Persistence Dependence in Empirical Relations: The Velocity of Money

Richard Ashley
Department of Economics
Virginia Tech
Blacksburg, VA 24061-0316
Email: ashleyr@vt.edu

Randal J. Verbrugge
Economic Research
Federal Reserve Bank of Cleveland
1455 E. 6th St.
Cleveland, OH 44114
Email: randal.verbrugge@clev.frb.org

December 15, 2015

JEL Classification Codes: E00, E31, C22, C32, E5

Keywords: Velocity, money demand, frequency-dependence, spectral regression

Acknowledgement: This paper was inspired by Cochrane (2012). All errors, misinterpretations and omissions are ours. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

Brief Abstract

Standard theory predicts persistence dependence in numerous economic relationships. But achieving credible inference about this was not possible until recently. New tools allow one to elegantly quantify the variation in a time-series relationship across persistence levels. We apply these tools to study the velocity of money. Standard theory predicts that velocity should be positively correlated with the nominal interest rate. But Cochrane (2012) demonstrates that the linkage appears to be weak upon first-differencing. We here demonstrate that this follows from the relationship's persistence-dependence: this relationship is substantially (and statistically significantly) stronger at *low* frequencies – clearly evident despite first-differencing.

Extended Abstract

Standard theory predicts persistence dependence in numerous economic relationships. (For example, persistence dependence is precisely the kind of nonlinear relationship posited in the Permanent Income Hypothesis; persistence dependence is the inverse of 'frequency dependence' in a relationship.) Until recently, however, it was challenging to achieve credible inference about persistence dependence in an economic relationship using available methods. However, recently developed econometric tools (Ashley and Verbrugge, 2009a) allow one to elegantly quantify the variation in a time-series relationship across persistence levels, even when the data must be first-differenced because they are I(1), or nearly so. We apply these tools to study the velocity of money. Standard theory predicts that velocity should be positively correlated with the nominal interest rate: a high nominal interest rate raises the opportunity cost of holding wealth in liquid form, prompting agents to economize on money holdings. But as Cochrane (2012) pointed out, the velocity-interest rate linkage appears to be weak upon first-differencing. We argue that the root cause of this phenomenon is a particularly intuitive form of nonlinear dependence in the relationship: the strength of the relationship depends on the persistence level of a particular interest rate fluctuation. In particular, this relationship is substantially (and statistically significantly) stronger at *low* frequencies – i.e., at high interest rate fluctuation persistence levels. Because we allow for persistence dependence in the estimated relationship, this strong association is apparent despite the first-difference transformation applied to these data.

1 Introduction

> "It is often useful to regard time series as a combination of transitory stochastic and more permanent underlying components and to regard the two components as reflecting two different sets of forces; e.g., purely random measurement errors may have a far larger impact on the transitory component than on the permanent component. The process of obtaining serially uncorrelated residuals may in effect simply eliminate the permanent components, leaving the analyst to study the relation among the stochastic components of his series, which may be pure noise, when what is of economic interest is the relation between the permanent components he has discarded in the process of seeking to satisfy mechanical statistical tests." Friedman (1988, p. 230)

Empirical studies in economics generally assume linear relationships between variables. There are many reasons for this: Occam's razor; the fact that a first order approximation is often satisfactory; the fact that linear tools are more readily available; and so on. Perhaps of chief importance is this reason: while there are many potential forms of nonlinearity, relatively few are suggested by a compelling economic theory. Because of the multiplicity of ways in which a time-series relationship can be nonlinear, a search over a variety of them is bound to detect one form or another, simply by chance. Thus, if a particular form of nonlinearity is detected based primarily on statistical grounds, it is reasonable to suspect that it may be artifactual.

In contrast, one particular form of nonlinearity is often well-supported by theory in various contexts, and usually straightforward to interpret: persistence dependence. This refers to the notion that the relationship between the more transitory components of two time series is distinct from the relationship between their more persistent components. As Friedman (1988) and Cochrane (2012) suggest, measurement error may routinely give rise to persistence dependence, as a given time series might be plagued by highly transitory noise that is unrelated to other variables. Or persistence dependence might arise in an economically meaningful fashion, as in the Permanent Income Hypothesis of macroeconomic consumption theory.

Such persistence dependence gives rise to inference problems that we discuss below. It has not been addressed much because, until very recently, extant tools were difficult to use and limited in their ability to deliver clean and credible inference.

In this paper, we demonstrate that linear econometric analysis can be very misleading in the presence of persistence dependence. Key relationships can be missed, and it is even possible that the *sign* of a coefficient can be deceptive. This is because inference from linear tools is only valid if the relationship is not persistent-dependent. We demonstrate that the problems with standard techniques (such as differencing) extend beyond those noted by Friedman (1988) and Cochrane (2012). Further, we discuss newly-developed techniques that allow one to easily and gracefully detect, and model, persistence-dependence in the data. We apply these tools to the compelling example used in Cochrane (2012): the velocity of money. In that paper, Cochrane demonstrated that first-differencing appears to give rise to misleading results. He suggested using a levels-specification, with a correction for standard errors, as a superior approach. We provide a much more thorough explication of the the relationship between velocity and interest rates, a relationship that the tools used by Cochrane can only hint at.

2 Persistence dependence and linear analysis

2.1 Persistence dependence in economics

Either economic theory or an informal inspection of the data frequently suggests that y_t and x_t have one kind of relationship at high persistence levels (i.e., at low frequencies), but a different kind of relationship at low persistence levels (i.e., at high frequencies). One prominent theoretical example is the Permanent Income Hypothesis; it posits that consumption is strongly related to low-frequency variations in income, whereas high-frequency variations in income do not affect consumption. But other theoretical examples abound in economics. In the context of monetary policy, theory dictates that a central bank should not respond to movements in the natural rate of unemployment – and furthermore, that high-frequency movements in inflation should also be ignored (see Ashley, Tsang and Verbrugge 2015 for evidence); the standard Phillips curve posits that inflation itself responds to slack, and hence, does not respond to variations in the NAIRU (see Ashley and Verbrugge 2013); forward-looking optimizing firms must attempt to distinguish between high-frequency and low-frequency movements in demand and cost conditions; and so on. According to standard macroeconomic theory, the relationship between variables at "business-cycle" frequencies (or the "cyclical components", in the terminology of Kydland and Prescott 1983) might differ from the relationship at lower frequencies. Indeed, this notion motivated the development of the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997) in 1980, and its subsequent routine use in business-cycle analysis for decades. (We discuss below why the common practice of examining the relationship of HP-filtered data does not allow one to credibly draw conclusions about business-cycle relationships.)

2.2 Persistence dependence and the inadequacy of ordinary linear regression

We here clarify further the meaning of the term "frequency dependence" in the context of a regression coefficient, to distinguish it from related concepts, and to demonstrate how linear specifications cannot capture this kind of serial dependence. Ashley and Verbrugge (2009a) observes that a linear relationship between c_t and y_{t-1} implies that the relationship between c_t and y_{t-1} is the same at all frequencies; that is, if y_{t-1} experiences a one-standard-deviation increase, then c_t responds in the same way regardless of whether that one-standard-deviation increase is part of an unusually persistent movement in y_{t-1} or whether it is part of an unusually transient movement in y_{t-1}.

Now consider the following hypothetical consumption function:

$$c_{it} = \gamma_{o,i} + \gamma_1 y_{it-1} + \gamma_2 y_{it-2} + \gamma_3 c_{it-1} + \beta X_{it-1} + \varepsilon_{it} \tag{1}$$

where c_{it} and y_{it} are the log of consumption spending and disposable income of individual i in period t, X_{it} represents control variables (such as the number of children, etc.), and ε_{it} is a covariance-stationary error term. In this model γ_1 may be interpreted as the "short-run marginal propensity to consume," characterizing how consumption spending (on average) responds to fluctuations in y_{it-1}. In contrast, $\frac{(\gamma_1+\gamma_2)}{(1-\gamma_3)}$ might be interpreted as the "long-run marginal propensity to consume," in that it represents the eventual total response of consumption to a one unit change in income. The distinction between γ_1 and $\frac{(\gamma_1+\gamma_2)}{(1-\gamma_3)}$ is <u>*not*</u> frequency-dependence.

