Why You Should Never Use the Hodrick-Prescott Filter*

James D. Hamilton

jhamilton@ucsd.edu

Department of Economics, UC San Diego

July 30, 2016

Revised: November 15, 2016

ABSTRACT

Here's why. (1) The HP filter produces series with spurious dynamic relations that have no basis in the underlying data-generating process. (2) A one-sided version of the filter reduces but does not eliminate spurious predictability and moreover produces series that do not have the properties sought by most potential users of the HP filter. (3) A statistical formalization of the problem typically produces values for the smoothing parameter vastly at odds with common practice, e.g., a value for λ far below 1600 for quarterly data. (4) There's a better alternative. A regression of the variable at date $t+h$ on the four most recent values as of date t offers a robust approach to detrending that achieves all the objectives sought by users of the HP filter with none of its drawbacks.

---

*I thank Daniel Leff for outstanding research assistance on this project.

1 Introduction.

Hodrick and Prescott (1981, 1997) proposed a method for separating an observed series y_t into components typically labeled trend and cycle. The drawbacks to the approach have been known for some time. Nevertheless, the method continues today to be very widely adopted in academic research, policy studies, and analysis by private-sector economists. For this reason it seems useful to collect these results in this paper and remind potential users of the HP filter of both the pitfalls and the existence of superior alternatives.

2 Characterizations of the Hodrick-Prescott filter.

Given T observations on a variable y_t, Hodrick and Prescott (1981, 1997) proposed interpreting a trend component g_t as a very smooth series that does not differ too much from the observed values.[1] It is calculated as[2]

$$\min_{\{g_t\}_{t=-1}^{T}} \left\{ \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=1}^{T} [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\}. \tag{1}$$

When the smoothness penalty $\lambda \to 0$, g_t would just be the series y_t itself, whereas when $\lambda \to \infty$ the procedure amounts to a regression on a linear time trend (that is, produces a series whose second difference is exactly 0). The common practice is to use a value of $\lambda = 1600$ for quarterly time series.

[1] Phillips and Jin (2015) reviewed the rich prior history of generalizations of this approach.

[2] There are some slight differences across different authors in treatment of the endpoints. For example, de Jong and Sakarya (2016) and Cornea-Madeira (forthcoming) take the second summation in (1) to be over $t = 3$ to T and characterize the exact inference for that case, while King and Rebelo (1993) take the summation over $t = 2$ to $T + 1$. Our expression follows Hodrick and Prescott (1997) and the algorithms coded in Stata and RATS.

A closed-form expression for the resulting series for trend and cycle can be written in vector notation by defining $\tilde{T} = T + 2$ and

$$\underset{(T\times 1)}{y} = (y_T, y_{T-1}, ..., y_1)'$$

$$\underset{(\tilde{T}\times 1)}{g} = (g_T, g_{T-1}, ..., g_{-1})'$$

$$\underset{(T\times \tilde{T})}{H} = \begin{bmatrix} \underset{(T\times T)}{I_T} & \underset{(T\times 2)}{0} \end{bmatrix}$$

$$\underset{(T\times \tilde{T})}{Q} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix}.$$

The problem (1) can then equivalently be written

$$\min_g \{(y - Hg)'(y - Hg) + \lambda(Qg)'(Qg)\}$$

whose solution[3] is

$$g^* = (H'H + \lambda Q'Q)^{-1}H'y = A^*y. \tag{2}$$

The inferred trend g_t^* for any date t is thus a linear function of the full set of observations on y for all dates.

As noted by Hodrick and Prescott (1981) and King and Rebelo (1993), the identical inference can alternatively be motivated from particular assumptions about the time-series

[3] The derivative with respect to g is $-2H'(y - Hg) + 2\lambda Q'Qg$. Cornea-Madeira (forthcoming) provided further details on A^* and a convenient algorithm for calculating it.

behavior of the growth and cyclical components. Suppose our goal was to choose a value of a $(T \times 1)$ vector a_t such that the estimate $\hat{g}_t = a_t' y$ has minimum expected squared difference from the true trend:

$$\min_{a_t} E(g_t - a_t' y)^2. \tag{3}$$

The solution depends on what we assume about the variance of y and its covariance with the trend, and is given in the general case by $\tilde{a}_t = [E(yy')]^{-1} E(yg_t)$, the population analog to a sample regression coefficient.[4] Stacking the estimates $\tilde{g}_t$ into a $(\tilde{T} \times 1)$ vector produces

$$\tilde{g} = E(gy') [E(yy')]^{-1} y = \tilde{A} y. \tag{4}$$

Let c_t denote the cyclical component and v_t the second difference of the trend component:

$$y_t = g_t + c_t \tag{5}$$

$$g_t = 2g_{t-1} - g_{t-2} + v_t \tag{6}$$

Suppose that we assume that v_t and c_t are uncorrelated white noise processes that are also uncorrelated with (g_0, g_{-1}), and let C_0 denote the (2×2) variance of those first two pre-sample values. For concreteness we write these assumptions formally as equations (25)-(29)

---

[4] For any a_t we have

$$\begin{aligned} E(g_t - a_t' y)^2 &= E(g_t - \tilde{a}_t' y + a_t' y - a_t' y)^2 \\ &= E(g_t - \tilde{a}_t' y)^2 + 2E[(g_t - \tilde{a}_t' y) y'](\tilde{a}_t - a_t) + (\tilde{a}_t - a_t)' E(yy')(\tilde{a}_t - a_t). \end{aligned}$$

The middle term equals 0 by the definition of $\tilde{a}_t$:

$$E[(g_t - \tilde{a}_t' y) y'](\tilde{a}_t - a_t) = \{E(g_t y') - E(g_t y')[E(yy')]^{-1} E(yy')\}(\tilde{a}_t - a_t).$$

Hence the expression is minimized by setting $a_t = \tilde{a}_t$. Note that this result does not assume stationarity, in that each element of the $(T \times T)$ matrix $E(yy')$ could be distinct.

in the appendix. These assumptions imply a certain structure to the covariance matrices in (4). If we adopted these assumptions but had no information about the initial states (represented as $C_0^{-1} \to 0$),then the following proposition establishes that in any sample of any size T, the inference (4) would be numerically identical to expression (2).

Proposition 1. *For* $\lambda = \sigma_c^2/\sigma_v^2$ *and any fixed* T, *under conditions* (5)-(6) *and* (25)-(29), *the matrix* $\tilde{A}$ *in* (4) *converges to the matrix* A^* *in* (2) *as* $C_0^{-1} \to 0$.

Proposition 1 establishes that if researcher 1 sought to identify a trend by solving the minimization problem (1) while researcher 2 found the optimal linear estimate of a trend process that was assumed to be characterized by the particular assumption that v_t and c_t were both white noise, the two researchers would arrive at the numerically identical series for trend and cycle provided the ratio of σ_c^2 to σ_v^2 assumed by researcher 2 was identical to the value of λ used by researcher 1.

The Kalman smoother is an iterative algorithm for calculating the population linear projection (4) for models where the variance and covariance can be characterized by some recursive structure.[5] In this case, (5) is the observation equation and (6) is the state equation. Thus as noted by Hodrick and Prescott, applying the Kalman smoother to the above state-space model starting from a very large initial variance for $(g_0, g_{-1})'$ offers a convenient algorithm for calculating the HP filter, and is in fact a way that the HP filter is often calculated in practice. Nevertheless, this observation should also be a bit troubling for users of the HP filter, in that they never defend the claim that the particular structure

[5] See for example Hamilton, 1994, equation [13.6.3].

assumed in Proposition 1 is an accurate representation of the true data-generating process. Indeed, if a researcher did know for certain that these equations were the true data-generating process, and further knew for certain the value of the population parameter $\lambda = \sigma_c^2/\sigma_v^2$, he would probably be unhappy with using (2) to separate cycle from trend! The reason is that if this state-space structure was the true DGP, the resulting estimate of the cyclical component $c_t = y_t - \tilde{g}_t$ would be white noise– it would be random and exhibit no discernible patterns. By contrast, users of the HP filter hope to see suggestive patterns in plots of the series that is to be labeled as the cyclical component of y_t.

We can characterize some further aspects of the HP filter by rewriting (2) as

$$(H'H + \lambda Q'Q)g^* = H'y. \tag{7}$$

The tth element of this system can be written[6]

$$[1 + \lambda(1 - L^{-1})^2(1 - L)^2]g_t^* = y_t \qquad \text{for } t = 1, 2, ..., T - 2 \tag{8}$$

for L the lag operator ($L^k x_t = x_{t-k}$, $L^{-k} x_t = x_{t+k}$). Expression (8) states that $F(L)g_t^* = y_t$ for

$$F(L) = 1 + \lambda(1 - L^{-1})^2(1 - L)^2. \tag{9}$$

The following proposition establishes some properties of this filter.[7]

[6] Note that for any $(\tilde{T} \times 1)$ vector x, the tth element of Qx corresponds to $x_t - 2x_{t-1} + x_{t-2} = (1 - L)^2 x_t$. Likewise for w a $(T \times 1)$ vector the tth element of $Q'w$ corresponds to $(1 - L^{-1})^2 w_t$. Thus the multiplication $Q'Qx$ applies the compound operator $(1 - L^{-1})^2(1 - L)^2 x_t$.

[7] Related results have been developed by King and Rebelo (1989, 1993), Cogley and Nason (1995), and McElroy (2008). Unlike these papers, here I provide simple direct expressions for the values of ϕ_1 and ϕ_2, and my expressions of the HP filter entirely in terms of real parameters in (10) and (11) appear to be new.

Proposition 2. *For any $\lambda : 0 < \lambda < \infty$, the inverse of the operator* (9) *can be written*

$$[F(L)]^{-1} = C\left[\frac{1-(\phi_1^2/4)L}{1-\phi_1 L-\phi_2 L^2} + \frac{1-(\phi_1^2/4)L^{-1}}{1-\phi_1 L^{-1}-\phi_2 L^{-2}} - 1\right] \tag{10}$$

where

$$\frac{1}{1-\phi_1 z-\phi_2 z^2} = \textstyle\sum_{j=0}^{\infty} R^j[\cos(mj)+\cot(m)\sin(mj)]z^j \tag{11}$$

$$\frac{1}{1-\phi_1 z^{-1}-\phi_2 z^{-2}} = \textstyle\sum_{j=0}^{\infty} R^j[\cos(mj)+\cot(m)\sin(mj)]z^{-j}$$

$$\phi_1(1-\phi_2) = -4\phi_2 \tag{12}$$

$$(1-\phi_1-\phi_2)^2 = -\phi_2/\lambda \tag{13}$$

$$C = \frac{-\phi_2}{\lambda(1-\phi_1^2-\phi_2^2+\phi_1^3/2)} \tag{14}$$

$$R = \sqrt{-\phi_2}$$

$$\cos(m) = \phi_1/(2R). \tag{15}$$

Roots of $(1-\phi_1 z-\phi_2 z^2) = 0$ *are complex and outside the unit circle, with* ϕ_1 *a real number between* 0 *and* 2, ϕ_2 *a real number between* -1 *and* 0, *and* R *a real number between* 0 *and* 1.

