Journal of Econometrics 146 (2008) 202–206

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

The Beveridge–Nelson decomposition in retrospect and prospect

Charles R. Nelson

Department of Economics, University of Washington, Box 353330, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Available online 28 August 2008

JEL classification:
C52
E32
E37

Keywords:
Beveridge–Nelson decomposition
Trend and cycle
Forecasting

ABSTRACT

Beveridge and Nelson [Beveridge, Stephen, Nelson, Charles R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. Journal of Monetary Economics 7, 151–174] proposed that the long-run forecast is a measure of trend for time series such as GDP that do not follow a deterministic path in the long run. They showed that if the series is stationary in first differences, then the estimated trend is a random walk with drift that accounts for growth, and the cycle is stationary. In contrast to linear de-trending, the smoother of Hodrick and Prescott (1981) and Hodrick and Prescott [Hodrick, Robert, Prescott, Edward C., 1997. Post-war US business cycles: An empirical investigation. Journal of Money Credit and Banking 29 (1), 1–16] and the unobserved components model of Harvey, [Harvey, A.C., 1985. Trends and cycles in macroeconomic time series. Journal of Business and Economic Statistics 3, 216–227]. Watson [Watson, Mark W., 1986. Univariate detrending methods with stochastic trends Journal of Monetary Economics 18, 49–75] and Clark [Clark, Peter K., 1987. The cyclical component of US economic activity. The Quarterly Journal of Economics 102 (4), 797–814], the BN decomposition attributes most variation in GDP to trend shocks while the cycles are short and brief. Since each is an estimate of the transitory part of GDP that will die out, it seems natural to compare cycle measures by their ability to forecast future growth. The results presented here suggest that cycle measures contain little if any information beyond the short-term momentum captured by BN.

1. Genesis of the Beveridge–Nelson decomposition

What does 'trend' mean for a time series that is not deterministic in the long run but nevertheless is 'trending' in the sense that it grows over time? This seemed an obvious question in the early 1970s when the modeling strategy of Box and Jenkins (1970)lead economists to model GNP and other trending economic time series as ARMA models in their first differences. Instead of representing the data as temporary fluctuations around a fixed trend, these models imply that the future will diverge from any pre-specified path, although the forecasted path is readily computed. It seemed to Stephen Beveridge and me that a satisfactory definition of trend for these 'I(1)' time series would preserve the property of trend that it is the best estimate of where the variable will be in the distant future; so why not define trend as simply the long-horizon forecast? Rather than being fixed and pre-determined, this trend will shift as each new data point reveals new information about the future. That implies that trend is a source of stochastic (unpredictable) variation in, say, GDP, and that it is meaningful to think of parsing its fluctuations into a part due to trend and a part due to the business cycle. Further, we were able to show that the trend is always a random walk with drift and the deviation from trend is stationary.

The first draft of the paper that finally became Beveridge and Nelson (1981) was dated July 1972 and was presented at the Western Economic Association Meetings in August. What distinguishes that draft from the final paper is considerably more attention to the algebra of the decomposition. There is little evidence in my files that the paper generated much interest (a letter from the chief labor economist for the State of Nevada stands out), and we put the paper aside until a draft dated August 1977. A hand-written letter from Steve Beveridge dated May 18, 1978 says "I sent the paper to Brunner last Friday so it is now in the hands of the gods. If it's accepted I'll buy you two lunches. I'm curious as to why you want five copies of the paper. Are scratch pads hard to come by at UW?" Karl Brunner was of course the editor of the *Journal of Monetary Economics*, and I am still waiting for the lunches.

A letter dated March 15, 1979 from Karl Brunner reports the reaction of a referee: 'While he found the paper of some interest . . . . He also felt the paper might be more suitable for an NBER journal. I would be willing to take another look at it'. Then on Oct. 15: 'I am interested in publishing the paper. . .'. A three page referee report was enclosed, and of greatest concern to us was the request that we reverse the sign of the cycle component. What bothered the referee was our result that the cycle component

E-mail address: cnelson@u.washington.edu.