Rather, frequency-dependence, of the form asserted by the permanent-income hypothesis, implies that *this regression is misspecified:* the value of γ_1 itself depends upon frequency. In particular, this hypothesis asserts that consumption should *not* change appreciably if the previous period's fluctuation in income is highly transitory (i.e., high-frequency), whereas consumption *should* change significantly if the previous period's fluctuation in income is part of a persistent (low-frequency) movement in income. Thus, the inverse of frequency dependence is what we here call "persistence dependence." γ_1, then, should be approximately equal to zero for high frequencies, and close to one for very low frequencies. Notice that equation (1) contains an implicit assumption which violates the permanent income theory: a restriction that γ_1 is the same across all frequencies.[1]

When the coefficient on y_{it-1}, say, is frequency-dependent, it is thus history-dependent; that is, it depends upon y_{it-2}, y_{it-3}, and so forth. Therefore the linear specification in equation (1) must be incorrect. In that case, inference based upon the linear specification, i.e. equation (1), is clearly also not correct.

A useful analogy is the standard example of a break in a coefficient. To focus ideas, consider the simpler consumption model,

$$c_{it} = \gamma_{o,i} + \gamma_1 y_{it-1} + \gamma_3 c_{it-1} + \varepsilon_{it} \tag{2}$$

The parameter γ_1 can be interpreted as the conditional expectation $\partial E c_{it}/\partial y_{it-1}$. But suppose that the coefficient γ_1 actually takes on two values: γ_1^* in the first half of the sample and γ_1^{**} in the second half of the sample, for example. Then this regression is clearly mis-specified, and the usual statistical machinery for testing hypotheses about γ_1 is invalid. Indeed, the hypotheses themselves are essentially meaningless, since γ_1 does not have a single well-defined value to test. Similarly, the least-squares estimate of γ_1 is clearly neither a consistent estimator for γ_1^* nor for γ_1^{**}. In particular, if the sign of the relationship is positive in the first part of the sample and negative later on, then the least squares estimate of γ_1 might well be close to zero, even if both γ_1^* and $|\gamma_1^{**}|$ are quite

[1] One might wonder why this issue does not seem to arise in many theoretical models of consumption. The explanation is that the filtering problem in these models is trivial: no filtering is necessary, since all fluctuations in income are treated equivalently. This will be the case when income is specified as a known, essentially linear stochastic process (such as AR(1) in levels or first differences). Of course, such assumptions need not hold in reality. For example, some fluctuations in after-tax income are known to be transitory. See Arellano, Blundell and Bonhomme (2014), who discuss the shortcomings of linear consumption models and develop a particular nonlinear model which incorporates persistence-dependence in consumption.

large.

If the value of γ_1 different at low frequencies than at high frequencies, then all of the same unhappy properties hold. In particular, the least squares estimator of γ_1 is an inconsistent estimator of $\partial Ec_{it}/\partial y_{it-1}$, and – since γ_1 does not have a unique value – hypothesis tests about γ_1 are of doubtful value, and in many instances quite misleading.

2.3 Other tools

Frequency or persistence based decompositions are not, in themselves, new to economics. Beveridge and Nelson (1981) provide an early example, in which the time series is decomposed into a ('permanent') integrated – i.e., I(1) – series and a remainder ('transitory') covariance stationary series. This decomposition allows for only two levels of persistence: arbitrarily large and finite. Wavelet analysis (e.g., Ramsey and Lampart, 1998) provides another example. Wavelet analysis projects the data on one of several families of time-dependent basis sets, in each of which the persistence of a fluctuation varies. The choice of which wavelet family to use is not clear, however. Also, the focus in a wavelet decomposition is on the detection of time-dependent variations in persistence, with little opportunity for economic interpretation as to how this dependence and variation arise.

A species of what we call 'pseudo frequency dependence' in a regression coefficient was proposed by Geweke (1982), fundamentally based on the cross-spectrum between the an explanatory variable and the dependent variable. A discussion of this technique is relegated to Appendix 1 below, as this is the analysis of a mathematical decomposition (albeit in the frequency domain) of a fixed-coefficient linear time-series relationship, rather than the nonlinear frequency dependence considered here. The estimated gain and phase functions resulting from this kind of analysis are, in addition to being inconsistently estimated where there is feedback, very difficult to interpret in economic terms.[2]

Consequently, we focus below on a method which is fundamentally based on a one-sided band-pass filtering of the data. However, as explained below, band-pass filtering as currently practiced cannot be used.

2.4 Differencing and persistence dependence

Under persistence dependence, differencing can lead to distorted inference. If the relationship between two variables is linear, and therefore the same at all frequencies, the estimated dynamic relationship between these two variables will not be distorted by the first-differencing. But as Baxter (1994) pointed out, the first-difference filter emphasizes the high-frequency variation in the data (see Figure 1) at the expense of *all* other variation. Not only is the zero-frequency component of the variation removed by first-differencing, but much of the low- and medium-frequency variation is greatly attenuated, leaving mostly the high-frequency variation in the variable.

Thus, regardless of whether the variables y_t and x_{t-1} are mostly related via their most persistent components (as in the examples of Friedman 1988 and Cochrane 2012), or alternatively are mostly related via "business-cycle" components (as in the example of Baxter 1994), a linear regression of a first-differenced variable y_t on a first-differenced variable x_{t-1} will yield a coefficient estimate that pertains mainly to the high-frequency relationship between the two variables – which may well be quite weak.

[2] Granger (1969) notes, "in many realistic economic situations, however, one suspects that feedback is occurring. In these situations the coherence and phase diagrams become difficult or impossible to interpret, particularly the phase diagram."

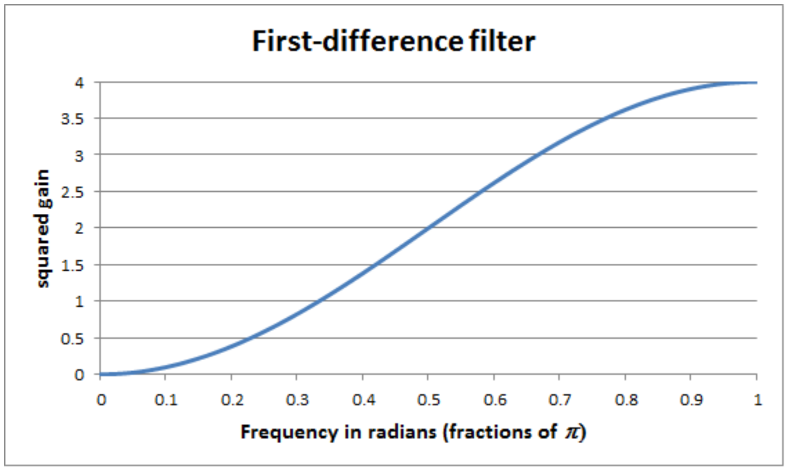

Figure 1: Squared gain of the first-difference filter.

As noted above, persistence dependence is common in economics. The solution is not "don't first difference." First-differencing is necessary if the data are non-stationary – i.e., I(1) – and failing to difference in that circumstance leads to the well known "spurious regression" phenomenon with regard to any levels variables that are not in a cointegrating relation (if there is one).[3] Under persistence dependence, any analysis that assumes linearity in the relationship is fraught with danger for both consistent parameter estimation and for statistical inference. Linearity assumes that the relationship between two variables is the same at all frequencies – that is, for all persistence levels. If this assumption is violated, then linear tools are not appropriate.

2.5 Two-sided filtering and the problem with feedback

It is natural to consider detecting and modeling frequency dependence in a regression relationship by repeatedly band-pass filtering both sides of the equation. To this end, excellent band-pass filters have been developed by Baxter and King (1999), Christiano and Fitzgerald (2003) and Iacobucci

[3]See Ashley and Verbrugge (2009b) and Granger, Hyung and Jeon (2010). They show that distorted inference will occur even for I(0) data, if it is highly persistent and the sample length is not huge.

and Noullez (2005). Indeed, the use of these and other two-sided filters is common, particularly in the RBC literature, ostensibly because researchers wish to focus on the relationship between variables at business-cycle frequencies. However, there is a pitfall in utilizing such filtering whenever the relationship involves feedback.