Proposition 2 establishes that for observations in the middle of a large sample, the HP trend could be calculated by first constructing a linear function of the current and past values by iterating on $\xi_{1t} = w_{1t} + \phi_1\xi_{1,t-1} + \phi_2\xi_{1,t-2}$ for $t = 3, 4, ...$ and $w_{1t} = y_t - (\phi_1^2/4)y_{t-1}$, and next constructing a linear function of the current and future values by iterating on $\xi_{2t} = w_{2t} + \phi_1\xi_{2,t+1} + \phi_2\xi_{2,t+2}$ for $t = T-2, T-3, ...$ and $w_{2t} = y_t - (\phi_1^2/4)y_{t+1}$.[8] We then

[8] For large T, these iterations will converge to the HP trend for observations around $t = T/2$ from any starting values for ξ_{11} and ξ_{2T}. It will be a better approximation near the endpoints if started from $\xi_{1t} = [(1-\phi_1^2/4)/(1-\phi_1-\phi_2)]y_t$ for $t = 1, 2$ and $\xi_{2t} = [(1-\phi_1^2/4)/(1-\phi_1-\phi_2)]y_t$ for $t = T, T-1$.

add these together, subtract the current value y_t to avoid double counting, and multiply by a constant C so that the coefficients sum to unity. The resulting value $g_t = C(\xi_{1t} + \xi_{2t} - y_t)$ will equal the HP trend g_t^* for t near the middle of a large sample. The proposition also gives us a closed-form expression for the results of the iterations:

$$\xi_{1t} = \textstyle\sum_{j=0}^{\infty} R^j[\cos(mj) + \cot(m)\sin(mj)]w_{1,t-j}$$

$$\xi_{2t} = \textstyle\sum_{j=0}^{\infty} R^j[\cos(mj) + \cot(m)\sin(mj)]w_{2,t+j}.$$

Figure 1 plots the values of ϕ_1 and ϕ_2 generated by different values of λ. For $\lambda = 1600$, $\phi_1 = 1.777$ and $\phi_2 = -0.7994$. These imply $R = 0.8941$, so that the absolute value of the weights decay with a half-life of about 6 quarters.[9]

From (8), the cyclical component $c_t = y_t - g_t^*$ is then characterized by

$$c_t = \lambda(1 - L^{-1})^2(1 - L)^2 g_t^* = \frac{\lambda(1 - L^{-1})^2(1 - L)^2}{F(L)} y_t = \frac{\lambda(1 - L)^4}{F(L)} y_{t+2}. \tag{16}$$

As noted by King and Rebelo, obtaining the cyclical component thus amounts to taking fourth differences of the original y_{t+2} and applying the operator $[F(L)]^{-1}$ to the result, so that the HP cycle should produce a stationary series as long as fourth-differences of the original series are stationary.[10]

[9] The other parameters for this case are $C = 0.056075$, $m = 0.111687$ and $\cot(m) = 8.9164$.

[10] De Jong and Sakarya (2016) and Phillips and Jin (2015) provided more details on the relation between the large T expression (16) and the exact finite T formula (2) and the conditions under which the HP cyclical component is a weakly dependent series. Phillips and Jin concluded that for $\lambda = 1600$ and typical sample sizes, the HP filter may not successfully remove the trend even if the true series is only $I(1)$.

3 Consequences of using the HP filter.

The presumption by users of the HP filter is that it offers a reasonable approach to detrending for a range of commonly encountered economic time series. The leading example of a time-series process for which we would want to be particularly convinced of the procedure's appropriateness would be a random walk. Simple economic theory suggests that variables such as stock prices (Fama, 1965), futures prices (Samuelson, 1965), long-term interest rates (Sargent, 1976; Pesando, 1979), oil prices (Hamilton, 2009), consumption spending (Hall, 1978), inflation, tax rates, and money supply growth rates (Mankiw, 1987) should all follow martingales or near martingales. To be sure, hundreds of studies have claimed to find evidence of statistically detectable departures from pure martingale behavior.[11] Even so, there is indisputable evidence that a random walk is often extremely hard to beat in out-of-sample forecasting comparisons, as has been found for example by Meese and Rogoff (1983) and Cheung, Chinn, and Pascual (2005) for exchange rates, Flood and Rose (2010) for stock prices, Atkeson and Ohanian (2001) for inflation, or Balcilar, et al. (2015) for GDP, among many others. Certainly if we are not comfortable with the consequences of applying the HP filter to a random walk, then we should not be using it as an all-purpose approach to economic time series.

For $y_t = y_{t-1} + \varepsilon_t$, where ε_t is white noise and $(1 - L)y_t = \varepsilon_t$, Cogley and Nason (1995)[12] noted that expression (16) means that when the HP filter is applied to a random walk, the

[11] See for example Flavin (1981) on consumption, Baumeister and Kilian (2015) on oil prices, and Bauer and Hamilton (2016) on long-term interest rates, among many, many others.

[12] Harvey and Jaeger (1993) also have a related discussion.

cyclical component will be characterized by

$$c_t = \frac{\lambda(1-L)^3}{F(L)}\varepsilon_{t+2}.$$

For $\lambda = 1600$ this is

$$c_t = 89.72\left\{-q_{0,t+2} + \textstyle\sum_{j=0}^{\infty}(0.8941)^j[\cos(0.1117j) + 8.916\sin(0.1117j)](q_{1,t+2-j} + q_{2,t+2+j})\right\}$$

with $q_{0t} = \varepsilon_t - 3\varepsilon_{t-1} + 3\varepsilon_{t-2} - \varepsilon_{t-3}$, $q_{1t} = \varepsilon_t - 3.79\varepsilon_{t-1} + 5.37\varepsilon_{t-2} - 3.37\varepsilon_{t-3} + 0.79\varepsilon_{t-4})$,[13] and $q_{2t} = -0.79\varepsilon_{t+1} + 3.37\varepsilon_t - 5.37\varepsilon_{t-1} + 3.79\varepsilon_{t-2} - \varepsilon_{t-3}$. The underlying innovations ε_t are completely random and exhibit no patterns, whereas the series c_t is both highly predictable (as a result of the dependence on lags of ε_{t-j}) and will in turn predict the future (as a result of dependence on future values of ε_{t+j}). Since the coefficients that make up $[F(L)]^{-1}$ are determined solely by the value of λ, these patterns in the cyclical component are entirely a feature of having applied the HP filter to the data rather than reflecting any true dynamics of the data-generating process itself.

For example, consider the behavior of stock prices and real consumption spending.[14] The top panels of Figure 2 show the autocorrelation functions for first-differences of these series, confirming that there is little ability to predict either from its own past values, as we might have expected from the literature cited at the start of this section. The lower panels show cross correlations. Consumption has no predictive power for stocks, though stock prices may have a modest ability to anticipate changes in aggregate consumption.

[13] The term q_{1t} is the expansion of $(1-L)^3[1-(\phi_1^2/4)L]\varepsilon_t$.

[14] Stock prices were measured as 100 times the natural log of the end-of-quarter value for the S&P 500 and consumption from 100 times the natural log of real personal consumption expenditures from the U.S. NIPA accounts. All data for this figure are quarterly for the period 1950:1 to 2016:1.

Figure 3 shows the analogous results if we tried to remove the trend by HP filtering rather than first-differencing. The HP cyclical components of stock prices and consumption are both extremely predictable from their own lagged values as well as each other. The rich dynamics in these series are purely an artifact of the filter itself and tell us nothing about the underlying data-generating process. Filtering takes us from the very clean understanding of the true properties of these series that we can easily see in Figure 2 to the artificial set of relations that appear in Figure 3. The values plotted in Figure 3 summarize the filter, not the data.

4 A one-sided HP filter.

The HP trend and cycle have an artificial ability to "predict" the future because they are by construction a function of future realizations. One way we might try to get around this would be to restrict the minimization problem in (3), forcing a_t to load only on values $(y_t, y_{t-1}, ..., y_1)'$ that have been observed as of date t, rather than also using future values as was done in the HP filter (4). This restricted solution is in fact easy to calculate using popular software packages. We again could use the state-space model assumed in Proposition 1 with C_0 large and $\sigma_c^2/\sigma_v^2 = 1600$. Whereas the Kalman smoother would yield the two-sided linear projection, which is numerically identical to the usual HP filter, the Kalman filter gives the one-sided linear projection.

The top panel of Figure 4 shows the result of applying the usual two-sided HP filter to stock prices. The trend is identified to have been essentially flat throughout the 2000s, with the pre-recession booms and post-recession busts in stock prices viewed entirely as cyclical phenomena. The bottom panel shows the results of applying a one-sided HP filter to the same data. This would instead label the trend component as rising during economic expansions and falling during recessions. The reason is that a real-time observer would not know in early 2009, for example, that stock prices were about to appreciate remarkably, and accordingly would have judged much of the drop observed up to that date as permanent. It is only with hindsight that we are tempted to label the 2008 stock-market crash as a temporary phenomenon. Making use of unknowable future values in this way is in fact a fundamental reason that HP-filtered series exhibit the visual properties that they do, precisely because they impose patterns that are not a feature of the data-generating process and could not be recognized in real time.

Moreover, although a one-sided filter would eliminate the problem of generating a series that is artificially able to predict the future, changes in both the one-sided trend and its implied cycle are readily forecastable from their own lagged values, and likewise by values of any other variables. Again this is not a feature of the stock prices themselves, but instead is an artifact of choosing to characterize the cycle and trend in this particular way.

5 Estimating λ by quasi-maximum likelihood.

A separate question is what value we should use for the smoothing parameter λ. Hodrick and Prescott motivated their choice of $\lambda = 1600$ based on the prior belief that a large change in the cyclical component within a quarter would be around 5%, whereas a large change in the trend component would be around (1/8)%, suggesting a choice of $\lambda = \sigma_c^2/\sigma_v^2 = (5/(1/8))^2 = 1600$. Ravn and Uhlig (2002) showed how to choose the smoothing parameter for data at other frequencies if indeed it would be correct to use 1600 on quarterly data. These rules of thumb are almost universally followed.

It's worth noting that if the state-space model in Proposition 1 were indeed an accurate characterization of the trend that we were trying to infer, we would not need to make up a value for λ but could in fact estimate it from the data. If for example we assumed a Normal distribution for the innovations $(v_t, c_t)'$ we could use the Kalman filter to evaluate the likelihood function for the observed sample $(y_1, ..., y_T)'$ and find the values for σ_v^2 and σ_c^2 that maximize the likelihood function.[15] This could alternatively be given a quasi-maximum likelihood interpretation as a GLS minimization of the squared forecast errors weighted by their model-implied variance.

Table 1 reports MLEs of σ_v^2, σ_c^2, and λ for a number of commonly studied macroeconomic series. For every one of these we would estimate a value for σ_c^2 whose magnitude is similar to, and in fact often smaller than, σ_v^2, and certainly not 1600 times as large. If we used a value of $\lambda = 1$ instead of $\lambda = 1600$, the resulting series for g_t would differ little from the original data y_t itself; $\lambda = 1$ implies a value for R in expression (11) of 0.48, which decays with a half-life of less than one quarter.