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.08.008

of GDP is negative when the economy is growing rapidly. This followed from the empirical fact that the growth rate of GDP tends to persist, so during times of unusually rapid growth the forecasted level of GDP (adjusted for average growth) is above the current actual level, implying the cycle is negative. By redefining 'cycle' as 'forecastable momentum' (trend minus actual instead of the usual actual minus trend) we reversed the sign, satisfied the referee, and laid a trap for future readers. A July 1980 letter from Brunner said 'When you have dealt with these relatively minor editorial points…we will plan to publish the paper in the March 1981 issue of the *JME*'. It is hard to imagine a mere nine month publication lag today, in spite of all the new technology.

Since the paper was not exactly an instant hit, a reasonable question might be, why did we think it was worth plugging away at it for nine years? I don't know. A question I would like to try to answer here is this: How well has the Beveridge–Nelson decomposition weathered the test of time? By that I mean, is it a useful method of trend cycle decomposition, particularly for US GDP? The BN decomposition also stands as a useful statistical result, the fact that any I(1) time series may be expressed as random walk plus a stationary component, but that is not my focus here.

2. How should we assess the effectiveness of alternative trend-cycle estimates?

What are appropriate criteria for judging the BN and other trend-cycle decompositions? Is there a meaningful ranking of them? Or is the choice of method simply a matter of taste, picked to match the priors of the user? I wish to argue that the objective of decomposition is to separate temporary movements from those that permanently shift the level of a time series, most importantly of aggregate output and ask: How successful are the BN and competing decompositions in doing that?

If the measured cycle component is temporary then it predicts future growth rates of the opposite sign. For example, if the economy is in recession today, we mean that output is below trend, and recovery will require future growth at an above-average rate. Conversely, if output is above trend we can reasonably expect tepid growth in coming quarters. Predictability is the essence of 'transitory' variation, as it is for seasonal variation—both may be expected to be reversed in future periods. Indeed, the business cycle is like a seasonal cycle except that the seasonal frequency is known *a priori*. Predictability of the cycle implies a metric then for measuring the effectiveness of alternative decompositions: how well do they predict future growth or future turning points in the economy? This idea is not new, and it is central to several articles in the references including Cogley (2002), Hodrick and Zhang (2003), Orphanides and van Norden (2005), Rotenberg and Woodford (1996) and Wakerly et al. (2006).

However, the criterion adopted in much of the literature on business cycles is very different from predictability; namely that the objective is to isolate variation at 'the business cycle frequency'. Filters, basically moving averages of the data, can be designed to remove variation at other frequencies and so in principle reveal variation at the business cycle frequency. Thus, if recessions occur about every seven years, we would want to filter out frequencies outside a range around seven years. Certainly, the zero frequency – that of the trend component – is to be excluded, as should be the seasonal frequency. However, the theory of filters applies to stationary time series, and application of that theory to non-stationary time series such as GDP is problematic; see Cogley and Nason (1995) and Murray (2003). In particular, the fact that the filtered series has a spectral peak around the business cycle frequency does not establish that this component contains the transitory variation in GDP, or that there is a transitory component.

And how do we know that the business cycle frequency is seven years? Estimates of the frequency seem to derive from the interval between NBER turning points. Recessions correspond to periods of two or more quarters of decline in the economy, and generally correspond to declines in GDP though many indicators are examined. But periods of decline in a non-stationary time series do not establish the existence of a transitory component; for example, a random walk with drift will exhibit periods of decline but has no transitory component. More relevant perhaps, the interval between turning points will reflect properties of both the stochastic trend and the cycle component. Thus, the NBER chronology does not in itself tell us much about the statistical properties of the cyclical component of GDP, nor does it even establish the existence of a business cycle. Looking outside the US, Cerra and Saxena (2008) find no tendency for output to rebound following recessions in a virtually exhaustive sample of economies.

