PITFALLS IN THE ANALYSIS OF
FREQUENCY FILTERED TIME SERIES

Steven N. Durlauf

TECHNICAL REPORT NO. 32
February 1988

Prepared Under
NATIONAL SCIENCE FOUNDATION GRANT No. SES 85-20313

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
PITFALLS IN THE ANALYSIS OF
FREQUENCY FILTERED TIME SERIES

Steven N. Durlauf

TECHNICAL REPORT NO. 32
February 1988

PREPARED UNDER

NATIONAL SCIENCE FOUNDATION GRANT No. SES 85-20313

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Abstract

This paper explores a number of issues associated with analyzing frequency filtered time series - series where a positive Lebesgue measure of frequencies are removed from the data. These filters generate a number of interpretational difficulties in the context of structural models. First, in the stationary case, the estimates from linear regression will in general be inconsistent. Second, the data will be rendered deterministic, which implies that a researcher will no longer be able to adjust his model specification in response to the data. Third, if the data are integrated, then filtering will be an inefficient way to look for long run relationships. These arguments imply that frequency filtering should be employed cautiously.
PITFALLS IN THE ANALYSIS OF FREQUENCY FILTERED TIME SERIES

by

Steven N. Durlauf

In an important and influential paper, Robert Engle (1974) developed the technique of band spectrum regression. Engle observed that in a correctly specified model, such that

\[ y_t = x_t^b + \varepsilon_t \quad x_t, \varepsilon_t \text{ independent} \]

that \( \hat{b}_{OLS} \) is not alone as a consistent estimator of \( b \). In particular, if \( y_t \) and \( x_t \) are converted to new series \( y^*_t \) and \( x^*_t \) by transforming their respective spectral densities such that

\[ f_{y^*_t}(\omega) = f_y(\omega); f_{x^*_t}(\omega) = f_x(\omega) \text{ if } \omega \in S \]

\[ f_{y^*_t}(\omega) = f_{x^*_t}(\omega) = 0 \text{ otherwise,} \]

the regression coefficient \( \hat{b}_S \) estimated from

\[ y^*_t = x^*_t^b + \varepsilon^*_t \]

will converge to \( b \) as well. In other words, if data are passed through a frequency filter, consistency is preserved when (1) is the correct model.

On the other hand, for alternative hypotheses to (1) such as

\[ H_1: \text{cov}(x, \varepsilon) \neq 0 \]

*I would like to thank Deborah Johnston and Rebecca A. Robinson for preparing the manuscript. This paper has benefited greatly from discussions with Robert Hall, Thomas Sargent and John Taylor. David Bizer has provided helpful comments. All errors are my own. NSF grant SES 85-20313 is gratefully acknowledged.
or

\[ H_1: y_t = \int_{-\pi}^{\pi} e^{it\omega} b(\omega) \, dz_x(\omega) + \epsilon_t \]

\( \hat{b}_S - b_{OLS} \) or \( \hat{b}_S - \hat{b}_S \) need not converge to zero. Filtered regression coefficients could therefore provide useful tests of misspecification.

Alternatively, if the variable \( x \) is measured with error, and if the variance of \( x \) is concentrated at low frequencies, then after filtering out high frequencies, a more accurate estimate of \( b \) could be obtained. Engle's original paper provided such an example, where the low frequency components of income are treated as permanent part relevant to consumption. A recent example of the use of filtering to mitigate measurement error is Poterba and Summers (1982) who employ frequency filtering to guard against errors in measuring Tobin's \( q \).