Thus not only is the HP filter very inappropriate if the true process is a random walk. As

[15] See for example Hamilton (1994, equations [13.4.1]-[13.4.2]). Note that although the inferred value for the trend g_t depends only on the ratio σ_c^2/σ_v^2, the parameters σ_c^2 and σ_v^2 are separately identifiable because σ_c^2 can be inferred from the average observed size of $(y_t - g_t)^2$.

commonly applied, the HP filter is not even optimal for the only example for which anyone has claimed that it might provide the ideal inference!

6 A better alternative.

Here I suggest an alternative concept of what we might mean by the cyclical component of a trending series: how different is the value at date $t+h$ from the value that we would have expected to see based on its behavior through date t?[16] This concept of the cyclical component has several attractive features. First, as noted by den Haan (2000), the forecast error is stationary for a wide class of nonstationary processes. Second, the primary reason that we would be wrong in predicting the value of most macro and financial variables at a horizon of $h = 8$ quarters ahead is cyclical factors such as whether a recession occurs over the next two years and the timing of recovery from any downturn.[17]

While it might seem that calculating this concept of the cyclical component requires us already to know the nature of the trend and to have the correct model for forecasting the series, neither of these is the case. We can instead always rely on very simple forecasts within a restricted class, namely, the population linear projection of y_{t+h} on a constant and the 4 most recent values of y as of date t. This object exists and can be consistently estimated for a wide range of nonstationary processes, as I now show.

Note first that as long as the dth difference of y_t is stationary, we can write the value

[16] This definition corresponds to the Beveridge-Nelson (1981) characterization of the cyclical component of a time series for the special case when y_t is assumed to be an $I(1)$ process and the forecast horizon h and number of conditioning observations as of date t (denoted p below) both go to infinity.

[17] This same consideration suggests using $h = 24$ for monthly data and $h = 2$ for annual data.

of y_{t+h} as a linear function of initial conditions at time t plus a stationary process. For example, when $d = 1$, letting $u_t = \Delta y_t$ we can write

$$y_{t+h} = y_t + w_t^{(h)} \tag{17}$$

where the stationary component is given by $w_t^{(h)} = u_{t+1} + \cdots + u_{t+h}$. For $d = 2$ and $\Delta^2 y_t = u_t$, we have

$$y_{t+h} = y_t + h\Delta y_t + w_t^{(h)} \tag{18}$$

where now $w_t^{(h)} = u_{t+h} + 2u_{t+h-1} + \cdots + hu_{t+1}$. This result holds for general d, as demonstrated in the following proposition.

Proposition 3. *If $(1-L)^d y_t$ is stationary for some $d \geq 1$, then for all finite $h \geq 1$,*

$$y_{t+h} = \kappa_h^{(1)} y_t + \kappa_h^{(2)} \Delta y_t + \cdots + \kappa_h^{(d)} \Delta^{d-1} y_t + w_t^{(h)}$$

with $\Delta^s = (1-L)^s$, $\kappa_\ell^{(1)} = 1$ for $\ell = 1, 2, ..$ and $\kappa_j^{(s)} = \sum_{\ell=1}^{j} \kappa_\ell^{(s-1)}$ for $s = 2, 3, ..., d$ and $w_t^{(h)}$ is a stationary process.

It further turns out that if $\Delta^d y_t \sim I(0)$ and we regress y_{t+h} on a constant and the d most recent values of y as of date t, the coefficients will be forced to be close to the values implied by the coefficients $\kappa_h^{(j)}$ in Proposition 3. For example, if $\Delta^2 y_t$ is $I(0)$ then in a regression of y_{t+h} on $(y_t, y_{t-1}, 1)'$, the fitted values will tend to $y_t + h(y_t - y_{t-1}) + \mu_h$ for $\mu_h = E(w_t^{(h)})$ as the sample size gets large; that is, the coefficient on y_t will go to $1 + h$ and the coefficient on y_{t-1} will go to $-h$. The implication is that the residuals from a regression of y_{t+h} on $(y_t, y_{t-1}, 1)'$ will be stationary whenever y itself is $I(2)$. The reason is that any other values for these coefficients would imply a nonstationary series for the residuals, whose sum of squares become arbitrarily large relative to those implied by the coefficients $1 + h$ and $-h$ as the sample size grows large.

If $\Delta^d y_t$ is stationary and we regress y_{t+h} on a constant and the p most recent values of y as of date t for any $p > d$, the regression will use d of the coefficients to make sure the residuals are stationary and the remaining $p + 1 - d$ coefficients will be determined by the parameters that characterize the population linear projection of the stationary variable $w_t^{(h)}$ on the stationary regressors $(\Delta^d y_t, \Delta^d y_{t-1}, ..., \Delta^d y_{t-p+d+1}, 1)'$. The following proposition provides a formal statement of these claims. In the proof of this proposition I have followed Stock (1994, p. 2756) in defining a series u_t to be $I(0)$ if it has fixed mean μ and satisfies a Functional Central Limit Theorem.[18] This requires that the sample mean of u_t has a Normal distribution as the sample size T gets large, as does a sample mean that used only Tr observations for $0 < r \leq 1$. Formally,

$$T^{-1/2} \sum_{s=1}^{[Tr]} (u_t - \mu) \Rightarrow \omega W(r), \tag{19}$$

where $[Tr]$ denotes the largest integer less than or equal to Tr, $W(r)$ denotes Standard Brownian Motion, and "$\Rightarrow$" denotes weak convergence in probability measure. I will show that if either the dth difference $(u_t = \Delta^d y_t)$ satisfies (19) or if the deviation from a dth-order deterministic polynomial in time $(u_t = y_t - \delta_0 - \delta_1 t - \delta_2 t^2 - \cdots - \delta_d t^d)$ satisfies (19), then we can remove the nonstationary component with the same simple regression.[19]

[18] Stock (1994, p. 2749) demonstrated that an example of sufficient conditions that imply (19) is that $u_t = \mu + \sum_{j=0}^{\infty} \psi_j \eta_t$ where η_t is a martingale difference sequence with variance σ^2 and finite fourth moment, $\psi(1) \neq 0$, and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$, in which case ω^2 in (19) is given by $\sigma^2[\psi(1)]^2$. Alternatively, Phillips (1987, Lemma 2.2) derived (19) from primitive moment and mixing conditions on u_t.

[19] The reason to state these as two separate possibilities is that if the nonstationarity is purely deter-

Proposition 4. *Suppose that either* $u_t = \Delta^d y_t$ *satisfies* (19) *or that* $u_t = y_t - \sum_{j=0}^{d} \delta_j t^j$ *with* $\delta_d \neq 0$ *satisfies* (19) *for some unknown* d. *Let* $x_t = (y_t, y_{t-1}, ..., y_{t-p+1}, 1)'$ *for some* $p \geq d$ *and consider OLS estimation of* $y_{t+h} = x_t'\beta + v_{t+h}$ *for* $t = 1, ..., T$ *with estimated coefficient*

$$\hat{\beta} = \left(\sum_{j=1}^{T} x_t x_t'\right)^{-1} \left(\sum_{j=1}^{T} x_t y_{t+h}\right). \tag{20}$$

If $p = d$, *the OLS residuals* $y_{t+h} - x_t'\hat{\beta}$ *converge to the variable* $w_t^{(h)} - E(w_t^{(h)})$ *in Proposition 3. If* $p > d$, *the OLS residuals converge to the residuals from a population linear projection of* $w_t^{(h)}$ *on* $(\Delta^d y_t, \Delta^d y_{t-1}, ..., \Delta^d y_{t-p+d+1}, 1)'$.

Proposition 4 establishes that if we estimate an OLS regression of y_{t+h} on a constant and the $p = 4$ most recent values of y as of date t,

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+h}, \tag{21}$$

the residuals

$$\hat{v}_{t+h} = y_{t+h} - \hat{\beta}_0 - \hat{\beta}_1 y_t - \hat{\beta}_2 y_{t-1} - \hat{\beta}_3 y_{t-2} - \hat{\beta}_4 y_{t-3} \tag{22}$$

offer a reasonable way to remove an unknown trend for a broad class of underlying processes. Like the HP filter, this will take out the trend provided that fourth differences of y_t are stationary. But whereas the HP filter *imposes* all 4 unit roots in equation (16), the sample regression would only use 4 differences if it is warranted by observed features of the data. The proposed procedure has a number of other advantages over HP. First, any finding that

ministic, then the dth differences will not satisfy the Functional Central Limit Theorem. For example, if $y_t = \gamma_0 + \gamma_1 t + \varepsilon_t$ with ε_t white noise, then $\Delta y_t = \gamma_1 + \psi(L)\varepsilon_t$ for $\psi(L) = 1 - L$ and $\psi(1) = 0$. Of course when $u_t = \Delta^d y_t$ satisfies (19) with $\mu \neq 0$, the series y_t has both dth-order stochastic as well as dth-order deterministic polynomial trends, so that case, along with pure stochastic trends ($\mu = 0$) and pure deterministic trends are all allowed by Proposition 4.

$\hat{v}_{t+h}$ predicts some other variable x_{t+h+j} represents a true ability of y to predict x rather than an artifact of the way we chose to detrend y, by virtue of the fact that $\hat{v}_{t+h}$ is a one-sided filter. Second unlike the HP cyclical series c_{t+h}, the value of $\hat{v}_{t+h}$ will by construction be difficult to predict at time t. If we find such predictability, it tells us something about the true data-generating process, for example, that x Granger-causes y. Third, the value of $\hat{v}_{t+h}$ is a model-free and essentially assumption-free summary of the data. Regardless of how the data may have been generated, as long as $(1-L)^d y_t$ is covariance stationary for some $d \leq 4$, there exists a population linear projection of y_{t+h} on $(y_t, y_{t-1}, y_{t-2}, y_{t-3}, 1)'$. That projection is a characteristic of the data-generating process that can be used to define what we mean by the cyclical component of the process and can be consistently estimated from the data. Given a dynamic stochastic general equilibrium or any other theoretical model that would imply an $I(d)$ process, we could calculate this population characteristic of the model and estimate it consistently from the data.

Given the literature cited at the beginning of Section 3 it is instructive to examine the consequences if this procedure were applied to a random walk: $y_t = y_{t-1} + \varepsilon_t$. In this case, $d = 1$ and $w^{(1)}_{t+h} = \varepsilon_{t+h} + \varepsilon_{t+h-1} + \cdots + \varepsilon_{t+1}$. For large samples, the OLS estimates of (21) converge to $\beta_1 = 1$ and all other $\beta_j = 0$, and the resulting filtered series would simply be the difference

$$\tilde{v}_{t+h} = y_{t+h} - y_t, \tag{23}$$

that is, how much the series changes over an $h = 8$-quarter horizon, or equivalently the sum of the observed changes over h periods. Note that for $h = 8$ the filter $1 - L^h$ wipes out any cycles with frequency of exactly one year, and thus is taking out both the long-run trend as well as any strictly seasonal components.[20] This also fits with the common understanding of what we would mean by the cyclical component. Because the simple filter (23) does not require estimation of any parameters, it can also be used as a quick robustness check for concerns about the small-sample applicability of the asymptotic claims in Proposition 4, as will be illustrated in the applications below.