Consider the analysis of possible frequency dependence in the parameter λ_2 of the following bivariate equation system:

$$y_t = \lambda_1 y_{t-1} + \lambda_2 x_{t-1} + \varepsilon_t \tag{3}$$

$$x_t = \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \eta_t$$

Clearly, this is a feedback relationship only if α_2 is nonzero. But note that the x_t equation implies that

$$\begin{aligned} x_t &= \alpha_1 x_{t-1} + \alpha_2 y_{t-1} + \eta_t \\ &= \alpha_1 x_{t-1} + \alpha_2 \left(\lambda_1 y_{t-2} + \lambda_2 x_{t-2} + \varepsilon_{t-1}\right) + \eta_t \\ &= \alpha_1 x_{t-1} + \alpha_2 \lambda_1 y_{t-2} + \alpha_2 \lambda_2 x_{t-2} + \alpha_2 \varepsilon_{t-1} + \eta_t \end{aligned}$$

so that x_t is correlated with ε_{t-1} if there is feedback from past y_t to x_t. But, two-sided filtering implies that x^*_{t-1} depends upon x_t, x_{t+1}, x_{t+2}, etc., so that x^*_{t-1} is thus correlated with $\varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \varepsilon_{t+1}, ...$, which (under two-sided filtering) are correlated with ε^*_t. Thus, in the presence of feedback, a two-sided transformation of x_{t-1} will in general produce a transformed explanatory variable, x^*_{t-1}, which is correlated with the tranformed error term, ε^*_t, yielding inconsistent least-squares parameter estimates (including correlation estimates). Putting this differently, using a two-sided filter means that one cannot reliably uncover the true relationship between y_t and x_t. A one-sided filter is necessary.

A second potential pitfall relates to the filtering problem faced by agents. Frequency dependence often arises from the intertemporal optimizing behavior of agents facing uncertainty. In real-time, agents might misinterpret a highly transitory fluctuation as a persistent fluctuation, and might respond accordingly. A two-sided filter implicitly grants perfect foresight to agents with respect to this type of filtering problem.

Below we use filtration based on rolling windows moving through the sample data; this effectively renders our filtering one-sided, and hence still valid in the presence of feedback. Our approach partitions the sample data on an explanatory variable (and only this variable) into set of frequency/persistence components which add up to the original sample data for this variable. Consequently, in our approach, one can simply replace the data on this variable in the original regression model by a linear form in these persistence components and estimate/test the resulting new coefficient estimates.[4] This makes the approach easy to implement and straightforward to interpret in economic terms.

3 Methodology: The Ashley/Verbrugge approach

In essence, the Ashley/Verbrugge (2009a) approach consists of applying a one-sided band-pass filter solely to the explanatory variable – depicted Δx_t below – whose coefficient is being examined for possible frequency dependence. As described below, this filter simultaneously decomposes Δx_t into k separate time series, where the first series $\left(z_t^{\{1\}}\right)$ corresponds to the lowest (zero) frequency portion of Δx_t, the next series $\left(z_t^{\{2\}}\right)$ corresponds to the next-lowest frequency portion of Δx_t, and so on, with $\sum_{i=1}^{k} z_t^{\{i\}} = \Delta x_t$.[5] Then a regression equation such as $\Delta y_t = \alpha + \beta \Delta x_{t-1} + e_t$ can

[4] Further, one can simultaneously partition any number of explanatory variables in any otherwise conventional specification; for example, Ashley et al. (2015) investigate nonlinearity in a Taylor rule, which involves partitioning both the unemployment rate and the inflation rate.

[5] Software to accomplish this decomposition is available from the authors, either as a standalone Windows executable or as RATS code, which can be easily customized. We discuss the upper bound on k momentarily.

be reformulated as:

$$\Delta y_t = \alpha + \sum_{i=1}^{k} \beta_i z_{t-1}^{\{i\}} + e_t$$

The coefficient β_1 estimates the lowest-frequency relationship between y and x. Similarly, the coefficient β_k estimates the highest-frequency relationship. A rejection of the hypothesis that these $\beta_1 ... \beta_k$ coefficients are all equal, using the usual F-test, implies frequency dependence in the relationship.

We provide specifics below on the relatively simple and intuitive set of bandpass filters that we use here to obtain the k frequency component time series $\left(z_t^{\{1\}}, ..., z_t^{\{k\}}\right)$ from the data on Δx_t itself. Like most bandpass filters, this filter is fundamentally two-sided in nature, but (for the reasons given above) one-sided filtering is needed here. We obtain a one-sided filtering by repeatedly applying this filter to a set of $m < N$ sample observations on Δx_t at a time. The length of this subset (m) is kept constant, but its starting point is repeatedly incremented, moving a 'rolling window' of length m through the sample data set.

Suppose (as is the case in the empirical work below) that m is set equal to 16 quarters, so that each of these moving windows is 16 quarters in length. And consider the first such window, consisting of the sample observations of $\Delta x_1 ... \Delta x_{16}$. Applying k bandpass filters to this window extracts k time series $\left(\chi_t^1, \chi_t^2, ..., \chi_t^k\right)$ – one time series for each of the k frequency components – with each of these k time series running from period 1 to period 16. Each of the $\chi_t^1, \chi_t^2, ..., \chi_t^k$ components obtained from this 16-month long window depends upon all 16 observations of Δx_t, reflecting the two-sided nature of bandpass filtering. But the last observation, dated '16,' depends only upon data in its own past. This final observation alone is retained; in particular, we set $\left(z_{16}^{\{1\}}, ..., z_{16}^{\{k\}}\right)$ equal to $\left(\chi_{16}^1, ..., \chi_{16}^k\right)$, the 16$^{\text{th}}$ (last) observation on each of the k frequency component time series extracted by the bandpass filters from the first window of the sample data, $\Delta x_1 ... \Delta x_{16}$.

In the next step of the procedure, the window is moved one month farther in the sample, so that it now contains observations $\Delta x_2 ... \Delta x_{17}$. Once again, k time series are obtained, this time running from observation number 2 to observation number 17. As above, we retain only the last observation in each of the k frequency components, i.e., $\left(z_{17}^{\{1\}}, ..., z_{17}^{\{k\}}\right) = \left(\chi_{17}^1, ..., \chi_{17}^k\right)$. Next the window is again moved ahead one month, to yield the vector $\left(z_{18}^{\{1\}}, ..., z_{18}^{\{k\}}\right)$. This process is continued until the last observation in the window is Δx_T, at which point the k frequency component time series observations $\left(z_s^{\{1\}}, z_s^{\{2\}}, ..., z\left\{{k \atop s}\right\}\right)$, which run from date $s = 16$ until date $s = T$, have been obtained.[6] Notice that at each value of s, the vector $\left(z_s^{\{1\}}, z_s^{\{2\}}, ..., z_s^{\{k\}}\right)$ is constructed using only data from its own past. Thus, this moving window procedure uses a two-sided filter in each window, but produces frequency components that are, by construction, backward-looking – i.e., the product of one-sided filtering.

A window size of m allows a partitioning of the data into, at most, $\left(\frac{m}{2} + 1\right)$ components; i.e., $k \leq \left(\frac{m}{2} + 1\right)$. Thus a window 16 quarters in length allows for 9 frequency components. The first of these corresponds to the sample mean of $\Delta x_1 ... \Delta x_{16}$, i.e., to a frequency of zero. The next component is comprised of oscillations in the window that occur at a frequency of $2\pi/16$. (As explained below, the filter we use builds this component from an associated cosine and sine function at this frequency.) The next frequency component corresponds to frequency $2\,(2\pi/16)$; the next to frequency $3\,(2\pi/16)$; and so forth. The highest-frequency component, $8\,(2\pi/16) = \pi$, is just a sequence of the changes in the series Δx_t.[7]

It is more intuitive to index each of these k components by its reversal period – typically

[6] Reference to the A matrix defined in equation (5) below shows that the first frequency component extracted (i.e., $z_{16}^{\{1\}}, ..., z_T^{\{1\}}$), which corresponds to a frequency of zero and the first row of A, is here just a 16-quarter backward-looking moving average of Δx_t.

