. Finally, if the apparent break is, in fact, ''spurious''—reflecting the fact that, ex post, the realized means of the technology shocks look quite different across subsamples—then the approach is also likely to be appropriate. That is, if people knew the (unchanged) 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.

Other ways to control for low-frequency correlations should be less sensitive to concerns about the dynamics induced by growth shocks. First, if one estimates over subsamples that discard the periods where the growth shocks occur, then the observed dynamics should not be driven by those shocks. The subsample responses, noted earlier, suggest hours fall when technology improves (though for the post-1982 period, they are sensitive to the variables included in the VAR).

Second, in the break specification, one can omit years around the breaks themselves. Hence, the transitory dynamics associated with the breaks should have little effect on estimates. Doing so 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, one still finds a substantially negative response. (This sample discards enough data to ensure that one controls for expectations and learning.)

In sum, the restrictions imposed on the dynamics of growth shocks do not appear to affect results on the response to shocks to the level of technology.

5.2. Similarity of shocks and responses across specifications

Table 2 compares correlations of estimated technology residuals from various specifications and shows the estimated impact effect of technology on hours. The table

[12]In DSGE models, the response to shocks to technology *growth* generally differs from the response to shocks to its *level* (e.g., Campbell, 1994; Pakko, 2002, Edge et al., 2007. The restrictions on the responses do not require that the sign of the response be the same.

Table 2
Robustness across specifications: correlations, impact effect, and granger causality tests

	Specification	Correlation of technology residuals			Impact effect on hours
		2-variable, level, breaks	4-variable, level, breaks	BFK annual residuals	
1	2-Variable, level, breaks	1	0.85	0.50	−0.43
2	4-Variable, level, breaks	0.85	1	0.48	−0.45
3	6-Variable, level, breaks	0.84	0.87	0.30	−0.40
4	2-Variable, quadratic trend, no breaks	0.97	0.81	0.48	−0.40
5	4-Variable, quadratic trend, no breaks	0.90	0.90	0.45	−0.35
6	6-Variable, quadratic trend, no breaks	0.87	0.83	0.41	−0.34
7	2-Variable, non-farm business, level, breaks	0.85	0.74	0.47	−0.45
8	4-Variable, non-farm business, level, breaks	0.49	0.79	0.32	−0.29
9	2-Variable, Francis–Ramey level, no breaks	0.90	0.72	0.53	−0.21
10	4-Variable, Francis–Ramey level, no breaks	0.48	0.25	0.33	−0.06
11	2-Variable, Francis–Ramey level, breaks	0.99	0.86	0.49	−0.48
12	4-Variable, Francis–Ramey level, breaks	0.61	0.90	0.34	−0.36
13	2-Variable, difference, no breaks	0.92	0.75	0.54	−0.29
14	2-Variable, difference, breaks	0.93	0.73	0.38	−0.18
15	2-Variable, level, no breaks	0.48	0.28	0.18	0.28
16	4-Variable, level, no breaks	0.33	0.56	0.20	0.17
17	6-Variable, level, no breaks	0.24	0.30	0.38	0.10
18	2-Variable, non-farm business, level, no breaks	0.60	0.42	0.34	0.11
19	4-Variable, non-farm business, level, no breaks	0.25	0.59	0.22	0.04

Notes: 2-Variables are productivity growth and hours per capita; 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 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 (2005b) demographically adjusted measure of the population available to work. Basu et al. (2006) residuals are available annually. To calculate the correlation of the BFK and VAR specifications, the quarterly VAR residuals are annualized (by converting to a level index, taking the annual average, then calculating annual growth rates) from 1953 to 1996. The impact effect shows the percent response of hours in the quarter for which a 1% technology improvement occurs. For annual correlations (44 observations), 0.30 is significant at the 5% level and 0.25 is significant at the 10% level.

shows correlations with the residuals from the two- and four-variable VARs with trend breaks, and with the (annual) technology innovations estimated by Basu et al. (2006). Basu et al. use a completely different identification scheme—estimating industry production functions and aggregating residuals. Their data-intensive approach relaxes some of the assumptions underlying the long-run restriction, such as perfect competition and constant returns. The BFK innovations are available annually through 1996. For the BFK correlations, one must first annualize the quarterly VAR residuals.

Table 2 shows that once one controls for low-frequency correlations, results are consistent across a range of specifications. The top of table (lines 1–14) controls in some way for low-frequency movements: removing subsample 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.[13] The bottom of the table shows the unmodified CEV specifications with no control for breaks. For both the break and no-break specifications, the table also shows non-farm business, where hours per person 16+ shows less of a U-shape (though it is still there).

After controlling for low-frequency movements, the correlations of estimated technology shocks across methods are very high and are statistically significantly correlated with the BFK shocks. In addition, all of those specifications imply a negative impact response of hours to a one% technology innovation, typically in the range of −0.4% to −0.5%.

so CEV is bad after all! so we need C 2008

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). Correlations with the BFK shocks are generally much lower; the two- and four-variable business-sector residuals are not statistically correlated. In all cases, these approaches yield positive impulse responses.

Finally, the existing literature focuses 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.

6. Conclusions

This paper seeks to better understand the strengths and weaknesses of VARs identified with long-run restrictions. Simple analytics and simulations document the sensitivity to low-frequency data correlations in empirically relevant cases. Indeed, in the so-called levels specification, the common high–low–high pattern of productivity growth and hours per person drive the apparent positive response of hours worked to a technology shock. 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. As Francis and Ramey (2005b) discuss, 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.[14]

A key recommendation for practitioners is to check sensitivity to alternative detrending methods. If results are sensitive, that should raise warning flags. Practitioners use several

[13]Differencing might help control for structural change, even if hours were, in fact, levels stationary. Over-differencing can potentially cause problems. In some simulated DSGE models, Erceg et al. (2005) find that over-differencing leads to bias; in others it does not. Francis and Ramey (2005a) and Galí and Rabanal (2004) suggest removing a quadratic trend from hours. Christiano et al. (2003) suggest, instead, removing a quadratic trend from all variables.