Another instructive example is a pure deterministic time trend of order $d = 1$: $y_t = \delta_0 + \delta_1 t + \varepsilon_t$ for ε_t white noise. In this case $\Delta y_t = \delta_1 + \varepsilon_t - \varepsilon_{t-1}$ is stationary and $w_t^{(h)} = \Delta y_{t+1} + \cdots + \Delta y_{t+h} = \delta_1 h + \varepsilon_{t+h} - \varepsilon_t$ is also stationary for any h. I show in the appendix that for this case the limiting coefficients on $y_t, .., y_{t-p+1}$ described by Proposition 4 are each given by $1/p$ and the implied trend for y_{t+h} is

$$\delta_0 + \delta_1(t+h) + p^{-1}(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-p+1}). \tag{24}$$

Even for $p = 1$ this is not a bad estimate and for $p = 4$ should not differ much from the true trend $\delta_0 + \delta_1(t+h)$. Again regardless of the choice of p, the difference between y_{t+h} and (24) will be stationary.

A third instructive example is when y_t is an element of a theoretical dynamic stochastic general equilibrium model that is stationary around some steady-state value μ. If the effects of shocks in the theoretical model die out after h periods, then the linear projection (21) in the theoretical model is characterized by $\beta_0 = \mu$ and $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. In other

[20] As in Hamilton (1994, pp. 171-172), the filter $1 - L^8$ has power transfer function $(1 - e^{-8i\omega})(1 - e^{8i\omega}) = 2 - 2\cos(8\omega)$ which is zero at $\omega = 0, \pi/4, \pi/2, 3\pi/4, \pi$ and thus eliminates not only cycles at the zero frequency but also cycles that repeat themselves every 8,4,8/3, or 2 quarters. See also Hamilton (1994, Figures 6.5 and 6.6).

words, the component v_{t+h} is exactly the deviation from the steady state. If shocks have not completely died out after h periods, then part of what is being labeled trend by this method would include the components of shocks that persist longer than h periods. But for any value of h, the linear projection is a well-defined population characteristic of the theoretical stationary model, and there is an exactly analogous object one can calculate in the possibly nonstationary observed data. The method thus offers a way to make an apples-to-apples comparison of theory with data of the sort that users of the HP filter often desire, but which the HP filter itself will always fail to deliver.

Figure 5 shows the results when this approach is applied to data on U.S. total employment. The raw seasonally adjusted data (y_t) are plotted in the upper left panel. The residuals from regression (21) estimated for these data are plotted in black in the lower-left panel, while the 8-lag difference (23) is in red. The latter two series behave very similarly in this case, as indeed I have found for most other applications. The primary difference is that the regression residual has sample mean zero by construction (by virtue of the inclusion of a constant term in the regression) whereas the average value of (23) will be the average growth rate over a two-year period.

One interesting observation is that the cyclical component of employment starts to decline significantly before the NBER business cycle peak for essentially every recession. Note that this inference from Figure 5 is summarizing a true feature of the data and is not an artifact of any forward-looking aspect of the filter.

The right panels of Figure 5 show what happens when the same procedure is applied to seasonally unadjusted data. The raw data themselves exhibit a very striking seasonal pattern, as seen in the top right panel. Notwithstanding, the cyclical factor inferred from seasonally unadjusted data (bottom right panel) is almost indistinguishable from that derived from seasonally adjusted data, confirming that this approach is robust to methods of seasonal adjustment.

Figure 6 applies the method to the major components of the U.S. national income and product accounts. Investment spending is more cyclically volatile than GDP, while consumption spending is less so. Imports fall significantly during recessions, reflecting lower spending by U.S. residents on imported goods, and exports substantially less so, reflecting the fact that international downturns are often decoupled from those in the U.S. Detrended government spending is dominated by war-related expenditures– the Korean War in the early 1950s, the Vietnam War in the 1970s, and the Reagan military build-up in the 1980s.

Table 2 reports the standard deviation of the cyclical component of each of these and a number of other series, along with their correlation with the cyclical component of GDP. We find very little cyclical correlation between output and prices.[21] Both the nominal fed funds rate and the ex ante real fed funds rate (the latter based on the measure in Hamilton, et al., 2015) are modestly procyclical, whereas the 10-year nominal interest rate is not.

[21] Identifying the sign of this correlation was one of the primary interests of den Haan's (2000) application of a related methodology. In contrast to the results in Table 2, he found a positive correlation between the cyclical components of these series. I attribute the difference to differences in sample period.

7 Conclusion.

The HP filter will construct a stationary component from any $I(4)$ series, but at a great cost. It introduces spurious dynamic relations that are purely an artifact of the filter and have no basis in the true data-generating process, and there exists no plausible data-generating process for which common popular practice would provide an optimal decomposition into trend and cycle. There is an alternative approach that can also isolate a stationary component from any $I(4)$ series but that preserves the underlying dynamic relations and consistently estimates well defined population characteristics for a broad class of possible data-generating processes.

References

Atkeson, Andrew and Lee E. Ohanian. 2001. "Are Phillips Curves Useful for Forecasting Inflation?," *Quarterly Review*, Federal Reserve Bank of Minneapolis, 25(1): 2-11.

Balcilar, Mehmet, Rangan Gupta, Anandamayee Majumdar and Stephen M. Miller. 2015. "Was the Recent Downturn in US Real GDP Predictable?," *Applied Economics,* 47(28): 2985-3007.

Bauer, Michael D. and James D. Hamilton. 2016. "Robust Bond Risk Premia," UCSD working paper.

Baumeister, Christiane, and Lutz Kilian. 2015. "Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach," *Journal of Business and Economic Statistics*, 33(3): 338-351.

Beveridge, Stephen, and Charles R. Nelson. 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.

Cheung, Yin-Wong, Menzie D. Chinn & Antonio Garcia Pascual. 2005. "Empirical exchange rate models of the nineties: Are any fit to survive?," *Journal of International Money and Finance,* 24(7): 1150-1175.

Choi, In. 1993. "Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with Some Applications," *Econometric Theory*, 9:263-282.

Cogley, Timothy, and James M. Nason. 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(1-2): 253-278.

Cornea-Madeira, Adriana. Forthcoming. "The Explicit Formula for the Hodrick-Prescott Filter in Finite Sample," *Review of Economics and Statistics.*

De Jong, Robert M., and Neslihan Sakarya. 2016. "The Econometrics of the Hodrick-Prescott Filter," *Review of Economics and Statistics*, 98: 310–317.

Den Haan, Wouter J. 2000. "The Comovement between Output and Prices," *Journal of Monetary Economics* 46: 3-30.

Fama, Eugene F. 1965. "The Behavior of Stock-Market Prices," *The Journal of Business*, 38(1): 34-105.

Flavin, Marjorie A. 1981. "The Adjustment of Consumption to Changing Expectations About Future Income," *The Journal of Political Economy,* 89(5): 974-1009.

Flood, Robert P. and Andrew K. Rose. 2010. "Forecasting International Financial Prices with Fundamentals: How do Stocks and Exchange Rates Compare?," *Globalization and Economic Integration*, Chapter 6, Edward Elgar Publishing.

Hall, Robert E. 1978. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86(6): 971-987.

Hamilton, James D. 1994. *Time Series Analysis,* Princeton: Princeton University Press.

Hamilton, James D. 2009. "Understanding Crude Oil Prices," *Energy Journal,* 30(2): 179-206.

Hamilton, James D., Ethan S. Harris, Jan Hatzius, and Kenneth D. West. 2015. "The

Equilibrium Real Fed Funds Rate: Past, Present, and Future," NBER Working Paper 21476.

Harvey, Andrew C., and Albert Jaeger. 1993. "Detrending, Stylized Facts and the Business Cycle," *Journal of Applied Econometrics*, 8(3): 231-247.

Hodrick, Robert J. and Edward C. Prescott. 1981. "Postwar U.S. Business Cycles: An Empirical Investigation," working paper, Northwestern University.

Hodrick, Robert J. and Edward C. Prescott. 1997. "Postwar U.S. Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29(1): 1-16.

King, Robert and Sergio T. Rebelo. 1989. "Low Frequency Filtering and Real Business Cycles," working paper, University of Rochester.

King, Robert and Sergio T. Rebelo. 1993. "Low Frequency Filtering and Real Business Cycles," *Journal of Economic Dynamics and Control,* 17: 207-231.

Mankiw, N. Gregory. 1987. "The Optimal Collection of Seigniorage: Theory and Evidence," *Journal of Monetary Economics*, 20(2): 327-341.

McElroy, Tucker. 2008. "Exact Formulas for the Hodrick-Prescott Filter," *Econometrics Journal*, 11(1): 209-217.

Meese, Richard A. and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies : Do they Fit Out of Sample?," *Journal of International Economics,* 14(1-2): 3-24.

Park, Joon Y., and Peter C.B. Phillips. 1989. "Statistical Inference in Regressions with Integrated Processes: Part 2," *Econometric Theory*, 5:95-131.

Phillips, Peter C.B. 1987. "Time Series Regression with a Unit Root," *Econometrica*, 55: 277-301.

Phillips, Peter C.B., and Sainan Jin. 2015. "Business Cycles, Trend Elimination, and the HP Filter," working paper, Yale University.

Pesando, James E. 1979. "On the Random Walk Characteristics of Short- and Long-Term Interest Rates in an Efficient Market," *Journal of Money, Credit and Banking,* Blackwell Publishing, 11(4): 457-466.

Ravn, Morten O., and Harald Uhlig. 2002. "On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations," *Review of Economics and Statistics*, 84(2): 371-380.

Samuelson, P. 1965. "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review,* 6(2): 41-49.

Sargent, Thomas J. 1976. "A Classical Macroeconometric Model for the United States," *Journal of Political Economy,* 84(2): 207-237.

Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica* 58: 113-144.

Stock, James H. 1994. "Unit Roots, Structural Breaks, and Trends," in *Handbook of Econometrics*, Volume 4, pp. 2739-2841, edited by Robert F. Engle and Daniel L. McFadden. Amsterdam: Elsevier.

Appendix.

Proof of Proposition 1.

