Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements: Appendix

John G. Fernald
Federal Reserve Bank of San Francisco
June 2007

This is the Online Appendix for the paper, "Trend Breaks, Long-Run Restrictions, and Contractionary Technology Improvements." It is posted at www.frbsf.org/economics/economists/jfernald.html, and is also available in the working paper version of the paper.

1. Data
2. Encompassing
3. Restrictions imposed on dynamics of growth shocks

APPENDIX 1. DATA

I follow Christiano, Eichenbaum, and Vigfusson (2003) 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 makes little difference). Other national accounts data were downloaded via Haver Analytics. For the fed funds rate, I use the average effective fed funds rate 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, my sample starts in 1950:1, which, with four lags of growth rates, implies that regressions begin 1951:2. I discount earlier data for two reasons. First, the aftermath of World War II led to enormous adjustments that affected labor productivity, employment, and other variables; the dynamics of post-war adjustment are not necessarily well captured by the same model that captures the dynamic response to more "normal" shocks (conceptually, post-war dynamics may reflect a "post-war shock" that we don't control for). Second, Young (1974) cites several studies suggesting (from both statistical analysis and "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. 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 (2003) 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.[1]

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 A1 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 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.)

---

[1] 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.

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.[2] 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.

APPENDIX 3. RESTRICTIONS IMPOSED ON DYNAMICS OF GROWTH SHOCKS

When is it appropriate to simply remove the subsample means prior to 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:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} C_{11}(L) & C_{12}(L) & C_{13}(L) \\ C_{21}(L) & C_{22}(L) & C_{23}(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t^Z \\ g_t \\ \varepsilon_t^N \end{pmatrix} \tag{1}$$

μ_t incorporates the time-varying constant term in the productivity equation; $g_t \equiv \mu_t - \mu_{t-1}$ captures the transition dynamics associated with such changes. Removing subsample means (to estimate $\Delta p_t - \mu_t$) still leaves the dynamics associated with g_t. As Faust and Leeper (1997) and Blanchard and Quah (1989) show, with the following conditions on the matrix *C(L)*, the system has a composite aggregate supply shock (a linear function of the two underlying supply shocks):

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

If D_{ij}(L)=D_{ij}, all *i*,*j*, then the supply shock captures contemporaneous shocks alone:

$$\begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{bmatrix} G_{11}(L) & G_{12}(L) \\ G_{21}(L) & G_{22}(L) \end{bmatrix} \begin{bmatrix} D_{11}\varepsilon_t^Z + D_{12} g_t \\ D_{23}\varepsilon_t^N \end{bmatrix} \tag{2}$$

Up to a scalar, the two sources of supply shocks must have the same impulse responses for 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, 2002, and Edge, Laubach, and Williams, 2007), implying that the coefficients D_{11} and D_{12} must have opposite signs. Since it is

[2] 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, 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.

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, such a path is consistent with the transitional dynamics associated with capital accumulation. Thus, removing subsample means might lead to estimates that properly capture the response to a levels shock ε_t^Z. In contrast, the VAR would be misspecified if there is structural change that we do not account for.

In any case, one can relax the restrictions that allow for a composite technology shock if one assumes that the econometrician knows the break dates. Suppose we write equation (1) as follows:

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

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

(3)

We can now express this in VAR form:

$$\tilde{C}(L)^{-1} \begin{bmatrix} \Delta p_t - \mu_t \\ n_t \end{bmatrix} = \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \tilde{C}(L)^{-1} \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t \qquad (4)$$

or, defining $[I - B(L)] \equiv \tilde{C}_0 \tilde{C}(L)^{-1}$, we can write this as:

$$\begin{bmatrix} \Delta p_t \\ n_t \end{bmatrix} = \left(I - B(L)\right) \begin{bmatrix} \mu_t \\ 0_t \end{bmatrix} + B(L) \begin{bmatrix} \Delta p_t \\ n_t \end{bmatrix} + \tilde{C}_0 \begin{pmatrix} \varepsilon_t^Z \\ \varepsilon_t^N \end{pmatrix} + \left(I - B(L)\right) \begin{bmatrix} C_{12}(L) \\ C_{22}(L) \end{bmatrix} g_t \qquad (5)$$

If we assume the break dates are known at the time they occurred, then we can estimate this equation. In particular, we impose the long-run restriction and augment the SVAR with a dummy variable (including lags) for the break dates. Conceptually, this corresponds to the approach taken to government spending dummies by Ramey and Shapiro (1998) and others.

In real time, there is 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 detects the change in aggregate productivity data. In addition, coefficients may capture lags that arise from slow learning.

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.

Table A1: 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) shows 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 and 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) and have an F statistic greater than 6.83. Sample period is 1950:2-2004:2.

Appendix References

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.

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

Edge, R.M., Laubach, T., Williams, J., 2007. Learning and shifts in long-run productivity growth. Forthcoming, Journal of Monetary Economics.

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

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.

Ramey, V., Shapiro, M., 1998. Costly capital reallocation and the effects of government spending. Carnegie-Rochester Conference Series on Public Policy 48, 145-194

Young, A. H., 1974. Reliability of the quarterly national income and product accounts of the United States, 1947-71. Review of Income and Wealth 20, 1-40.