A second use of frequency filtering has centered on the testing of trend macroeconomic relationships. Many theoretical relationships, such as the neutrality of money, purchasing power parity, or diminishing marginal product of labor, are expected to hold only in the long run. Assessing the validity of such relationships with time series data is therefore difficult, due to the high frequency components of the data. This concern has led applied researchers to attempt to uncover long run relationships through the removal of high frequencies. The hope of a number of authors (see Lucas (1980), Summers (1983), Geweke (1986)), has been that frequency filtering will provide clearer tests of particular macroeconomic hypotheses.
The purpose of this paper is to argue that for standard macroeconomic contexts, frequency filtering is an inappropriate technique for analyzing time series properties. We shall derive circumstances where the consistency of coefficient estimates of a correctly specified model is not preserved after filtering. As a result, filtered regression coefficients will be inconsistent even when OLS coefficients are consistent. Specification tests which assume that \( \hat{\beta}_{\text{OLS}} - \hat{\beta}_{\text{S}} \) converges to a zero vector under the null will therefore erroneously lead to model rejection.

In particular we analyze the properties of band spectrum regression for time series contexts currently popular in applied macroeconomics where the modelling assumptions underlying Engle's analysis are relaxed. We shall demonstrate that for very general time series models, band frequency filtering will generate uninterpretable estimates. In particular, we shall describe a number of macroeconomic problems where, due to the imposition of orthogonality rather than independence conditions on regressors and errors, do not fall under Engle's original model. Low frequency filtering of these time series may not only be irrelevant, but harmful. In particular, we argue that inconsistent estimates may result.

The argument that frequency filtering can lead to inferential problems is most clearly seen when \( S = [k,-k] \). In this case, frequency filtering is equivalent to a two sided convolution of the data with the sequence \( \{b_j\} \) such that \( b_j = -b_{-j} = \frac{\sin(jk)}{j} \). As discussed in Sims (1974), in the context of distributed lag estimation consistency, ordinary least squares estimation of time series is in general not preserved after convolution when regression
errors Granger cause regressors. This paper will therefore formalize the dangers of such convolutions to statistical inference.

In addition, the arguments in this paper will apply to more general contexts than the sorts of filters discussed. For example, the use of band pass filters for seasonal adjustment will also generate inconsistent estimates if a positive measure of frequencies is distorted.

The paper is organized as follows. Section 2 analyzes the properties of linear regression when errors are orthogonal to regressors prior to frequency filtering but not after frequency filtering. The difficulties in estimation when regressors are predetermined was initially recognized by Hansen and Hodrick (1980) and Cumby, Huizinga, and Obstfeld (1983) and others in the context of instrumental variables estimation of rational expectations models. We provide examples where the sorts of inconsistency noted by these authors is generated through the use of band pass filters on the data. Section 3 argues that frequency filtered data is not amenable to certain types of multiple regression analysis, since the transformed time series are deterministic. Section 4 relaxes the stationarity assumptions for the data. Regressors are allowed to possess stochastic trends. We shall demonstrate that if stochastic trends exist, then the long run dynamics of the system will show up both in the time and frequency domains. Thus the trend cycle decomposition is irrelevant to the interpretation of the data. Section 5 will provide summary and conclusions.
2. Filtering with Predetermined Regressors

This section argues that for many of the sorts of time series analyses normally performed in macroeconomics, frequency filtering of data will generate inconsistent estimates. This inconsistency stems from the relatively weak conditions which macroeconomic theory imposes upon the relationship between regressors and errors. In particular, regressors will be orthogonal to but not independent of the underlying model errors. An obvious example of this feature occurs when regressands Granger cause regressors. However, whereas frequency filtering will preserve the independence of two time series, it will not preserve orthogonality. Therefore inconsistent regression estimates may occur with filtered data. Consider the model

\[ y_t = x_t^\beta + \varepsilon_t \]

where the vector \( v_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} \)

possesses a Cramer representation

\[ v_t = \int_{-\pi}^{\pi} e^{i\omega t} d\zeta(\omega) \text{ where } E[\zeta_j(\omega_1)\zeta_k(\omega_2)] = f_{jk}(\omega_1)\omega_1 \text{ if } \omega_1 = \omega_2, \]

\[ 0 \text{ otherwise} \]