The assumptions about c_t, v_t, and the initial states can be written formally as

$$E(v_t) = E(c_t) = 0 \tag{25}$$

$$E\begin{bmatrix} v_t \\ c_t \end{bmatrix}\begin{bmatrix} v_{t-j} & c_{t-j} \end{bmatrix} = \begin{cases} \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & \sigma_c^2 \end{bmatrix} & \text{if } j = 0 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

and for the initial conditions we assume

$$E(g_0) = E(g_{-1}) = 0 \tag{27}$$

$$E\begin{bmatrix} g_0 \\ g_{-1} \end{bmatrix}\begin{bmatrix} g_0 & g_{-1} \end{bmatrix} = C_0 \tag{28}$$

$$E\begin{bmatrix} v_t \\ c_t \end{bmatrix}\begin{bmatrix} g_0 & g_{-1} \end{bmatrix} = 0 \quad \text{for } t = 1, ...T. \tag{29}$$

I first establish that under (5)-(6) and (25)-(29),

$$(Q'Q)E(gg')H' \to \sigma_v^2 H' \tag{30}$$

as $C_0^{-1} \to 0$. To do so write (6) as $Qg = v$ for $v = (v_T, v_{T-1}, ..., v_1)$ and $Q_0 g = v_0$ for v_0 a (2×1) vector with mean 0 and variance $\sigma_v^2 I_2$. Also from (29), v_0 is uncorrelated with v and

$$\underset{(2\times\tilde{T})}{Q_0} = \begin{bmatrix} \underset{(2\times T)}{0} & \underset{(2\times 2)}{P_0^{-1}} \end{bmatrix}$$

where P_0 is the Cholesky factor of C_0 ($P_0P_0' = C_0$). Stacking these,

$$\begin{bmatrix} Q \\ Q_0 \end{bmatrix} g = \begin{bmatrix} v \\ v_0 \end{bmatrix}$$

so

$$E(gg') = \sigma_v^2 \begin{bmatrix} Q \\ Q_0 \end{bmatrix}^{-1} \begin{bmatrix} Q' & Q_0' \end{bmatrix}^{-1}$$

$$\begin{bmatrix} Q' & Q_0' \end{bmatrix} \begin{bmatrix} Q \\ Q_0 \end{bmatrix} E(gg') = (Q'Q + Q_0'Q_0)E(gg') = \sigma_v^2 I_{\tilde{T}}$$

$$(Q'Q)E(gg')H' = \sigma_v^2 H' - (Q_0'Q_0)E(gg')H'$$

which goes to $\sigma_v^2 H'$ as $P_0^{-1} \to 0$, as claimed in (30).

Notice next from

$$\underset{(T\times 1)}{y} = \underset{(T\times\tilde{T})}{H}\underset{(\tilde{T}\times 1)}{g} + \underset{(T\times 1)}{c}$$

that $E(yy') = HE(gg')H' + \sigma_c^2 I_T$ and $E(gy') = E(gg')H' + E(gc') = E(gg')H'$. Hence

$$\begin{aligned} \tilde{A} &= E(gy')[E(yy')]^{-1} \\ &= E(gg')H'[HE(gg')H' + \sigma_c^2 I_T]^{-1}. \end{aligned} \tag{31}$$

Combining (2) and (31),

$$\begin{aligned} & (H'H + \lambda Q'Q)(A^* - \tilde{A})[HE(gg')H' + \sigma_c^2 I_T] \\ = {} & H'[HE(gg')H' + \sigma_c^2 I_T] - (H'H + \lambda Q'Q)E(gg')H' \\ = {} & H'\sigma_c^2 - (\sigma_c^2/\sigma_v^2)(Q'Q)E(gg')H' \end{aligned} \tag{32}$$

which from (30) goes to 0 as $C_0^{-1} \to 0$. Since the matrices premultiplying and postmultiplying the left side of (32) are of full rank, this establishes that $A^* = \tilde{A}$ as claimed.

Proof of Proposition 2. Let $\theta_1, \theta_2, \theta_3, \theta_4$ be the roots satisfying $F(\theta_i) = 0$. As noted by King and Rebelo (1989), since $\lambda > 0$, $F(z)$ in (9) is positive for all real z meaning that θ_i comprise two pairs of complex conjugates. Since $F(z) = F(z^{-1})$, if θ_i is a root, then so is θ_i^{-1}. Thus the values of θ_i are given by $Re^{im}, Re^{-im}, R^{-1}e^{im}$, and $R^{-1}e^{-im}$ for some fixed R and m; one pair is inside the unit circle and the other is outside. Noting that the coefficients on z^2 and z^{-2} in $F(z)$ are both λ, it follows that $F(z)$ can be written

$$F(z) = \lambda(1 - \theta_1 z)(1 - \theta_2 z)(\theta_1^{-1} - z^{-1})(\theta_2^{-1} - z^{-1}).$$

From the symmetry of $F(z)$ in z and z^{-1} we can without loss of generality normalize θ_1 and θ_2 to be inside the unit circle and write

$$F(z) = \frac{\lambda}{\theta_1 \theta_2}(1 - \theta_1 z)(1 - \theta_2 z)(1 - \theta_1 z^{-1})(1 - \theta_2 z^{-1}).$$

Define $(1 - \phi_1 z - \phi_2 z^2) = (1 - \theta_1 z)(1 - \theta_2 z)$, namely ϕ_1 is the real number $\theta_1 + \theta_2$ and ϕ_2 is the negative real number $-\theta_1 \theta_2$. Note also that the roots of $(1 - \phi_1 z - \phi_2 z^2) = 0$ are the complex conjugates θ_1^{-1} and θ_2^{-1}, which are both outside the unit circle. This gives the bounds on ϕ_1 and ϕ_2 stated in Proposition 2 as in Hamilton (1994, Figure 1.5), and allows us to write

$$F(z) = \frac{\lambda}{-\phi_2}(1 - \phi_1 z - \phi_2 z^2)(1 - \phi_1 z^{-1} - \phi_2 z^{-2}). \tag{33}$$

Evaluating (9) and (33) at $z = 1$ gives

$$F(1) = 1 = (1 - \phi_1 - \phi_2)^2 \lambda / (-\phi_2) \tag{34}$$

as claimed in (13), Likewise evaluating (9) and (33) at $z = -1$ gives

$$F(-1) = 1 + 16\lambda = (1 + \phi_1 - \phi_2)^2 \lambda/(-\phi_2). \tag{35}$$

Taking the difference between these last two equations establishes $(4\phi_1 - 4\phi_1\phi_2)\lambda/(-\phi_2) = 16\lambda$ or $\phi_1(1 - \phi_2) = -4\phi_2$ as claimed in (12). Note that since $\phi_2 < 0$ (required by complex roots), from (12) $\phi_1 > 0$.

I next establish that

$$\frac{1}{(1 - \phi_1 z - \phi_2 z^2)(1 - \phi_1 z^{-1} - \phi_2 z^{-2})} = \frac{C_0 + C_1 z}{1 - \phi_1 z - \phi_2 z^2} + \frac{C_0 + C_1 z^{-1}}{1 - \phi_1 z^{-1} - \phi_2 z^{-2}} + B_0. \tag{36}$$

Combining terms on the right-hand side over a common denominator shows that (36) will hold provided

$$\begin{aligned} 1 &= (C_0 + C_1 z)(1 - \phi_1 z^{-1} - \phi_2 z^{-2}) + (C_0 + C_1 z^{-1})(1 - \phi_1 z^1 - \phi_2 z^2) \\ &\quad + B_0(1 - \phi_1 z - \phi_2 z^2)(1 - \phi_1 z^{-1} - \phi_2 z^{-2}) \\ &= [2C_0 - 2C_1\phi_1 + B_0(1 + \phi_1^2 + \phi_2^2)] \\ &\quad + [C_1 - C_0\phi_1 - C_1\phi_2 - B_0\phi_1 + B_0\phi_1\phi_2](z + z^{-1}) \\ &\quad - [C_0\phi_2 + B_0\phi_2](z^2 + z^{-2}). \end{aligned}$$

The coefficient on $(z^2 + z^{-2})$ will be zero provided $B_0 = -C_0$. Substituting this back in, we then require

$$1 = [C_0 - 2C_1\phi_1 - C_0\phi_1^2 - C_0\phi_2^2] + [C_1 - C_1\phi_2 - C_0\phi_1\phi_2](z + z^{-1}). \tag{37}$$

The coefficient on $(z + z^{-1})$ will be zero provided

$$C_1 = \frac{C_0\phi_1\phi_2}{1 - \phi_2} = -C_0\phi_1^2/4 \tag{38}$$

where the last equation made use of (12). Substituting (38) into (37), we see that (36) will be true provided we set

$$1 = C_0(1 - \phi_1^2 - \phi_2^2 + \phi_1^3/2).$$

Combining these results we conclude that

$$\frac{1}{(1 - \phi_1 z - \phi_2 z^2)(1 - \phi_1 z^{-1} - \phi_2 z^{-2})} = C_0 \left[\frac{1 - (\phi_1^2/4)z}{1 - \phi_1 z - \phi_2 z^2} + \frac{1 - (\phi_1^2/4)z^{-1}}{1 - \phi_1 z^{-1} - \phi_2 z^{-2}} - 1 \right].$$

From (33) we then obtain (10) with $C = -C_0\phi_2/\lambda$ as claimed in (14).

To derive (11), recall from Hamilton (1994, pp. 16 and 33) that

$$\frac{1}{1 - \phi_1 z - \phi_2 z^2} = \sum_{j=0}^{\infty} R^j [2\alpha \cos(mj) + 2\beta \sin(mj)] z^j. \tag{39}$$

We know that the coefficient on z^j for $j = 0$ must be 1, requiring $[2\alpha \cos(0) + 2\beta \sin(0)] = 1$ or $\alpha = 1/2$. We likewise know that the coefficient on z^j for $j = 1$ is given by ϕ_1, so $R[\cos(m) + 2\beta \sin(m)] = \phi_1$, which from (15) gives $R2\beta \sin(m) = \phi_1/2$ or $2\beta \sin(m) = \cos(m)$ so $2\beta = \cot(m)$. Substituting these values for α and β into (39) gives (11).

Proof of Proposition 3. Recall the identity

$$y_{t+h} = y_t + \sum_{j=1}^{h} \Delta y_{t+j} \tag{40}$$

which immediately gives the result of Proposition 3 for the case $d = 1$ as stated in (17). We likewise have the identity

$$\Delta y_{t+j} = \Delta y_t + \sum_{s=1}^{j} \Delta^2 y_{t+s}. \tag{41}$$

Substituting (41) into (40) gives

$$y_{t+h} = y_t + \Delta y_t \sum_{j=1}^{h} 1 + w_t^{(h)}$$

for $w_t^{(h)} = \sum_{j=1}^{h} \sum_{s=1}^{j} \Delta^2 y_{t+s}$ as claimed in (18) for the case $d = 2$. We can proceed recursively using the identity $\Delta^k y_{t+s} = \Delta^k y_t + \sum_{r=1}^{s} \Delta^{k+1} y_{t+r}$ and substituting into the preceding expression. For any d the resulting $w_t^{(h)}$ is a finite sum of stationary variables and therefore is itself stationary.

Proof of Proposition 4. Note that the fitted values and residuals implied by the coefficients in (20) are numerically identical to those if we were to do the (infeasible) regression $y_{t+h} = \tilde{x}_t' \alpha + v_{t+h}$ for

$$\tilde{x}_t = (\tilde{u}_t, \tilde{u}_{t-1}, ..., \tilde{u}_{t-p+d+1}, 1, \Delta^{d-1} y_t, \Delta^{d-2} y_t, ..., \Delta y_t, y_t)'$$

with $\tilde{u}_t = \Delta^d y_t - \mu$. The latter regression is infeasible because we do not know the true values of μ and d. But because the fitted values are the same, once we find the properties of the second regression, we will also know the properties of the first. For example, for $d = 2$ and $p = 4$,

$$\tilde{x}_t = \begin{bmatrix} 1 & -2 & 1 & 0 & -\mu \\ 0 & 1 & -2 & 1 & \mu \\ 0 & 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ 1 \end{bmatrix} \equiv H x_t \tag{42}$$

and $\hat{\alpha} = \left(\sum H x_t x_t' H'\right)^{-1} \left(\sum H x_t y_{t+h}\right)$ so $\hat{\beta} = H'\hat{\alpha}$ for every sample. When $p = d$ we define the $(p+1) \times 1$ vector as $\tilde{x}_t = (1, \Delta^{d-1} y_t, \Delta^{d-2} y_t, ..., \Delta y_t, y_t)'$, that is, none of the $\tilde{u}_{t-j}$ variables appear in $\tilde{x}_t$ when $p = d$.