[7] See equation (5) below; and see Ashley, Tsang and Verbrugge (2015) for a more detailed account and an example with a window ten periods in length. Table 4 in the Appendix lists the frequencies and reversal periods corresponding to a window 16 quarters in length.

denoted just 'period' – which is defined as 2π times the reciprocal of the frequency; the reversal period corresponds to the number of months needed for a sinusoid with this frequency to complete one full oscillation. Thus, the lowest non-zero-frequency $\left(z_t^{\{2\}}\right)$ corresponds to a period of 16 quarters, and the largest frequency corresponds to a period of 2 quarters. The general idea of the bandpass filter being described here is that fluctuations in $\Delta x_1 ... \Delta x_{16}$ that vary quickly – i.e., tend to self-reverse within a couple of quarters – are mainly placed in the highest-frequency component, $z_t^{\{9\}}$. In contrast, fluctuations in $\Delta x_1 ... \Delta x_{16}$ that vary slowly – i.e., which tend to persist over most of the 16 quarter window – mainly end up in the lowest non-zero-frequency component $z_t^{\{1\}}$. With 16-quarter moving windows, fluctuations in $\Delta x_1 ... \Delta x_{16}$ that occur at the seasonal frequency (i.e., that more or less recur every four quarters) would mainly end up in the frequency component $z_t^{\{5\}}$.[8]

In principle, one could apply any two-sided bandpass filter within a moving window as described above. But most available filters only partition the data into two parts, greatly limiting the ability of the data to speak to the true nature of the frequency dependence in the data, or requiring iterative use of the filter to obtain frequency components which add up to the original Δx_t series. The filter we use is less sophisticated than, e.g., that of Christiano and Fitzgerald (2003). However, it is intuitively appealing and has the distinct advantage of automatically yielding a set of k frequency

[8] As the zero-frequency component, $z_t^{\{1\}}$, is just the sample mean of Δx_t over the observations in the window, any component of Δx_t which is either actually a constant or varies so slowly so as to not change appreciably in 16 quarters will have little impact on $z_t^{\{1\}}$. The frequency decomposition must be a bit more complex than described above, due to what are usually called 'edge effects.' When decomposing the Δx_t data, it is well known – e.g., see Dagum (1978) or Stock and Watson (1999) – that better results are obtained by augmenting the window data with projected observations for future periods. We find that at least four quarters of projected data (or, with monthly data, at least 5 months) are helpful. (Four quarters are used in the analysis described below.) Thus, for example, with a window length m of 16 quarters, the window for obtaining frequency components in period $t+12$ (i.e., for obtaining the vector $\left(z_{t+12}^{\{1\}}, ..., z_{t+12}^{\{9\}}\right)$ uses 12 quarters of actual data $(\Delta x_{t+1}, ..., \Delta x_{t+12})$ in conjunction with projections of $\Delta x_{t+13} ... \Delta x_{t+16}$, to yield a window of length 16 quarters. The frequency components extracted using the window are $z_{t+12}^{\{1\}}, ..., z_{t+12}^{\{9\}}$, the 12th element from the window; the components for observation number $t+12$ are thus still based only on data up through observation number $t+12$. We also use a linear detrending procedure within each window, to address the fact that the sample data within a relatively short window will appear trended for some windows.

components that, at each time period s, add up to the original series; that is, by construction, $z_s^{\{1\}} + z_s^{\{2\}} + ... + z_s^{\{k\}} = \Delta x_s$. This filter was introduced in Tan and Ashley (1999a and 1999b) and is also discussed in detail in Ashley and Verbrugge (2009a), so we only briefly review it here.

Tan and Ashley (1999a and 1999b) developed a real-valued re-formulation of the Engle's (1974) framework. It is based on the ordinary regression model:

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N\left(0, \sigma^2 I\right) \tag{4}$$

where Y is $m \times 1$ and, for illustration, X is $m \times 1$ (i.e., X consists of a single time series of length m). Now define a $m \times m$ real-valued matrix A with $(j,t)^{th}$ element:

$$a_{j,t} = \begin{cases} \frac{1}{\sqrt{m}} & j = 1 \\ \sqrt{\frac{2}{m}} \cos\left[\frac{\pi j(t-1)}{m}\right] & j = 2, 4, ..., m-2 \\ \sqrt{\frac{2}{m}} \sin\left[\frac{\pi (j-1)(t-1)}{m}\right] & j = 3, 5, ..., m-1 \\ \frac{1}{\sqrt{m}} \left(-1\right)^{t+1} & j = m \end{cases} \tag{5}$$

It can be shown that A is an orthogonal matrix. Premultiplying the regression model (4) by A yields

$$AY = AX\beta + A\varepsilon \;\rightarrow\; Y^* = X^*\beta + \varepsilon^* \qquad \varepsilon^* \sim N\left(0, \sigma^2 I\right)$$

The dimensions of the of Y^*, X^* and ε^* arrays are the same as those of Y, X and ε in (4), but the m components of Y^* and ε^* and the rows of X^* now correspond to frequencies instead of time periods, with the initial row corresponding to the lowest-frequency part of X. (Transformation back into the time domain will occur after the next step.)

Suppose one wishes to decompose frequency dependence between Y and X. In the frequency domain, this corresponds to testing whether the coefficient β is the same across the m "observations" of X_j^* in the frequency domain. But one generally wants to conduct regressions in the time domain.

To do so, the Ashley/Verbrugge approach strips apart X into k pieces. The first piece consists of the first "observation" of X. After that, we take *pairs* of observations; we use pairs because at a given frequency, there is usually both a sine and a cosine portion. We continue to take pairs individually, or possibly to group pairs of observations, until it comes to the very last observation. Hence if $k = 4$ and $m = 6$, we would obtain:[9]

$$\underset{X^*}{\begin{bmatrix} X_1^* \\ X_2^* \\ X_3^* \\ X_4^* \\ X_5^* \\ X_6^* \end{bmatrix}} = \underset{X^{1*}}{\begin{bmatrix} X_1^* \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}} + \underset{X^{2*}}{\begin{bmatrix} 0 \\ X_2^* \\ X_3^* \\ 0 \\ 0 \\ 0 \end{bmatrix}} + \underset{X_j^{3*}}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ X_4^* \\ X_5^* \\ 0 \end{bmatrix}} + \underset{X_j^{4*}}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ X_6^* \end{bmatrix}}$$

Back-transformation into the time domain is accomplished by pre-multiplying by A^T, the transpose of A. Thus, $A^T Y^* = Y$, etc. Similarly, we compute $z_t^{\{i\}} = A^T X^{i*}$ for each i. In each case, pre-multiplying the column with A^T yields a time series of length m that corresponds to the frequencies associated with the observations in X^*. Clearly, $X_t^* = \sum_{i=1}^{k} z_t^{\{i\}}$; the $z_t^{\{i\}}$ are effectively band-pass filtered versions of X, with two nice properties: they are orthogonal, and they add up precisely to X. To test for frequency dependence in the regression coefficient on this regressor, then, all that one need do is test the null hypothesis that $\beta^1 = \beta^2 = ... = \beta^k$.[10]

Selecting the number of frequency bands (k) and the particular set of frequencies to be included in each band is an important issue. A typical approach in the literature has been to completely

[9] Note that X_1^* corresponds to a frequency of zero; X_2^* and X_3^* correspond to a frequency of $2\pi/6$ and a period of 6; X_4^* and X_5^* correspond to a frequency of $2\,(2\pi/6)$ and a period of 3; and X_6^* corresponds to a frequency of $3\,(2\pi/6) = \pi$ and a period of 2. See equation (6) with $m = 16$, and Table 1 below.

[10] Here we use a moving window of length m and apply the filter described above within each moving window, so as to force the filter to be one-sided, as described earlier in this section. This use of the moving window filtration retains the "adding up" property for the frequency components noted here, but the k component time series are no longer precisely orthogonal.

limit consideration to an *ad hoc* set of frequencies thought to correspond to the 'business cycle.'[11] Here this would correspond to combining several (pairs of) components $z_t^{\{i\}}$ into one aggregated 'business cycle' component, say Z_t^{BC}. If the values of β_j^i are the same for all of the values of i included in this 'business cycle' component, then the coefficient β^{BC} on Z_t^{BC} will be consistently estimated. But if β_j^i actually vary substantially across the values of i, then it is clear that no single estimate of β^{BC} obtained in this manner can be consistent. Similarly, if a researcher chooses 3 bands (say), and β^r is inconsistently estimated for each r, then one could spuriously find, or fail to find, a population frequency pattern in β_j^i. On the other hand, such aggregation can lead to coefficient estiamtes which are somewhat inconsistent but more readily interpreted economically; there is a trade-off in this regard.