The conditions for consistency of OLS require that \( E(x_t \varepsilon_t) = 0 \forall t, \)

\[ \text{i.e.,} \]

\[ \int_{-\pi}^{\pi} f(x_t \varepsilon_\omega) d\omega = 0 \]
The discussion of this section will derive from the recognition that eq. (3) does not imply that

$$\int_S f_{xt}(\omega)d\omega = 0$$

for any set $S$ contained in $[-\pi, \pi]$. We immediately have

**Proposition 2.1**

Frequency filtered regressions will in general produce inconsistent estimates even if the nonfiltered regression is consistent, when errors and regressors are orthogonal but not independent.

As a result, band spectrum regression will be inconsistent for a wide variety of cases. As noted in the Introduction, the distinction between predetermined and exogenous regressors possesses a critical effect on the consistency of regression estimates obtained under autocorrelation corrections. Recall that autocorrelation in the time domain is equivalent to heteroskedasticity in the frequency domain. Observe further that band spectral regression is a frequency domain heteroskedastic correction (applying a weight of 1 to frequencies of interest and a weight of zero to other frequencies). It is clear that the inconsistency of band spectral regression derives from identical considerations as the Hansen and Hodrick, et al. results.

In order to demonstrate the significance of the potential inconsistency, we consider several examples.

**Example 1.** Band Spectrum Regression with White Noise
Suppose that \( \{ \varepsilon_t \} \) is a white noise \((\sigma, \sigma^2)\) process. A regression of \( \varepsilon_t \) on \( \varepsilon_{t-1} \) is, of course, consistent and will converge to zero. However, for any low band pass filter \([-k,k] \in [-\pi, \pi]\), a band spectrum regression will generate a non zero asymptotic regression coefficient.

This can be seen directly, when one considers the Cramer representation of \( \varepsilon_t \)

\[
\varepsilon_t = \int_{-\pi}^{\pi} e^{it\omega} d\omega
\]

\[
E(d\omega)^2 = \frac{\sigma^2}{2\pi}
\]

\[
E(d\omega_1 d\omega_2) = 0 \quad \omega_1 \neq \omega_2
\]

The band spectrum regression coefficient \( b_k \) corresponding to the \([-k,k]\) filter may be written (where we employ population moments)

\[
b_k = \frac{\int_{-k}^{k} e^{-i\omega} d\omega}{\int_{-k}^{k} d\omega} = \frac{\sin k}{k}
\]

For all \( k < \pi \),

Note, that as \( k \to 0 \), \( b_k \to 1 \). Thus the analysis of low frequency properties of a series could lead one to regard white noise as a random walk! This is not surprising, when one considers that all \( \varepsilon_t \) values possess a common \( dZ(0) \) component.

This example demonstrates how the tests of propositions which isolate low frequency behavior can give completely erroneous results.
Example 2. Long Run Relationships in Macroeconomics

Our next example reexplores the low frequency behavior of the feedback between macroeconomic time series which were originally discussed in McCallum. The first relationship was attributed to Lucas (1980). Suppose that the time series for inflation and money growth may be expressed as:

\[ \pi_t = b(L)m_t + \varepsilon_t \quad \pi_t = \text{inflation rate} \quad m_t = \text{money growth rate} \]

The long run feedback from money to prices in the time domain is \( \Phi_i \). As Lucas argued, a result that \( \Phi_i < 1 \) does not imply that the quantity theory is false, since (4) is a reduced form. Lucas proposed a variant of band spectrum regression to capture long run feedback. An extreme case of such a filtering procedure would calculate the zero frequency regression coefficient \( \beta = \frac{\pi_m}{m_0} \). McCallum provided an example where \( \beta = \Phi_i \), suggesting the irrelevance of the filtering procedure. Yet Lucas did find that low frequency analysis provided information which differed from time domain analysis. Does this suggest the low frequency analysis was instructive?