In contrast, if there is a cycle and a given trend-cycle decomposition is able to capture some of its variation, then a forecast based on that cycle measure will reflect the decay of the cycle over time as it reverts to its long run path of zero. An unobserved-components (UC) representation of the decomposition will even predict how successful it will be in forecasting, as a function of the rate of decay of the cycle and the relative variance of shocks to trend and cycle and their correlation. In practice that ideal is never achieved because we only have estimates of the components, whether those are based on a filter or a formal UC model with estimated parameters. Further, success in forecasting out-of-sample is not necessarily a property of a valid decomposition — by which I mean a representation of the data generating process that replicates its moments. For example, the correct decomposition of a random walk correctly assigns zero variation to the cycle component, and correctly predicts that future growth cannot be forecasted. The BN decomposition does assign some variation to the cycle but we would expect it to have little power to forecast GDP growth since trend variation is large and unpredictable if the BN decomposition is correct. Thus we are interested too in the consistency of the predicted ability to predict growth and the actual ability.

However, forecasting success within-sample may be very high even when completely meaningless. A retrospective decomposition based on the inference that the cycle must have been positive on the eve of NBER peaks is very successful predicting within-sample but has no implications for the future. Indeed, Nelson and Kang (1981) showed that a linear trend line fitted to the realization of a random walk will account for much of the variation *ex post* and produce a highly predictable cycle component within-sample, though the *ex ante* predictability of a random walk is zero. Within-sample predictability is not a reliable guide to the validity – or usefulness – of a decomposition method.

3. GDP growth in prospect: The relative contribution of alternative cycle measures

The question I would like to address now is this: Do alternative measures of the cycle component of GDP contain information that has been useful in forecasting GDP growth? The forecast comparisons presented here are limited to univariate methods so the information set for predicting the growth of GDP is only past GDP. I think it is obvious (though I do not have statistics to back this up) that the most popular method of trend-cycle decomposition is the filter of Hodrick and Prescott (1981) and Hodrick and Prescott (1997) which seeks to balance smoothness of the cycle against variance of the measured cycle. Further analysis of the 'HP filter' is given by Harvey and Jaeger (1993) and Schlicht (2005). Also highly influential have been the UC models of Harvey (1985), Watson (1986) and Clark (1987) which both model the trend as a random

Table 1
Descriptive statistics for quasi-real-time and final cycle estimates 1956.4–2005.3

	Autocorrelation lag 1		Std. Dev. x 100		Correlation
	QRT	Final	QRT	Final	QRT, Final
Beveridge–Nelson	0.29	0.29	0.54	0.45	0.98
Clark	0.87	0.95	1.38	1.94	0.64
Hodrick–Prescott	0.89	0.85	1.64	1.57	0.55
Linear	0.96	0.97	3.34	3.95	0.72
w/ Break in level	0.94	0.88	2.61	2.30	0.55
w/ Break in slope	0.93	0.94	2.84	2.99	0.59
w/ Break in both	0.93	0.89	2.40	2.16	0.67

walk and the cycle as an AR process; here we use the latter which allows the growth rate of the trend to evolve as a random walk as well. Certainly many practitioners still 'detrend' by fitting a linear trend line to the logs of the data so that is an essential benchmark. The influential work of Perron (1989) has kept the linear trend model in the running as a description of GDP as long as breaks in level, slope, or both are allowed to occur. Thus linear models with all three types of breaks are included in the comparison, and the break date is chosen to maximize fit.

The data are post-war US GDP 1947.1 through 2005.3 and we compute the measured cycle at each quarter, beginning with 1956.4, re-estimating parameters, if any, up to the current date, successively through the sample period. Orphanides and van Norden (2002) distinguish three measures of the cycle for any given historical quarter: one is the estimate that would have been made in real-time using preliminary data, another is the 'final' estimate made retrospectively with the benefit of all the subsequent data in revised form, and a third is the 'quasi-real-time' (QRT) forecast made by the researcher today using revised data but only the observations up to the historical date. Those authors found significant differences between final and real-time estimates, but little difference between quasi-real-time and real-time estimates. In other words, data revisions are not as important as the distinction between use of past data as opposed to future as well as past data. The results presented here use only revised data and we focus on the comparison between one-sided quasi-real-time (QRT) estimates of the cycle and two-sided 'final' estimates. When a decomposition requires estimation of parameters, they are re-estimated at each date before computing the cycle estimate, as is the break-date when relevant.