Another approach is to choose the number and composition of the frequency bands so as to minimize an adjusted goodness-of-fit criterion, such as the BIC. Such a procedure would require Monte Carlo simulation of the sampling distribution of the F statistic for testing equality of the coefficient across frequency bands to account for this quite extensive specification search, which would doubtless yield a test of very low power.

The least restrictive approach is to allow the regression equation to estimate a distinct coefficient for every possible frequency allowed by the limited length of the window used to implement the one-sided filtering. This is feasible given enough time periods of data, but the estimate of the coefficient on the associated variable z_t^j might be quite imprecise. When using monthly data, a parsimonious approach is to model the variation of the k coefficients by means of a lower-order Chebyshev polynomial. Finally, one might well decide a priori that one is not interested in variation at some frequencies.

[11] Of course, an even more typical approach is to exclude frequency dependence from consideration altogether.

In our model for velocity below, we aggregate all frequencies with period length of 4 quarters or lower into one band, but impose no further restrictions. As with considering just a single 'business cycle' band, this choice to aggregate all of the frequency components corresponding to reversal periods less than or equal to 4 quarters in inherently a bit *ad hoc* (and risks inconsistent estimation and concomitant inference distortion), but it yields estimated frequency dependence in β^j which is readily interpretable in economic terms.

4 The Velocity of Money

The velocity of money has been an important topic in macroeconomics for centuries. The first velocity function was posited in 1662 by Sir William Petty; Locke (1691) added the interest rate to this function.[12] Over the next several centuries, a huge literature arose involving the study of velocity (and its alter ego, the demand for money). Velocity was a central concept in two of the most famous studies of money in the 20th century, Fisher (1911) and Friedman (1956), and is central to the celebrated Cagan (1956) model.

The centrality of the velocity concept for monetary policy was unquestioned ... until disaster struck. In the 1980s, velocity in the U.S. fell markedly, and existing money demand functions displayed substantial underprediction. By the 1990s, velocity was increasingly viewed as being neither stable nor predictable. The perceived breakdown of the velocity-interest rate relationship prompted a significant change in the conduct of monetary policy: M2 would no longer be used as a policy target or indicator – despite the fact that the Humphrey-Hawkins Act required that the Federal Reserve specify growth ranges for money and credit (see Carlson, Craig and Schwarz 2000, Barnett 2012, and Belognia and Ireland 2015 for discussions of the intellectual history leading to

[12] Humphrey (1993) traces the historical (pre-1911) development of velocity functions.

the decline in the prominence of money in monetary policy discussions).

While velocity no longer plays the central role it once did in monetary policy discussions, it nonetheless continues to be of central interest to monetary economics and policy. For example, its behavior during the Great Recession prompted much discussion, and its behavior going forward is thought to be of crucial importance in managing the exit from the zero lower bound; it remains essential for understanding the welfare cost of inflation; and velocity (or money demand) often plays a role in theories which posit that the interest rate channel is not the only means by which monetary policy influences real activity (see, e.g., Lucas 2000, Wang and Shi 2006, Bae and de Jong 2007, Faig and Jerez 2007, Ireland 2009, Liu, Wang and Wright 2011, Ball 2012, Mbiti and Weil 2013, Wen 2014, Belongia and Ireland 2015, and Anderson, Bordo and Duca 2015).

Cochrane (2012) recently examined the velocity-interest rate relationship. As he notes, the standard theory of money demand holds that velocity $V = Y/M$ rises – that is, money demand falls – when interest rates rise. (In stochastic equilibrium models such as those of Lucas (2000) or of Teles and Zhou (2005), the inverse relationship between velocity and the interest rate is not precisely money demand, but is rather more properly described as an equilibrium relationship.) Following Cochrane, Figure 2 plots velocity against interest rates – more specifically, it plots $(Py/M^*) - 1$ (where Py is nominal GDP, and M^* is the St. Louis Fed MZM series)[13] against $i/10$, where i is the 3-month Treasury bill rate. Inspection of Figure 2 indicates that the prediction of the standard theory appears to hold up pretty well: velocity and the interest rate are clearly related at low frequencies.[14]

[13] We plot the series "Velocity of MZM Money Stock" from the Federal Reserve Bank of St. Louis. Cochrane notes, "There is an issue of what monetary aggregate to use and how to incorporate the vast expansion of highly liquid interest bearing assets. The St. Louis Fed's MZM definition tries to take account of this fact. The point here being econometric and not about the deep theory of money, I won't pursue the question." Teles and Zhou (2005) present a coherent argument in favor of using MZM; see also Belongia and Ireland (2015).

[14] Reynard (2012) also highlights the low-frequency relationship – a relationship which is, in fact, at the heart of "P*" models of inflation (see Orphanides and Porter, 2000). However, he makes use of the two-

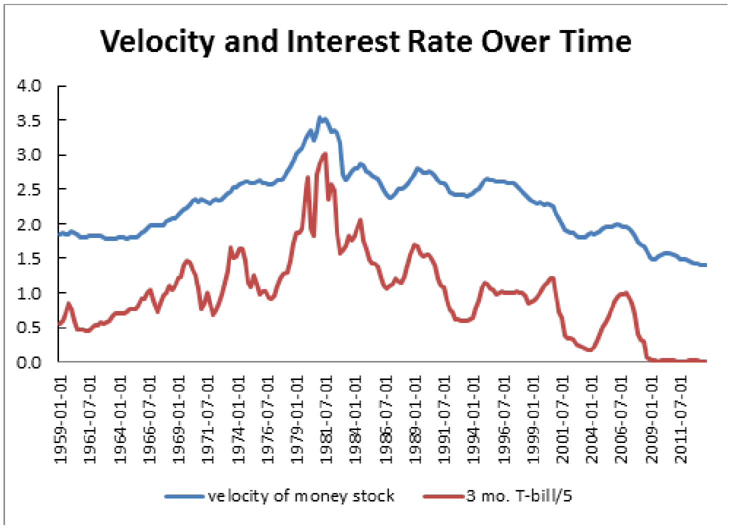

Figure 2: Velocity and the interest rate

However, both data series are highly persistent, and conventional unit-root tests suggest that both series are I(1). Standard econometric practice would suggest modeling the relationship in first-differences,[15] as in

$$\Delta V_t = \alpha + \beta \Delta i_{t-1} + e_t \tag{6}$$

Table 1 presents the results from the specification (6) over the period 1963:1-2011:3.[16]

Table 1. Regression of ΔV on Δi

sided HP filter, compromising his conclusions. As Ball (2012) notes, the conventional wisdom is that the long-run demand for money is stable – if one interprets long-run demand as a cointegrating relation – but the short-run demand for money is unstable.

[15] Or, perhaps, would suggest quasi-differencing the data (i.e. using GLS). Given the level of persistence in these data, there is little difference between the two practices. We sidestep the issue of cointegration for the moment, since we wish to focus attention on the relationship at various different frequencies, rather than the (distinct) issue of whether deviations from the trendline relationship impact the relationship between changes. Note that our procedure in no way precludes inclusion of an error-correction term in Equation 6. In fact, we examine such terms in our paper on the Taylor rule, Ashley, Tsang and Verbrugge (2015).

[16] Cochrane estimates this relationship over the period 1959:I-2011:III. Results are not materially different. In our analysis below, we must start our estimation period somewhat later; so for consistency throughout the paper, we use a common estimation period.

	OLS, Conventional s.e.	OLS, Newey-West s.e.
$\widehat{\alpha}$	-0.001 (0.004)	-0.001 (0.006)
$\widehat{\beta}$	0.027^{***} (0.005)	0.027^{***} (0.005)
Durbin-Watson statistic	1.44	
$\overline{R}^2$	0.14	

Newey-West standard errors are computed with Newey/West/Bartlett Window and 5 lags.