Define q to be the $(p+1) \times 1$ vector $q = (0, ..., 0, E(w_t^{(h)}), \kappa_h^{(d)}, \kappa_h^{(d-1)}, ..., \kappa_h^{(1)})'$, so that

$\tilde{w}_t^{(h)} = w_t^{(h)} - E(w_t^{(h)}) = y_{t+h} - \tilde{x}_t' q$ and

$$\begin{aligned}\hat{\alpha} &= \left(\sum \tilde{x}_t \tilde{x}_t'\right)^{-1} \sum \tilde{x}_t (\tilde{x}_t' q + \tilde{w}_t^{(h)}) \\ &= q + \left(\sum \tilde{x}_t \tilde{x}_t'\right)^{-1} \sum \tilde{x}_t \tilde{w}_t^{(h)}. \end{aligned} \tag{43}$$

We first consider the case when (19) holds for $\Delta^d y_t$ when there is further no drift and the initial value for all of the difference processes is zero, namely, the case when $\mu = 0$ and $\Delta^{d-j} y_t = \xi_t^{(j)}$ where $\xi_t^{(1)} = \sum_{j=1}^t \tilde{u}_j$ and $\xi_t^{(s)} = \sum_{j=1}^t \xi_j^{(s-1)}$ for $s = 2, 3, ..., d$. For this case define

$$\Upsilon_T = \begin{bmatrix} T^{1/2} I_{p-d+1} & 0 & 0 & \cdots & 0 \\ 0 & T & 0 & \cdots & 0 \\ 0 & 0 & T^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & T^d \end{bmatrix}. \tag{44}$$

Adapting the approach in Sims, Stock and Watson (1990), we have from (43) that

$$\begin{aligned} T^{-1/2}\Upsilon_T(\hat{\alpha} - q) &= T^{-1/2}\Upsilon_T \left(\sum \tilde{x}_t \tilde{x}_t'\right)^{-1} \sum \tilde{x}_t \tilde{w}_t^{(h)} \\ &= T^{-1/2} \left[\Upsilon_T^{-1} \sum \tilde{x}_t \tilde{x}_t' \Upsilon_T^{-1}\right]^{-1} \Upsilon_T^{-1} \sum \tilde{x}_t \tilde{w}_t^{(h)} \\ &= \left[\Upsilon_T^{-1} \sum \tilde{x}_t \tilde{x}_t' \Upsilon_T^{-1}\right]^{-1} \left[T^{-1/2}\Upsilon_T^{-1} \sum \tilde{x}_t \tilde{w}_t^{(h)}\right]. \end{aligned} \tag{45}$$

Consider first the last term in (45):

$$T^{-1/2}\Upsilon_T^{-1}\sum \tilde{x}_t\tilde{w}_t^{(h)} = \begin{bmatrix} T^{-1}\sum \tilde{u}_t\tilde{w}_t^{(h)} \\ \vdots \\ T^{-1}\sum \tilde{u}_{t-p+d+1}\tilde{w}_t^{(h)} \\ T^{-1}\sum \tilde{w}_t^{(h)} \\ T^{-3/2}\sum \xi_t^{(1)}\tilde{w}_t^{(h)} \\ T^{-5/2}\sum \xi_t^{(2)}\tilde{w}_t^{(h)} \\ \vdots \\ T^{-d-1/2}\sum \xi_t^{(d)}\tilde{w}_t^{(h)} \end{bmatrix}. \tag{46}$$

The first $p-d$ terms are just the sample means of stationary variables, which by the Law of Large Numbers converge in probability to their expectation $E(\tilde{u}_{t-j}\tilde{w}_t^{(h)})$. Term $p-d+1$ likewise converges to $E(\tilde{w}_t^{(h)}) = 0$. Calculations analogous to those behind Lemma 1(e) in Sims, Stock and Watson (1990) show that the last d terms in (46) also all converge in probability to zero.[22]

Turning next to the first term in (45), the upper-left $(p-d)\times(p-d)$ block of $\Upsilon_T^{-1}\sum \tilde{x}_t\tilde{x}_t'\Upsilon_T^{-1}$ is characterized by

$$\begin{bmatrix} T^{-1}\sum \tilde{u}_t^2 & \cdots & T^{-1}\sum \tilde{u}_t\tilde{u}_{t-p+d+1} \\ \vdots & \cdots & \vdots \\ T^{-1}\sum \tilde{u}_{t-p+d+1}\tilde{u}_t & \cdots & T^{-1}\sum \tilde{u}_{t-p+d+1}^2 \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \gamma_0 & \cdots & \gamma_{p-d-1} \\ \vdots & \cdots & \vdots \\ \gamma_{p-d-1} & \cdots & \gamma_0 \end{bmatrix}$$

for $\gamma_j = E(\tilde{u}_t\tilde{u}_{t-j})$. From Sims, Stock and Watson Lemma 1(a) and 1(b), the lower-right

[22] That is, before multiplying by $T^{-1/2}$ the terms are all $O_p(1)$; for similar calculations see Lemma 1(b) in Choi (1993) and Proposition 17.3(e) in Hamilton (1994).

$(d+1)\times(d+1)$ block satisfies

$$\begin{bmatrix} 1 & T^{-3/2}\sum\xi_t^{(1)} & T^{-5/2}\sum\xi_t^{(2)} & \cdots & T^{-d-1/2}\sum\xi_t^{(d)} \\ T^{-3/2}\sum\xi_t^{(1)} & T^{-2}\sum[\xi_t^{(1)}]^2 & T^{-3}\sum\xi_t^{(1)}\xi_t^{(2)} & \cdots & T^{-d-1}\sum\xi_t^{(1)}\xi_t^{(d)} \\ T^{-5/2}\sum\xi_t^{(2)} & T^{-3}\sum\xi_t^{(2)}\xi_t^{(1)} & T^{-4}\sum[\xi_t^{(2)}]^2 & \cdots & T^{-d-2}\sum\xi_t^{(2)}\xi_t^{(d)} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ T^{-d-1/2}\sum\xi_t^{(d)} & T^{-d-1}\sum\xi_t^{(d)}\xi_t^{(1)} & T^{-d-2}\sum\xi_t^{(d)}\xi_t^{(2)} & \cdots & T^{-2d}\sum[\xi_t^{(d)}]^2 \end{bmatrix}\Rightarrow$$

$$\begin{bmatrix} 1 & \omega\int_0^1 W^{(1)}(r)dr & \omega\int_0^1 W^{(2)}(r)dr & \cdots & \omega\int_0^1 W^{(d)}(r)dr \\ \omega\int_0^1 W^{(1)}(r)dr & \omega^2\int_0^1[W^{(1)}(r)]^2dr & \omega^2\int_0^1 W^{(1)}(r)W^{(2)}(r)dr & \cdots & \omega^2\int_0^1 W^{(1)}(r)W^{(d)}(r)dr \\ \omega\int_0^1 W^{(2)}(r)dr & \omega^2\int_0^1 W^{(2)}(r)W^{(1)}(r)dr & \omega^2\int_0^1[W^{(2)}(r)]^2dr & \cdots & \omega^2\int_0^1 W^{(2)}(r)W^{(d)}(r)dr \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \omega\int_0^1 W^{(d)}(r)dr & \omega^2\int_0^1 W^{(d)}(r)W^{(1)}(r)dr & \omega^2\int_0^1 W^{(d)}(r)W^{(2)}(r)dr & \cdots & \omega^2\int_0^1[W^{(d)}(r)]^2dr \end{bmatrix}$$

where $W^{(1)}(r)$ denotes Standard Brownian Motion and $W^{(j)}(r)=\int_0^r W^{(j-1)}(s)ds$. For the off-diagonal block of $\Upsilon_T^{-1}\sum\tilde{x}_t\tilde{x}_t'\Upsilon_T^{-1}$ we see using calculations analogous to Sims, Stock and Watson's Lemma 1(e) that

$$\begin{bmatrix} T^{-1}\sum\tilde{u}_t & \cdots & T^{-1}\sum\tilde{u}_{t-p+d+1} \\ T^{-3/2}\sum\xi_t^{(1)}\tilde{u}_t & \cdots & T^{-3/2}\sum\xi_t^{(1)}\tilde{u}_{t-p+d+1} \\ T^{-5/2}\sum\xi_t^{(2)}\tilde{u}_t & \cdots & T^{-5/2}\sum\xi_t^{(2)}\tilde{u}_{t-p+d+1} \\ \vdots & \cdots & \vdots \\ T^{-d-1/2}\sum\xi_t^{(d)}\tilde{u}_t & \cdots & T^{-d-1/2}\sum\xi_t^{(d)}\tilde{u}_{t-p+d+1} \end{bmatrix}\xrightarrow{p}0.$$

Bringing all these results together, it follows that

$$T^{-1/2}\Upsilon_T(\hat{\alpha}-q) \xrightarrow{p} \begin{bmatrix} g \\ 0 \end{bmatrix} \tag{47}$$

$$g = \begin{bmatrix} \gamma_0 & \cdots & \gamma_{p-d-1} \\ \vdots & \cdots & \vdots \\ \gamma_{p-d-1} & \cdots & \gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} E(\tilde{u}_t \tilde{w}_t^{(h)}) \\ \vdots \\ E(\tilde{u}_{t-p+d+1} \tilde{w}_t^{(h)}) \end{bmatrix}.$$

Note that g corresponds to the coefficients from a population linear projection of $\tilde{w}_t^{(h)}$ on $(\tilde{u}_t, \tilde{u}_{t-1}, ..., \tilde{u}_{t-p+d+1})'$.

Writing out (47) explicitly using (44) gives

$$\begin{bmatrix} I_{p-d+1} & 0 & 0 & \cdots & 0 \\ 0 & T^{1/2} & 0 & \cdots & 0 \\ 0 & 0 & T^{3/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & T^{d-1/2} \end{bmatrix} (\hat{\alpha}-q) \xrightarrow{p} \begin{bmatrix} g \\ 0 \end{bmatrix}.$$

This equation shows that the first $p-d$ elements of $\hat{\alpha}$ converge to the stationary population projection coefficients g, the $p-d+1$ term to $E(w_t^{(h)})$, and the last d elements of $\hat{\alpha}$ converge to the $\kappa_h^{(j)}$ terms in q. Indeed, the latter estimates are superconsistent– they still converge to the terms in q even when multiplied by some positive power of T.