Table 1 displays summary statistics for the QRT and final cycle estimates. There is a sharp distinction between the BN cycle and the other cycle estimates: it is much less strongly autocorrelated, it is much smaller in amplitude as measured by standard deviation, and the distinction between QRT and final cycle is much less important. The first two properties reflect the lack of smoothness priors for the trend; note that the smoothest trend – linear – also produces a cycle with the strongest autocorrelation and largest standard deviation. The high correlation between QRT and final BN estimates reflects the fact that it is inherently a one-sided estimate of cycle, so future data only influences estimation of the ARMA parameters. In this case, the model is AR(1) as suggested by lag selection based on SIC, and the AR parameter is very stable over the sample period. Future data matters the most for the HP filter and the linear with break-in-level model where the correlation between QRT and final cycles is only 0.55 for each.

Table 2 presents three sets of regressions in which the objective is to predict quarterly growth in GDP one quarter ahead using the various measures of the cycle. The first panel reports the least squares coefficients and p-values for BN alone (in this paper the BN cycle has the conventional interpretation as the actual minus trend) and then successively in combination with QRT cycle estimates using Clark, HP, linear trend, and linear with break in level, slope, and both. This exercise is in the spirit of Granger's composite prediction and is intended to suggest the marginal information content of cycle estimates. Explanatory power is low, R-square is only .08 for BN alone and none of the other cycle estimates are able to raise this number. While the p-value for BN is essentially zero, none of the other cycle estimates has a p-value lower (more significant) than .31 in the presence of BN, and in that case the sign is wrong.

The second exercise looks at the predictive value of each cycle estimate by itself. To separate the explanatory power of the level of each cycle from the dynamics of the cycle process I included the lagged first difference of the cycle along with the level. To motivate this specification, note that the predicted change in the AR(2) cycle process of Clark and Watson may be expressed in terms of the lagged level and the lagged change in the cycle. We expect that the sign of the coefficient on the lagged level will be negative; indeed for a stationary AR(p) cycle process the coefficient will be the sum of the AR coefficients (an amount less than 1) minus 1. More generally, if the cycle does represent a deviation from trend, then it will predict changes back toward trend. The results in the second panel of Table 2 suggest that only BN and HP have predictive power; however the sign is the opposite of what is expected in the case of HP. A positive HP cycle has signaled more rapid future growth rather than reversal towards trend.

The third panel of Table 2 asks how much difference the forecast comparison would be if we gave each methodology the advantage of hindsight in the form of using all the historical data to estimate the cycle retrospectively after the fact. These 'final' cycle estimates are generally highly significant in the regressions and all have the appropriately negative sign, though BN remains significant in each composite. The difference between R-squared in the first panel and this one is a measure of the value of hindsight for each of these methods, and it is by far the greatest for HP and next greatest for a linear trend with breaks in level and slope.

How well do any of the QRT cycle estimates signal NBER turning points? We expect that as GDP departs further from trend the probability of a turning point becomes more likely. In the first exercise, reported in the top panel of Table 3, the objective is to predict if a peak will occur the next quarter, given that an expansion is underway, using the duration of the expansion and, successively, each cycle estimate in a binary probit model. The table entries are ML p-values and the McFadden R-square. Duration alone has a p-value of only .12 which is consistent with the finding in the literature that expansions are not strongly duration-dependent; see Kim and Nelson (1998). BN by itself is highly significant and remains so in combination with duration. However none of the other cycle estimates is significant either alone or in combination with duration, though HP comes closest. The lower panel reports results for predicting a trough, given that a recession is underway and is perhaps less realistic simply because the lag in the information that the economy is officially in recession is long relative to the length of a recession. Duration is a highly significant predictor by itself, consistent with previous findings that recessions are duration dependent, and it is significant in combination with each of the QRT cycle estimates. The smallest p-value for any cycle estimate is for HP, but it loses its significance in the presence of duration, and none of the other cycle estimates have significant predictive ability.