As Cochrane notes, these results suggest that the relationship between velocity and the interest rate, while statistically significant, is quite weak; the estimated coefficient $\left(\widehat{\beta}\right)$ is far smaller than the "sensible" estimate derived from a regression involving levels of both variables, and the regression $\overline{R}^2$ is far smaller as well. The persuasive correlation in Figure 2 has been lost.[17]

What happened? Cochrane argues that measurement error can overwhelm the signal when data are differenced. He then argues that this "loss of signal" problem is why much recent research in finance has given up on "efficient" estimation in favor of running OLS regressions on levels data and correcting standard errors.[18]

Clearly the relationship between velocity and the interest rate is simply different at high frequencies than at low frequencies. Differencing the data thus leads to estimates of the high-frequency relationship – which, in the present case, is weak. But, using our new tools, we can do better. Rather than running the regression on levels data (and almost certainly incurring spurious regression), we next proceed to simultaneously estimate the relationship at every frequency, letting the

[17] Ironically, a number of researchers preferred the first-difference specification inasmuch as it appeared to generate more stable money demand function (see, e.g., Gordon 1984). Friedman (1959) indirectly called attention to persistence dependence in the velocity-interest rate relationship, by stating that the historical stability of velocity – i.e., its failure to mimic the volatility of interest rates – was evidence against the view that money demand was highly sensitive to the interest rate.

[18] Cochrane's note is full of insight and advice, and we enthusiastically recommend it, with the caveats we discuss herein.

data speak as to how the relationship varies across frequencies. This application of our modeling technique yields additional insight into the nature of the relationship between velocity and the interest rate. Further, our results demonstrate that – if frequency dependence in the relationship is accounted for – then first-differencing the data need not lead to weak or noninformative models.

We decompose Δi_t by frequency, using a window length equal to 16 quarters. As noted above, the number of bands (k) is reduced to five by aggregating all of the frequencies corresponding to reversal periods less than or equal to 4 quarters. The frequencies and periods associated with each band are listed in Table 2.[19] Velocity and interest rate data run from 1959:I-2011:III; given first-differencing and the 16-quarter window, the $z_t^{\{j\}}$ run from 1963:I-2011:III.[20]

[19] So as to deal with the "edge effect" discussed in footnote 8 above, each window uses 12 actual observations on Δi_t and is augmented by 4 projected quarters of "data." The projection model used here is the average of two univariate forecasting models of Δi_t: an AR(4) model, and an ARMA(2,2) model.The ARMA(2,2) model is chosen by the HQ criterion. Results are nearly identical if one uses solely the AR(4) model.

[20] I.e., analogous with using lagged variables in a regression model to account for dynamics and eliminate serial correlation in the model errors, our use of a 16-quarter moving window so as to make our bandpass filtering one-sided causes a "start-up" loss of 12 observations.

Table 2: The 5 persistence/frequency components into which Δi is partitioned

band	frequencies	periods	# of frequency (X^*) components	Row Number in A matrix
$z_t^{\{1\}}$	0-$\frac{\pi}{8}$	>16.0	1	1
$z_t^{\{2\}}$	$\frac{\pi}{8}$	16.0	2	2,3
$z_t^{\{3\}}$	$\frac{2\pi}{8}$	8.0	2	4,5
$z_t^{\{4\}}$	$\frac{3\pi}{8}$	5.3	2	6,7
$z_t^{\{5\}}$	$\frac{\pi}{2}$-π	2-4	9	8...16

Allowing for persistence/frequency dependence in the relationship, equation (6) now takes the form

$$\Delta V_t = \alpha + \sum_{i=1}^{5} \beta_i z_t^{\{i\}} + e_t \tag{7}$$

with OLS results given in the first column of Table 3.

We also perform some robustness checks, such as including ΔV_{t-1} (which addresses autocorrelation in the residuals, but also changes the interpretation of the coefficients), allowing for coefficients to be different in a subsample running from 1980-1983,[21] and aggregating $x^2 - x^4$ in one specification.[22] These alternative estimation results are reported in Table 3.

[21] Lucas and Nicolini, citing the work of Teles and Zhou (2005), state: "[T]he early eighties were a hectic period in terms of regulatory changes."
[22] A formal test of the equality of these coefficients yields a p-value of 0.69.

Table 3: Frequency dependence in the estimated velocity-interest rate relationship

	OLS	OLS *HAC s.e.*[a]	OLS *HAC s.e.*[a]	OLS *HAC s.e.*[a]	OLS *HAC s.e.*[a]
constant	-0.002 0.003	-0.002 0.004	-0.001 0.003	-0.001 0.003	-0.002 0.003
$z^{\{1\}}$	0.12** 0.019	0.12** 0.014	0.08** 0.016	0.08** 0.015	0.08** 0.020
$z^{\{2\}}$	0.10** 0.022	0.10* 0.049	0.08 0.044	0.05* 0.024	
$z^{\{3\}}$	-0.01 0.015	-0.01 0.013	-0.01 0.013	0.04** 0.013	0.04* 0.018
$z^{\{4\}}$	0.04** 0.012	0.04** 0.009	0.05** 0.013	0.03** 0.012	
$z^{\{5\}}$	-0.01 0.007	-0.01 0.009	0.00 0.010	0.00 0.007	-0.00 0.017
ΔV_{t-1}			0.26** 0.079	0.27** 0.069	0.27** 0.092
Coefficient change[b]				X	
Durbin-Watson statistic	1.84		2.28	2.34	2.29
$\overline{R}^2$	0.41		0.45	0.53	0.39
$F\ (H_0 : \beta_1 = ... = \beta_5)$ *(significance level)*	23.9 0.00	4.46 0.01	17.4 0.00	5.28 0.00	3.82 0.02

Notes:

[a]HAC standard errors are computed with the Newey-West/Bartlett window and 5 lags.

[b]In this regression, $\beta_1 ... \beta_5$ were all allowed to shift for the sub-sample 1980I to 1983IV.

There are two findings. First, we find a statistically-significant frequency-dependence in the velocity-interest rate relationship using all of these specifications. Second, we can characterize the nature of that frequency dependence. Across all specifications, the strongest relationship between velocity and the interest rate is a statistically significant positive one, at the lowest frequency – i.e., for those fluctuations with a reversal period of at least four years. But there is no evidence for a relationship at high frequencies – i.e., for fluctuations in the change of the interest rate with a reversal period equal to a year or less.

Because we find no evidence for a relationship at high frequencies, our results are consistent with a small estimated coefficient $\left(\widehat{\beta}\right)$ in the model specification (6), a first-differenced model that does not allow for persistence/frequency dependence in this coefficient. Thus, our findings are consistent with Cochrane (2012)'s interpretation: this apparently quite weak relationship in first-differenced data mainly reflects a low cross-correlation between the measurement errors in the money and velocity change variates. But Cochrane's result (essentially replicated here in Table 1) only hints at the true nature of the relationship (displayed here in Table 3) in which there is a strong positive relationship between money and velocity – even in this differenced formulation – at low frequencies, corresponding to reversal periods larger than four quarters.

5 Conclusion

Economic theory frequently implies that the low-frequency relationship between two variables differs from their high-frequency relationship. But inference from linear tools is not valid under these circumstances. Hence, the use of those tools readily leads to erroneous conclusions. For example, the relationship between two variables may appear to be weak or nonexistent if the high-frequency relationship is weak and first-differencing is necessary.

Also, we note above that common practices, such as pre-filtering using the HP filter, are also invalid – because two-sided filtering leads to inconsistent estimation if there is feedback in the relationship. But recent developments in econometric theory allow one to properly circumvent these difficulties and obtain correct inferences, despite the presence of persistence dependence in a relationship. In particular, here we demonstrate how the nonlinear tools of Ashley and Verbrugge (2009a) can be used to obtain a more complete (and richer) depiction of the relationship between the velocity of money and the interest rate, and to formally test for frequency-dependence in this relationship. We find that the relationship between velocity and interest rates is strong at low frequencies, but essentially nonexistent at high frequencies, even though the data needed to be differenced in this case, since the levels of these two variables are not covariance stationary.