Taking again the $p = 4$ and $d = 2$ example (42), the coefficients $\hat{\beta}$ from the actual regression of y_{t+h} on $(y_t, y_{t-1}, y_{t-2}, y_{t-3}, 1)'$ have plim

$$\hat{\beta} \xrightarrow{p} \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ -2 & 1 & 0 & -1 & 0 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -\mu & -\mu & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \mu h(h+1)/2 \\ h \\ 1 \end{bmatrix} = \begin{bmatrix} g_1 + h + 1 \\ g_2 - 2g_1 - h \\ g_1 - 2g_2 \\ g_2 \\ \mu\{[h(h+1)/2] - g_1 - g_2\} \end{bmatrix}.$$

The above derivation assumed $\mu = 0$ so that there was no drift in $\Delta^d y_t$. If instead we had $\mu \neq 0$, then $\Delta^{d-1} y_t = \sum_{s=1}^t u_s = \sum_{s=1}^t \tilde{u}_s + t\mu = \xi_t^{(1)} + t\mu$, which is dominated for large t by the drift term $t\mu$ rather than the random walk term $\xi_t^{(1)}$, and $\Delta^{d-j} y_t = \xi_t^{(j)} + (1/j)t^j\mu + o_p(t^j)$. In this case we would simply replace Υ_T in the above derivations with

$$\tilde{\Upsilon}_T = \begin{bmatrix} T^{1/2} I_{p-d+1} & 0 & 0 & \cdots & 0 \\ 0 & T^{3/2} & 0 & \cdots & 0 \\ 0 & 0 & T^{5/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & T^{d+1/2} \end{bmatrix}. \tag{48}$$

We would then arrive at the identical conclusion (47) this time using results (a), (c), and (g) from Sims, Stock and Watson Lemma 1.

Alternatively, adding a nonzero initial condition, e.g. replacing $\xi_t^{(1)}$ with $\xi_t^{(1)} + \xi_0^{(1)}$ for $\xi_0^{(1)}$ any fixed constant produces a term that is still dominated asymptotically by $\xi_t^{(1)}$, and as in Park and Phillips (1989), the original convergence claims again all go through.

Finally, the derivations are very similar for the case of purely deterministic time trends,

$y_t = \sum_{j=0}^{d} \delta_j t^j + u_t$. For this case we have $\mu = E(\Delta^d y_t) = \delta_d$ and

$$\tilde{x}_t = \begin{bmatrix} \Delta^d u_t - \delta_d \\ \vdots \\ \Delta^d u_{t-p+d+1} - \delta_d \\ 1 \\ \sum_{j=0}^{1} \delta_j^{(d-1)} t^j + \Delta^{d-1} u_t \\ \vdots \\ \sum_{j=0}^{d-1} \delta_j^{(1)} t^j + \Delta u_t \\ \sum_{j=0}^{d} \delta_j t^j + u_t \end{bmatrix}$$

where $\sum_{j=0}^{d-s} \delta_j^{(s)} t^j = \sum_{j=0}^{d-s+1} \delta_j^{(s-1)} t^j - \sum_{j=0}^{d-s+1} \delta_j^{(s-1)} (t-1)^j$ and $\delta_j^{(0)} = \delta_j$. Then for $\tilde{\Upsilon}_T$ as in (48), we again have

$$T^{-1/2} \tilde{\Upsilon}_T^{-1} \sum \tilde{x}_t \tilde{w}_t^{(h)} \xrightarrow{p} \begin{bmatrix} E(\Delta^d u_t - \delta_d) \tilde{w}_{t+h} \\ \vdots \\ E(\Delta^d u_{t-p+d+1} - \delta_d) \tilde{w}_{t+h} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The matrix $\tilde{\Upsilon}_T^{-1} \sum \tilde{x}_t \tilde{x}_t' \tilde{\Upsilon}_T^{-1}$ likewise has a block-diagonal plim, giving us again

$$\hat{\alpha} \xrightarrow{p} \begin{bmatrix} g \\ E(w_t^{(h)}) \\ \kappa_h^{(d)} \\ \vdots \\ \kappa_h^{(1)} \end{bmatrix} \tag{49}$$

for g the coefficients of the population linear projection of $\tilde{w}_{t+h}$ on $(\Delta^d y_t - \delta_d, ..., \Delta^d y_{t-p+d+1} - \delta_d)'$.

Derivation of equation (24).

For σ^2 the variance of ε_t, $\tilde{w}_t^{(h)} = \varepsilon_{t+h} - \varepsilon_t$, and $v_t = \varepsilon_t - \varepsilon_{t-1}$ we have

$$
g = \begin{bmatrix} E(v_t^2) & E(v_t v_{t-1}) & E(v_t v_{t-2}) & \cdots & E(v_t v_{t-p+2}) \\ E(v_{t-1} v_t) & E(v_{t-1}^2) & E(v_{t-1} v_{t-2}) & \cdots & E(v_{t-1} v_{t-p+2}) \\ E(v_{t-2} v_t) & E(v_{t-2} v_{t-1}) & E(v_{t-2}^2) & \cdots & E(v_{t-2} v_{t-p+2}) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ E(v_{t-p+2} v_t) & E(v_{t-p+2} v_{t-1}) & E(v_{t-p+2} v_{t-2}) & \cdots & E(v_{t-p+2}^2) \end{bmatrix}^{-1} \begin{bmatrix} E[v_t(\varepsilon_{t+h} - \varepsilon_t)] \\ E[v_{t-1}(\varepsilon_{t+h} - \varepsilon_t)] \\ E[v_{t-2}(\varepsilon_{t+h} - \varepsilon_t)] \\ \vdots \\ E[v_{t-p+2}(\varepsilon_{t+h} - \varepsilon_t) \end{bmatrix}
$$

$$
= \left\{ \sigma^2 \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} -\sigma^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} -(p-1)/p \\ -(p-2)/p \\ -(p-3)/p \\ \vdots \\ -1/p \end{bmatrix}
$$

where the last equation can be verified by premultiplying by

$$
\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}
$$

and confirming that the resulting vector is indeed $(-1, 0, ..., 0)'$. Hence the plim in (49) for this example is

$$\hat{\alpha} \xrightarrow{p} \begin{bmatrix} -(p-1)/p \\ -(p-2)/p \\ \vdots \\ -1/p \\ h\delta_1 \\ 1 \end{bmatrix}.$$

Also for this case we have

$$H = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & -\delta_1 \\ 0 & 1 & -1 & \cdots & 0 & 0 & -\delta_1 \\ 0 & 0 & 1 & \cdots & 0 & 0 & -\delta_1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 & -\delta_1 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}$$

so $\hat{\beta} = H'\hat{\alpha}$ has plim

$$\begin{bmatrix} 1/p \\ 1/p \\ \vdots \\ 1/p \\ \delta_1[h + (p-1)/p + (p-2)/p + \cdots + 1/p] \end{bmatrix}$$

implying a fitted value

$$
\begin{aligned}
\beta' x_t &= (1/p)(y_t + y_{t-1} + \cdots + y_{t-p+1}) + \delta_1[h + 1/p + 2/p + \cdots + (p-1)/p] \\
&= \delta_1 h + (1/p)\{y_t + [y_{t-1} + \delta_1] + [y_{t-2} + 2\delta_1] + \cdots + [y_{t-p+1} + (p-1)\delta_1]\} \\
&= \delta_1 h + (1/p)\{[\delta_0 + \delta_1 t + \varepsilon_t] + [\delta_0 + \delta_1 t + \varepsilon_{t-1}] + \cdots + [\delta_0 + \delta_1 t + \varepsilon_{t-p+1}]\} \\
&= \delta_0 + \delta_1(t+h) + (1/p)(\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-p+1}).
\end{aligned}
$$

Table 1. Maximum likelihood estimates of parameters of state-space formalization of the HP filter for assorted quarterly macroeconomic series.

	σ^2_c	σ^2_v	λ
GDP	0.115	0.468	0.245
Consumption	0.163	0.174	0.940
Investment	4.187	12.196	0.343
Exports	5.818	3.341	1.741
Imports	4.423	4.769	0.927
Government spending	0.221	1.160	0.191
Employment	0.006	0.250	0.023
Unemployment rate	0.014	0.092	0.152
GDP Deflator	0.018	0.081	0.216
S&P 500	21.284	15.186	1.402
10-year Treasury yield	0.135	0.054	2.486
Fed Funds Rate	0.633	0.116	5.458
Real Rate	0.875	0.091	9.596

Table 2. Standard deviation of cyclical component and correlation with cyclical component of GDP for assorted macroeconomic series.

	Regression Residuals		**Random walk**		**Sample**
	St. Dev.	*GDP Corr.*	*St. Dev.*	*GDP Corr.*	
GDP	3.38	1.00	3.69	1.00	1947:1-2016:1
Consumption	2.85	0.79	3.04	0.82	1947:1-2016:1
Investment	13.19	0.84	13.74	0.80	1947:1-2016:1
Exports	10.77	0.33	11.33	0.30	1947:1-2016:1
Imports	9.79	0.77	9.98	0.75	1947:1-2016:1
Government spending	7.13	0.31	8.60	0.38	1947:1-2016:1
Employment	3.09	0.85	3.32	0.85	1947:1-2016:2
Unemployment rate	1.44	-0.81	1.72	-0.79	1948:1-2016:2
GDP Deflator	2.99	0.04	4.11	-0.13	1947:1-2016:1
S&P 500	21.80	0.41	22.08	0.38	1950:1-2016:2
10-year Treasury yield	1.46	-0.05	1.51	0.08	1953:2-2016:2
Fed funds rate	2.78	0.33	3.03	0.40	1954:3-2016:2
Real rate	2.25	0.39	2.60	0.42	1958:1-2014:3

Notes to Table 2. Filtered series were based on the full sample available for that variable, while correlations were calculated using the subsample of overlapping values for the two indicators. Note that the regression residuals lose the first 11 observations and the baseline calculations lose the first 8 observations.

Figure 1. Values for ϕ_1 and ϕ_2 implied by different values of λ.