4. Conclusions

This paper presents evidence that the univariate trend-cycle decompositions widely used for macro-econometric analysis have little if any value as predictors of economic activity in real time. Only the modest momentum in growth, captured by the BN cycle estimates, allows for a very modest amount of predictability. Perhaps large and seemingly predictable business cycles are only

Table 2
Predictive regressions for real GDP growth using lag of cycle estimates 1956.4–2005.3

Quasi-real-time cycle measures vs. BN							
Beveridge–Nelson	−.46 (.00)	−.47 (.00)	−.37 (.01)	−.48 (.00)	−.49 (.00)	−.48 (.00)	−.51 (.00)
Clark		−.03 (.50)					
Hodrick–Prescott			.05 (.31)				
Linear				−.01 (.47)			
w/ Break in level					−.01 (.63)		
w/ Break in slope						−.01 (.76)	
w/ Break in both							−.03 (.37)
R-squared	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Quasi-real-time cycle measures in presence of change in cycle							
Beveridge–Nelson	−.54 (.00)						
Clark		−.03 (.56)					
Hodrick–Prescott			.09 (.02)				
Linear				−.01 (.71)			
w/ Break in level					.01 (.78)		
w/ Break in slope						.01 (.73)	
w/ Break in both							−.00 (.97)
R-squared	0.08	0.004	0.1	0.09	0.08	0.08	0.06
Final cycle estimates vs. BN							
Beveridge–Nelson	−.58 (.00)	−.63 (00)	−.77 (.00)	−.61 (.00)	−.65 (00)	−.62 (.00)	−.68 (.00)
Clark		−.09 (.005)					
Hodrick-Prescott			−.21 (.00)				
Linear				−.03 (.04)			
w/ Break in level					−.10 (.00)		
w/ Break in slope						−.05 (.01)	
w/ Break in both							−.11 (00)
R-squared	0.08	0.12	0.21	0.10	0.14	0.11	0.15

Least Squares Coefficient with p-value in parentheses.

Table 3
Probit models for NBER turning points using lag of quasi-real-time cycle estimates and duration 1956.4–2005.3

Predicting NBER peak during expansions															
Equation:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Duration	0.12		0.11		0.05		0.11		0.16		0.07		0.11		0.05
Beveridge–Nelson		0.02	0.02												
Clark				0.96	0.22										
Hodrick-Prescott						0.14	0.13								
Linear								0.46	0.93						
w/ Break in level										0.98	0.33				
w/ Break in slope												0.85	0.68		
w/ Break in both														0.83	0.21
McFadden R-sq.	0.04	0.10	0.14	0.00	0.06	0.03	0.07	0.01	0.04	0.00	0.05	0.00	0.04	0.00	0.06
Predicting NBER trough during recessions															
Duration	0.01		0.02		0.01		0.09		0.03		0.02		0.05		0.01
Beveridge–Nelson		0.09	0.13												
Clark				0.36	0.36										
Hodrick-Prescott						0.03	0.18								
Linear								0.08	0.51						
w/ Break in level										0.19	0.38				
w/ Break in slope												0.06	0.81		
w/ Break in both														0.35	0.13
McFadden R-sq.	0.25	0.12	0.38	0.03	0.28	0.22	0.32	0.10	0.27	0.06	0.28	0.12	0.26	0.03	0.33

Table entries are ML binary probit p-values.

apparent in retrospect, only statistical artifacts. If so, much of the variation observed in US GDP is due to permanent shocks which shift the level of the trend, itself a stochastic process, and are largely unpredictable.