On the surface, first-differencing the data appears to destroy the apparent strength of the relationship, due to the way that first-differencing emphasizes high-frequency variation. But – because we have appropriately allowed for frequency dependence in the relationship – first-differencing does not prevent us from uncovering both the strength of this relationship at low frequencies and its weakness at high frequencies.

6 Appendix

6.1 Pseudo frequency dependence

We here distinguish "true" frequency dependence in a relationship from a superficially similar concept in which the coefficients of the model quantifying the relationship are constant, but the *coherence* (closely related to the magnitude of the cross-spectrum of the variates) is frequency-dependent. This latter notion is used in Geweke (1982), Diebold, Ohanian and Berkowitz (1998), and a host of other studies. These decompositions are mathematically sound, but we call what they measure "pseudo frequency dependence" because – since the underlying model coefficients are assumed constant – such measures do not actually quantify frequency variation in the relationship itself.

A simple example clarifies this distinction. Consider the following consumption relation,

$$c_t = \gamma_1 y_{t-1} + u_t + \phi u_{t-1} \tag{8}$$

$$\begin{pmatrix} u_t \\ y_t \end{pmatrix} \sim NIID\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}\right]$$

The marginal propensity to consume in this relationship is clearly a constant (γ_1) and Fourier transforming both sides of this equation will do nothing to change that – it merely yields a relationship between the Fourier transform of c_t and the Fourier transform of y_{t-1}, still with a constant coefficient γ_1. But the cross-spectrum and coherence functions relating c_t and y_t are *not* constants: by construction, they depend explicitly upon the frequency parameter ω. In particular, Geweke (1982)'s measure of the strength of the linear dependence of c_t on y_{t-1} (a generalization of the coherence function) for this model is:

$$f_{y \to c}(\omega) = \frac{1}{2} \ln \left\{ \frac{\sigma_u^2 \left(1 + \phi^2 - 2\phi \cos(\omega)\right) + \gamma_1^2 \sigma_y^2}{\left[\sigma_u^2 \left(1 + \phi^2 - 2\phi \cos(\omega)\right)\right]^2} \right\}$$

which clearly *does* depend upon frequency so long as the moving average parameter ϕ is not zero.

Evidently, this frequency dependence in Geweke's measure (and in the other 'strength of association' measures based upon the cross-spectrum and the coherence function) is not quantifying the frequency variation in the *c*-*y* relationship itself, since in this case there is none to quantify. So what *is* it doing? These kinds of measures are usually interpreted as quantifying the degree to which the overall R^2 for the equation is due to sample variation at low frequencies versus high frequencies.

Suppose that ϕ is positive, in which case Geweke's measure indicates that low frequencies are important to the R^2 of the relationship. This says nothing about whether consumption and income are differently related at low versus high frequencies – that depends upon the marginal propensity to consume (γ_1), which is constant. Rather, it says that this dynamic relationship transforms serially uncorrelated fluctuations in y_{t-1} and u_t into positively correlated fluctuations in c_t. Alternatively, one could observe that c_t in that case has substantial spectral power at low frequencies, and interpret this result, to paraphrase Geweke (1982, p. 312), as indicating that the white noise innovations in y_{t-1} explain most of this low frequency portion of the variance in c_t.[23]

6.2 Frequency variation implies time variation; whither Wold?

The frequency dependence of γ_1 in (2) implied by the permanent income hypothesis concomitantly implies that γ_1 varies over time, depending upon the type of fluctuation that dominates y at time t. For example, with adaptive expectations, the implication is that γ_1 will be larger if the deviation y_{t-1} has the same sign as the deviation y_{t-2}, so that the deviation y_{t-1} is part of a smooth pattern. Note that this dependence of γ_1 on the recent history of y_{t-1} (and the resulting frequency dependence in γ_1) can thus be viewed as a symptom of unmodeled nonlinearity in the relationship between c_t and y_{t-1}. This aspect of frequency dependence is discussed at some length in Tan and Ashley (1999a); see also Ashley and Verbrugge (2013). Here, the essential point is that this frequency dependence in γ_1 further implies that the value of γ_1 is not a fixed constant, but rather varies over time due to its dependence on the recent past of y_t.

Similarly, viewing equation (2) as part of a bivariate VAR model, the impulse response function for c_t will be a function of past innovations in both equations, and c_t will depend differently on different lags in the y_t innovations. Frequency dependence alters the nature of the impulse response functions. In particular, if there is *no* frequency dependence in the $c_t - y_t$ relationship, then the moving average representation of the c_t process will be a linear function of serially independent innovations; this leads to a set of conventional *linear* impulse response functions in which the change in the expected value of c_{t+n} induced by an innovation in the y_t process of size δ is *unrelated* to the

[23] We do not deny that estimates of gain and phase, *taken together*, might contain much the same information as the sign and magnitude of the estimated coefficients $\widehat{\gamma}_1^j$. However, in practice these spectral measures can be extremely challenging to interpret, as opposed to the straightforward interpretation of (say) a negative coefficient on one of our persistence level components. See Granger (1969), quoted in footnote 2.

values of previous innovations. Conversely, frequency dependence in the $c_t - y_t$ relationship implies that the full moving average representation of the $c_t - y_t$ relationship (and hence, the impulse response functions also) are *nonlinear* functions of serially independent innovations. Thus, in that case, the change in the expected value of c_{t+n} induced by an innovation in the y_t process of size δ does depend on the values of previous innovations.[24] (Of course, the Wold Theorem still guarantees the existence of a linear MA(1) representation for c_t and y_t – and hence of a set of linear impulse response functions for these variables –but the innovations in this linear MA(1) representation are *not serially independent*.)

The following explicit example clarifies this point. Consider the particular case in which the linear moving average (Wold) representation for a series c_t can be approximated by the MA(1) process:

$$c_t = v_t + \gamma_1 v_{t-1}$$

in which the v_t innovation series is generated by the bilinear process:

$$v_t = 0.7 v_{t-2} u_{t-1} + u_t$$

where u_t is serially independent. It is easy to verify that the v_t generated by this bilinear process are serially uncorrelated, so this MA(1) process could in principle be the Wold representation for c_t. Now rewrite the moving average representation of c_t as a function of the current and past values of the serially independent innovations – $u_t, u_{t-1}, ...$ – by repeatedly substituting the bilinear model in to eliminate v_t, v_{t-1}, etc. from the model for c_t. In this way one obtains:

$$c_t = u_t + (\gamma_1 + 0.7u_{t-2} + \text{higher order terms})\, u_{t-1} + (0.7\gamma_1 u_{t-3} + \text{higher order terms})\, u_{t-2} + ...$$

where the higher order terms involve $(0.7)^2 v_{t-4}u_{t-3}$, $(0.7)^2 v_{t-5}u_{t-4}$, and so forth. Continued substitution would further elaborate these terms, but the point is clear: the coefficient on the

[24] See Potter (2000) for a formal treatment of nonlinear impulse response functions.

serially independent innovation u_{t-1} is no longer a constant. Instead, it is $(\gamma_1 + 0.7u_{t-2})$ plus higher order terms. Consequently, the impulse response function at lag one is frequency dependent in the sense discussed here: the coefficient on u_{t-1} will be different when the previous innovation (u_{t-2}) is of the same sign as u_{t-1}. Thus, estimating a linear moving average model for c_t yields an impulse response coefficient estimate at lag one which cannot be stable over time or across frequencies, since c_t responds differently to a lag-one shock which is part of a smooth pattern than to a lag-one shock which has just changed sign from the previous period.

6.3 Frequency components

Table 4: Frequency components for a 16 time-period rolling window

Frequency Component	Frequency	Reversion Period	Row Number(s) in A
1	0	> 16	1
2	$\pi/8$	$16/1 = 16$	2,3
3	$2\pi/8$	$16/2 = 8$	4,5
4	$3\pi/8$	$16/3 = 5.33$	6,7
5	$4\pi/8$	$16/4 = 4$	8,9
6	$5\pi/8$	$16/5 = 3.20$	10,11
7	$6\pi/8$	$16/6 = 2.67$	12,13
8	$7\pi/8$	$16/7 = 2.29$	14,15
9	π	$16/8 = 2$	16

7 References

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme (2014) "Household Earnings and Consumption: A Nonlinear Framework." Manuscript, University of Chicago.