[14]Uhlig (2004) and Francis et al. (2005) discuss medium-run identification schemes that might be less sensitive to low-frequency data movements.

approaches to justify specifications, none of which is perfect. One is *a priori* reasoning, e.g., hours per capita cannot have a unit root. This paper shows that low-frequency movements can lead to problems even if they do not literally reflect a unit root. A second is 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. A third is Monte Carlo simulations of specific economic models (e.g., Erceg et al., 2005). Like this paper, that literature also suggests that long-run restrictions need to be used with care. However, 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. 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 suggest that in the post-war period, increases in the level of technology reduce hours worked on impact. Research continues on what model features can best explain this contractionary effect.

References

Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.

Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18, 1–22.

Barlevy, G., 2004. The cost of business cycles under endogenous growth. American Economic Review 94, 964–990.

Basu, S., Fernald, J., Kimball, M., 2006. Are technology improvements contractionary? American Economic Review 96, 1418–1448.

Blanchard, O., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. American Economic Review 79, 655–673.

Campbell, J.Y., 1994. Inspecting the mechanism: an analytical approach to the stochastic growth model. Journal of Monetary Economics 33, 463–506.

Canova, F., Lopez-Salido, D., Michelacci, C., 2007. The Labor Market Effects of Technology Shocks, manuscript, CEMFI.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2005. Are structural VARs useful guides for developing business cycle theories? Working Paper 631, FRB Minneapolis.

Christiano, L., Fitzgerald, T., 2003. The band pass filter. International Economic Review 44, 435–465.

Christiano, L., Eichenbaum, M., Vigfusson, R., 2003. What happens after a technology shock? NBER Working Paper No. 9819.

Christiano, L., Eichenbaum, M., Vigfusson, R., 2006. Assessing structural VARs. International Finance Discussion Papers 866, Board of Governors of the Federal Reserve System.

Cooley, T., Dwyer, M., 1998. Business cycle analysis without much theory: a look at structural VARs. Journal of Econometrics 83, 57–88.

David, P.A., 1990. The dynamo and the computer: an historical perspective on the modern productivity paradox. American Economic Review, Papers and Proceedings 80, 355–361.

Edge, R.M., Laubach, T., Williams, J.C., 2007. Learning and shifts in long-run productivity growth. Journal of Monetary Economics, in press, doi:10.1016/j.jmoneco.2007.01.003.

Erceg, C., Guerrieri, L., Gust, C., 2005. Can long-run restrictions identify technology shocks? Journal of the European Economic Association 3, 1237–1278.

Faust, J., Leeper, E., 1997. When do long run restrictions give reliable results? Journal of Business and Economic Statistics 15, 345–353.

Fernald, J., 1999. Roads to prosperity? Assessing the link between public capital and productivity. American Economic Review 89, 619–638.

Fernald, J., 2007. Trend breaks, long-run restrictions, and contractionary technology improvements: appendix. Available online at ⟨http://www.frbsf.org/economics/economists/jfernald.html⟩.

Fisher, J., 2006. The dynamic effects of neutral and investment-specific technology shocks. Journal of Political Economy 114, 413–451.

Francis, N., Ramey, V., 2005a. Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited. Journal of Monetary Economics 52, 1379–1399.

Francis, N., Ramey, V., 2005b. Measures of per capita hours and their implications for the technology-hours debate. NBER Working Paper No. 11694.

Francis, N., Owyang, M., Roush, J., 2005. A flexible finite-horizon identification of technology shocks. International Finance Discussion Papers 832, Board of Governors of the Federal Reserve System.

Galí, J., 1999. Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations? American Economic Review 89, 249–271.

Galí, J., Rabanal, P., 2004. Technology shocks and aggregate fluctuations: how well does the real business cycle model fit postwar U.S. data? NBER Macroeconomics Annual, pp. 225–288.

Galí, J., Lopez-Salido, D., Vallés, J., 2003. Technology shocks and monetary policy: assessing the Fed's performance. Journal of Monetary Economics 50, 723–743.

Kahn, J.A., Rich, R.W., 2006. Tracking the new economy: using growth theory to detect changes in trend productivity. Journal of Monetary Economics, in press, doi:10.1016/j.jmoneco.2006.07.008.

King, R., Plosser, C., Stock, J., Watson, M., 1991. Stochastic trends and economic fluctuations. American Economic Review 81, 819–840.

Pakko, M., 2002. What happens when the technology growth trend changes? Transition dynamics, capital growth and the 'New Economy'. Review of Economic Dynamics 5, 376–407.

Perron, P., 1989. The great crash, the oil price shock and the unit root hypothesis. Econometrica 57, 1361–1401.

Roberts, J., 2001. Estimates of the productivity trend using time-varying parameter techniques. Contributions to Macroeconomics 1, Article 3.

Sarte, P., 1997. On the identification of structural vector autoregressions. Economic Quarterly, Federal Reserve Bank of Richmond, 45–68.

Shapiro, M., Watson, M., 1988. Sources of business cycle fluctuations. In: Fischer, S. (Ed.), NBER Macroeconomics Annual 1988, pp. 111–148.

Uhlig, H., 2004. Do technology shocks lead to a fall in total hours worked? Journal of the European Economic Association 2, 361–371.