Even if traditional business cycles do exist, the results in this and other papers cited here clearly show that they are not well measured in real time. This empirical fact poses a severe challenge to the conduct of monetary policy which aims to dampen the business cycle while giving weight to controlling inflation. Policy makers should pay little attention to real time estimates of the business cycle and focus on targeting inflation.

Acknowledgements

This paper is based on the author's after-dinner address at the Beveridge–Nelson 25th Anniversary Conference at the Federal Reserve Bank of Atlanta March 31–April 1, 2006. Thanks are due to organizers Tim Cogley, Steven Durlauf, and James Nason, and to Stephen Beveridge, Tim Cogley, Michael Dueker, James Morley, Chris Murray, James Nason, Jeremy Piger, and Richard Startz for helpful comments. Brian Donhauser provided excellent research assistance and valuable insights. Support from the Ford and Louisa Van Voorhis Endowment is gratefully acknowledged. Responsibility for the content is entirely the author's.

References

Beveridge, Stephen, Nelson, Charles R., 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. Journal of Monetary Economics 7, 151–174.
Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis; Forecasting and Control. Holden-Day, San Francisco.
Cerra, Valerie, Saxena, Sweta Chaman, 2008. Growth dynamics: The myth of economic recovery. American Economic Review 98, 439–457. IMF Working Paper WP/05/147.
Clark, Peter K., 1987. The cyclical component of US economic activity. The Quarterly Journal of Economics 102 (4), 797–814.
Cogley, Timothy, Nason, James, 1995. Effects of the Hodrick–Prescott filter on trend and difference stationary time series: Implications for business cycle research. Journal of Economic Dynamics and Control 19, 253–278.
Cogley, Timothy, 2002. A simple adaptive measure of core inflation. Journal of Money Credit and Banking 34 (1), 94–113.
Harvey, A.C., 1985. Trends and cycles in macroeconomic time series. Journal of Business and Economic Statistics 3, 216–227.
Harvey, A.C., Jaeger, A., 1993. Detrending, stylized facts and the business cycle. Journal of Applied Econometrics 8 (3), 231–247.
Hodrick, Robert, Prescott, Edward C., 1981. Post-war US business cycles: An empirical investigation. Working paper, Carnegie-Mellon University.
Hodrick, Robert, Prescott, Edward C., 1997. Post-war US business cycles: An empirical investigation. Journal of Money Credit and Banking 29 (1), 1–16.
Hodrick, Robert J., Zhang, Xiaoyan, 2003. Forecasting real GDP growth with Hodrick–Prescott filtered components: An out-of-sample analysis. Discussion paper.
Kim, Chang-Jin, Nelson, Charles R., 1998. Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. The Review of Economics and Statistics LXXX (2), 188–201.
Murray, Christian J., 2003. Cyclical properties of Baxter-King filtered time series. The Review of Economics and Statistics 85, 472–476.
Nelson, Charles R., Kang, Heejoon, 1981. Spurious periodicity in inappropriately detrended time series. Econometrica 49, 741–751.
Orphanides, Athanasios, van Norden, Simon, 2002. The unreliability of output-gap estimates in real time. The Review of Economics and Statistics 84 (4), 569–583.
Orphanides, Athanasios, van Norden, Simon, 2005. The reliability of inflation forecasts based on output gap estimates in real time. Journal of Money Credit and Banking 37 (3), 583–601.
Perron, Pierre, 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57, 1361–1401.
Rotenberg, Julio J., Woodford, Michael, 1996. Real business cycle models and the forecastable movements in output, hours and consumption. American Economic Review 86, 71–89.
Schlicht, Ekkehart, 2005. Estimating the smoothing parameter in the so-called Hodrick–Prescott filter. Journal of the Japan Statistical Society 35 (1), 99–119.
Wakerly, Elizabeth C., Byron, G.Scott, Nason, James M., 2006. Common trends and common cycles in Canada: Who know so much has been going on? Canadian Journal of Economics 39, 320–347.
Watson, Mark W., 1986. Univariate detrending methods with stochastic trends. Journal of Monetary Economics 18, 49–75.