Anderson, Richard G., Michael Bordo, and John V. Duca (2015) "Money and Velocity During Financial Crises: From the Great Depression to the Great Recession." Federal Reserve Bank of Dallas Working Paper 1503.

Ashley, Richard. (1984) "A Simple Test for Regression Parameter Instability," *Economic Inquiry* 22.2, 253-267.

Ashley, Richard, Kwok-Ping Tsang and Randal Verbrugge (2015) "Frequency Dependence in a Real Time Monetary Policy Rule." Manuscript, Virginia Tech.

Ashley, Richard, and Randal Verbrugge (2009a) "Frequency Dependence in Regression Model Coefficients: An Alternative Approach for Modeling Nonlinear Dynamic Relationships in Time Series." *Econometric Reviews*, 28(1–3), 4-20.

Ashley, Richard, and Randal Verbrugge (2009b) "To Difference or Not to Difference: a Monte

Carlo Investigation of Spurious Regression in Vector Autoregressive Models." *International Journal of Data Analysis Techniques and Strategies*, 1.3, 242-274.

Ashley, Richard, and Randal Verbrugge (2014) "The Phillips Curve Coefficient is Frequency-Dependent: A Sharper Look at the NIIRU." Manuscript in preparation, Federal Reserve Bank of Cleveland.

Bae, Youngsoo, and Robert M. de Jong (2007) "Money Demand Function Estimation by Non-linear Cointegration." *Journal of Applied Econometrics* 22, 767-793.

Ball, Laurence (2012) "Short-Run Money Demand." *Journal of Monetary Economics* 59.7, 622-633.

Barnett, William A. (2012) *Getting It Wrong: How Faulty Monetary Statistics Undermine the Fed, the Financial System, and the Economy.* Cambridge, MA: MIT Press.

Baxter, Marianne (1994) "Real Exchange Rates and Real Interest Differentials: Have we Missed the Business Cycle Relationship?" *Journal of Monetary Economics* 33.1, 5-37.

Baxter, Marianne, and Robert King (1999) "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *The Review of Economics and Statistics* 81, 575-593.

Belongia, Michael T., and Peter N. Ireland (2015) "Interest Rates and Money in the Measurement of Monetary Policy." Manuscript, Boston College.

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.

Cagan, P. (1956) "The Monetary Dynamics of Hyperinflation," in M. Friedman, ed., *Studies in* *the Quantity Theory of Money* (Chicago: University of Chicago Press, 1956), 25-117.

Carlson, John B., Ben Craig, and Jeffrey C. Schwarz (2000) "Structural Uncertainty and Breakpoint Tests: An Application to Equilibrium Velocity." *Journal of Economics and Business* 52, 101-115.

Chow, Gregory C. (1960) "Tests of Equality between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28, 591-605.

Christiano, Lawrence J., and Terry Fitzgerald (2003) "The Band-Pass Filter." *International Economic Review* 44.2, 435-465.

Cochrane, John H. (2012) "A Brief Parable of Over-Differencing." Manuscript, University of Chicago.

Dagum, Estela (1978) "Modelling, Forecasting and Seasonally Adjusting Economic Time Series with the X11 ARIMA Method." *The Statistician* 27.3/4, 203-16.

Diebold, Francis, Lee Ohanian and Jeremy Berkowitz (1998) "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies* 65.3, 433-51.

Engle, Robert (1974) "Band Spectrum Regression." *International Economic Review* 15, 1-11.

Faig, Miquel, and Belén Jerez (2007) "Precautionary Balances and the Velocity of Circulation of Money." *Journal of Money, Credit and Banking* 39.4, 843-873.

Fisher, Irving (1911) *The Purchasing Power of Money.* New and Revised edition, New York: Macmillan, 1913, reprinted New York: A.M. Kelley, 1963.

Friedman, Milton (1956) "The Quantity Theory of Money: A Restatement." In M. Friedman (ed.) *Studies in the Quantity Theory of Money.* Chicago: University of Chicago Press, 3-21.

Friedman, Milton (1988) "Money and the Stock Market." *Journal of Political Economy* 96, 221-245.

Geweke, John (1982) "Measurement of Linear Dependence and Feedback Between Multiple Time Series." *Journal of the American Statistical Association* 77, 304-314.

Gordon, Robert J. (1984) "The Short-Run Demand for Money: A Reconsideration." *Journal of Money, Credit and Banking* 16, 403-434.

Granger, Clive W.J. (1969) "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37, 424-438.

Granger, Clive W.J., Namwon Hyung and Yongil Jeon (2001) "Spurious regressions with stationary series." *Applied Economics* 33.7, 899–904.

Hodrick, Robert J., and Edward Prescott, (1997) "Postwar US business cycles: an empirical investigation." *Journal of Money, Credit and Banking* 29, 1-16.

Humphrey, Thomas M. (1993) "The Origins of Velocity Functions." *Federal Reserve Bank of Richmond Economic Quarterly* 79.4, 1-17.

Iacobucci, Alessandra, and Alain Noullez (2005) "A Frequency Selective Filter for Short-Length Time Series." *Computational Economics* 25, 75-102.

Ireland, Peter (2009) "On the Welfare Cost of Inflation and the Recent Behavior of Money Demand." *American Economic Review* 99.3, 1040-1052.

Liu, Lucy Qian, Liang Wang, and Randall Wright (2011) "On the 'Hot Potato' Effect of Inflation: Intensive Versus Extensive Margins." *Macroeconomic Dynamics* 15 (supp. 2), 191-216.

Locke, John. (1691) *Some Considerations of the Consequences of the Lowering of Interest and*

Raising the Value of Money. In Vol 5 of *The Works of John Locke*, pp. 3-116. London, 1823, reprinted, Aalen (Germany), 1963.

Lucas, Robert E., Jr. (2000) "Inflation and Welfare." *Econometrica* 68.2, 247-274.

Lucas, Robert E., Jr., and Juan Pablo Niclolini (2015) "On the Stability of Money Demand." Federal Reserve Bank of Minneapolois Working Paper 718.

Mbiti, Isaac, and David M. Weil (2013) "The Home Economics of E-Money: Velocity, Cash Management, and Discount Rates of M-Pesa Users." *American Economics Review* 103.3, 369-374.

Orphanides, Athanasios, and Richard D. Porter (2000) "P* Revisited: Money-Based Inflation Forecasts with a Changing Equilibrium Velocity." *Journal of Economics and Business* 52, 87-100.

Petty, William. (1662) *Treatise of Taxes and Contributions*. In Charles H. Hull, ed., *The Economic Writings of Sir William Petty*, Vol 1. Cambridge, 1900, reprinted, New York: A.M. Kelly, 1964.

Ramsey, James B., and Camille Lampart (1998) "Decomposition of Economic Relationships by Time Scale Using Wavelets." *Macroeconomic Dynamics* 2 (1), 49-71.

Reynard, Samuel (2012) "Assessing Potential Inflation Consequences of QE After Financial Crises." Peterson Institute for International Economics WP 12-22.

Tan, Hui-Boon, and Richard Ashley (1999a) "On the Inherent Nonlinearity of Frequency Dependent Time Series Relationships." in (P. Rothman, ed.) *Nonlinear Time Series Analysis of Economic and Financial Data*, Kluwer Academic Publishers: Norwell, 129-142.

Tan, Hui-Boon, and Richard Ashley (1999b). "Detection and Modeling of Regression Parameter Variation Across Frequencies With an Application to Testing the Permanent Income Hypothesis."

Persistence Dependence in Empirical Relations

Macroeconomic Dynamics 3, 69-83.

Teles, Pedro, and Ruilin Zhou (2005) "A Stable Money Demand: Looking for the Right Monetary Aggregate." *Federal Reserve Bank of Chicago Economic Perspectives* 1Q/2005, 50-63.

Wang, Weimin, and Shouyong Shi (2006) "The Variability of Velocity of Money in a Search Model." *Journal of Monetary Economics* 53, 537-571.

Wen, Yi (2014) "Evaluating Unconventional Monetary Policies: Why Aren't They More Effective?" Federal Reserve Bank of St. Louis Working Paper 2013-028B.