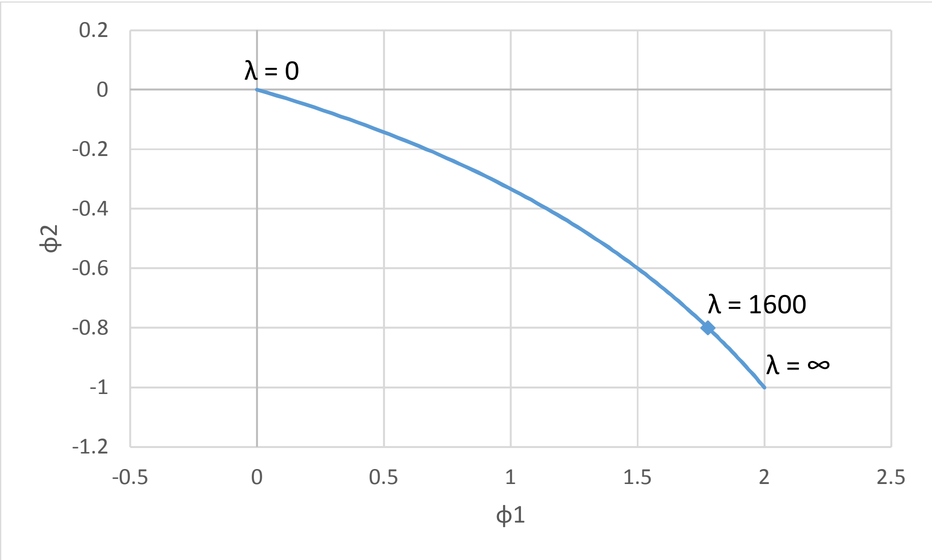
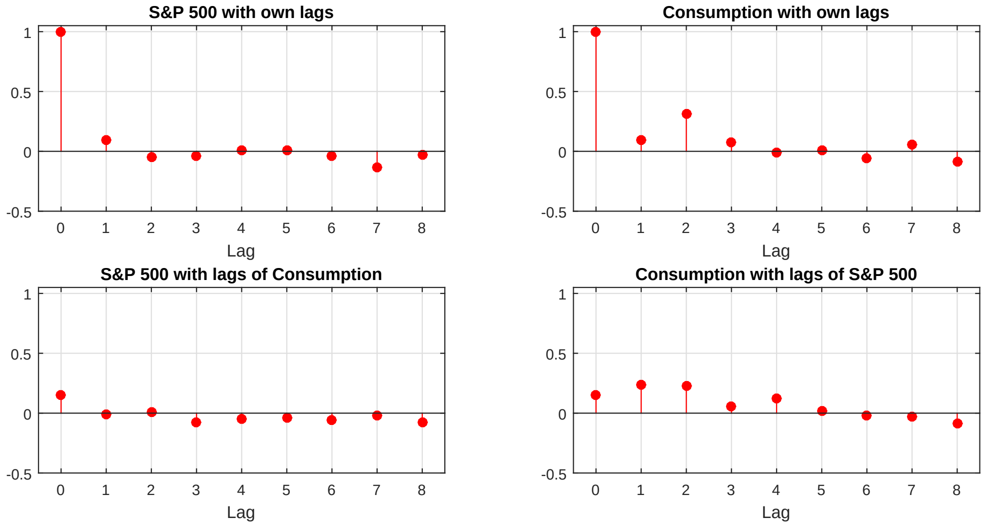
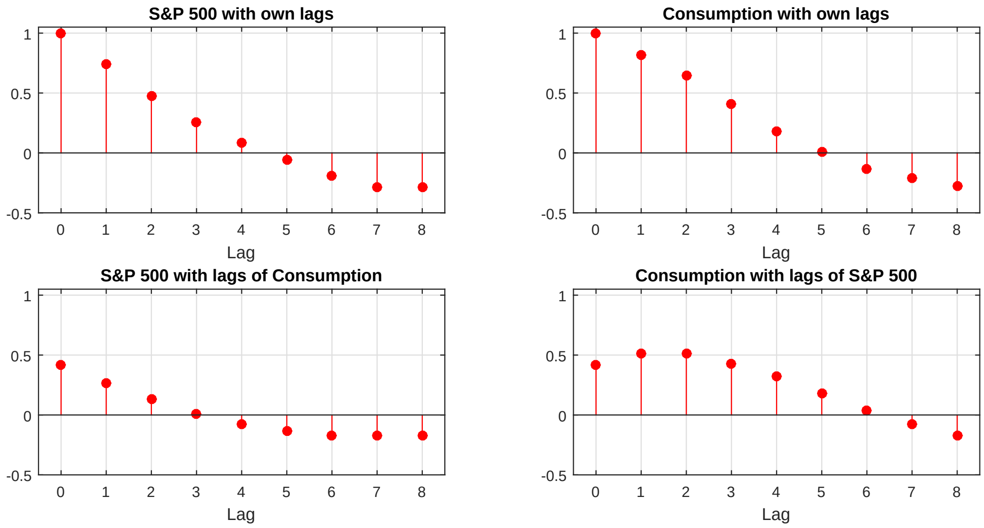


Figure 2. Autocorrelations and cross-correlations for first-difference of stock prices and real consumption spending.



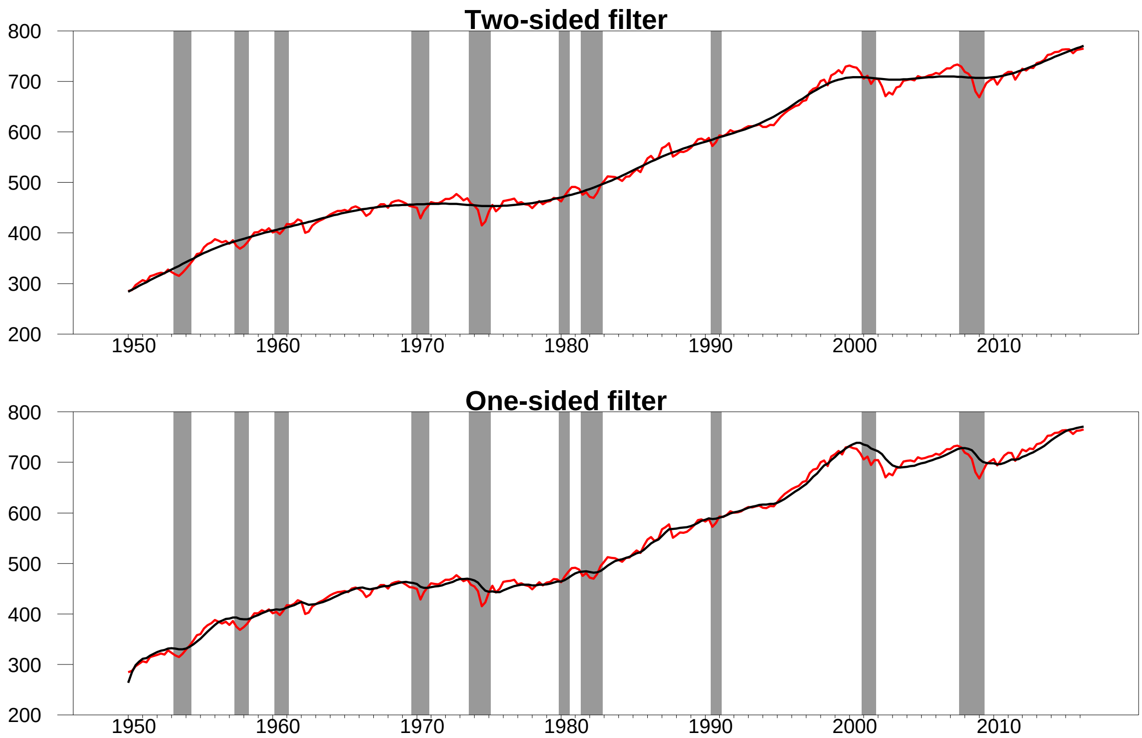
Notes to Figure 2. Upper left: autocorrelations of log growth rate of end-of-quarter value for S&P 500. Upper right: autocorrelations of log growth rate of real consumption spending. Lower panels: cross correlations.

Figure 3. Autocorrelations and cross-correlations for HP cyclical component of stock prices and real consumption spending.



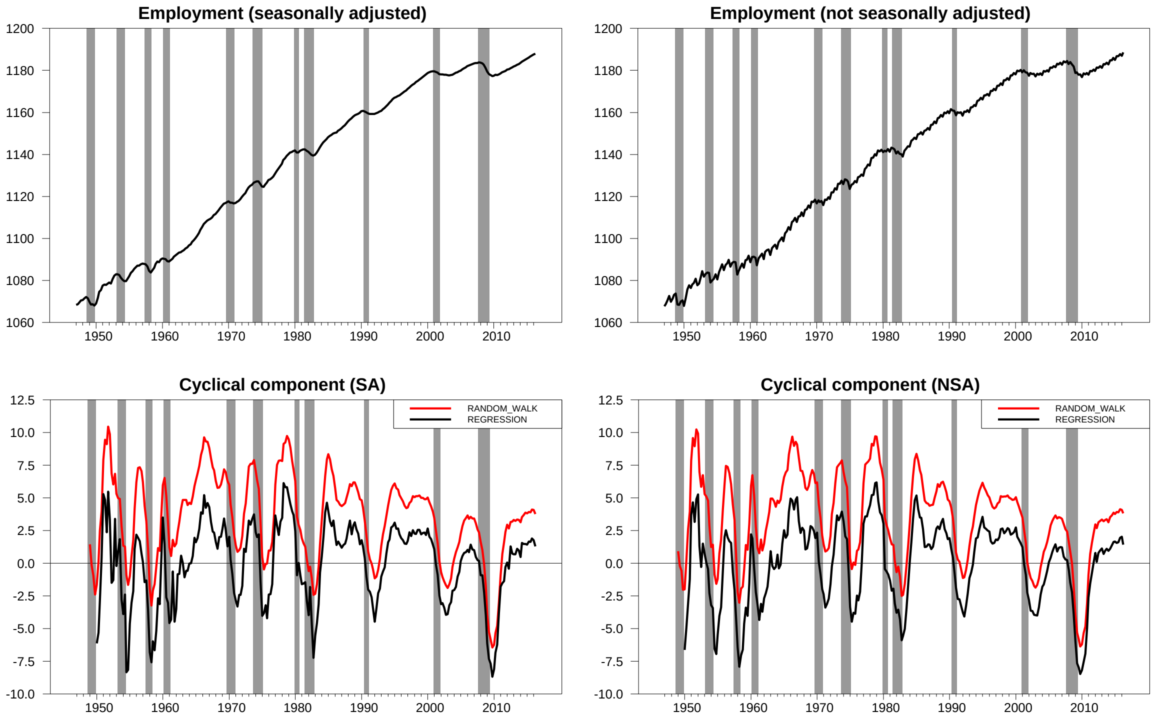
Notes to Figure 3. Upper left: autocorrelations of HP cycle for log of end-of-quarter value for S&P 500. Upper right: autocorrelations of HP cycle for log of real consumption spending. Lower panels: cross correlations.

Figure 4. Comparison of one-sided and two-sided HP filters.



Notes to Figure 4. Red line in both panels plots 100 times natural log of S&P 500 stock price index. The black curve in the top panel plots the HP estimate of trend as inferred using the usual two-sided filter (calculated using the Kalman smoother for the state-space model in Proposition 1), whereas the black curve in the bottom panel plots trend from a one-sided HP filter (calculated using the Kalman filter for the same model). Shaded regions denote NBER recession dates.

Figure 5. Regression and 8-quarter-change filters applied to seasonally adjusted and seasonally unadjusted employment data.



Notes to Figure 5. Upper left: 100 times the log of end-of-quarter values for seasonally adjusted nonfarm payrolls. Lower left: black plots $y_t - \hat{\beta}_0 - \hat{\beta}_1 y_{t-8} - \hat{\beta}_2 y_{t-9} - \hat{\beta}_3 y_{t-10} - \hat{\beta}_4 y_{t-11}$ as a function of t while red plots $y_t - y_{t-8}$. Right panels show results when the identical procedure is applied instead to seasonally unadjusted data.

Figure 6. Results of applying regression (black) and 8-quarter-change (red) filters to 100 times the log of components of U.S. national income and product accounts.