In fact, it is easy to construct examples where the quantity theory fails, in that money growth never fully feeds into price growth, yet the zero frequency regression coefficient suggests that it holds or indeed that the Mundell-Tobin effect occurs. Suppose that the money supply path follows an invertible AR process

\[ m_t = \rho(L)m_t + \eta_t = \gamma(L)\eta_t \]

The zero frequency regression coefficient \( \beta \) will therefore equal
\[ \frac{f_{mm}(0)}{f_m(0)} = D_i + \frac{D_i \ f_{\varepsilon \eta}(0)}{E \gamma_i \ f_{\eta}(0)} \]

The value of the zero frequency coefficient will exceed the value of \( \Sigma b_i \) so long as \( f_{\varepsilon \eta}(0) > 0 \). The properties of the cross spectrum of innovations of macroeconomic models are in virtually all cases unconstrained by theory. The McCallum example imposed a value of zero.

As a second low frequency hypothesis test, we consider the Fisher equation. Low frequency testing of the Fisher equation, as proposed by Summers (1983) and explored in detail by Barshy (1987), examined the zero frequency feedback between inflation and nominal rates. Similar considerations to the above example apply. For differing properties of the low frequency cross spectrum of interest rate and inflation innovations different feedback properties will result. For example, if we consider the equation

\[ r_t = C + \beta E \pi_t |_{t+1} + \varepsilon_t \]

\[ \pi_t = \gamma(L) \eta_t \]

\[ \varepsilon_t, \eta_t \] white noise

where \( \beta=1 \) under the null hypothesis, the zero frequency band spectrum regression coefficient will converge under the null to

\[ 1 + \frac{1}{E \gamma_i \ f_{\eta}(0)} \]

where \( \pi(L) = \left( \frac{\gamma(L)}{L} \right) \)
Again, the zero frequency cross spectrum will depend upon the specification of the structural relations which make up the inflation and interest rate innovations. For example, a negative value for $f_{\epsilon \eta}(0)$ is consistent with a positive relation between nonstochastic productivity and the real rate, and a negative relation between nonstochastic productivity and nonstochastic inflation.

**Example 3.** Multiple Regression with a Filtered Regressor

This example considers a common use of low band filtering to eliminate measurement error. Examples employing this idea include Poterba and Summers (1982) who employ a filtered series for Tobin's $q$ in order to estimate an investment equation.

For purposes of illustrating the impact of this sort of filtering on regression consistency, let

$$y_t = x_t^b + \epsilon_t$$

represent the true model and

$$\hat{y}_t = x_t^\ast b + \hat{\epsilon}_t$$

equal the regression relationship, where filtered $x_t^\ast$ is employed. It is easy to show that asymptotically

$$\hat{b} - b = [\int_{-\pi}^{\pi} f_{x^\ast x^\ast}(\omega)d\omega]^{-1} \int_{-\pi}^{\pi} f_{\epsilon x^\ast}(\omega)d\omega$$
\[ \int_{-\pi}^{\pi} f_{\varepsilon X}^*(\omega) d\omega = \begin{bmatrix} \text{cov}(\varepsilon X_1^*) \\ \vdots \\ \vdots \\ \text{cov}(\varepsilon X_k^*) \end{bmatrix} \]

which implies that if any element of this vector possesses a non-zero value, the potential for inconsistency exists. In addition, inconsistency will be transmitted to all coefficients, in an analogous fashion to measurement error in the multivariate model.

Two issues vis a vis the use of filtered regressors exist. The first point is that filtering will not necessarily reduce the correlation of regressors and regression errors. If

\[ \varepsilon_t = -b\eta_t + \xi_t \]

where \( \eta_t \) = measurement error in \( x_t \), \( \xi_t \) = structural error, then a filter may increase the covariance, if \( \int_{-k}^{k} f_{\varepsilon x}^*(\omega) d\omega \neq 0 \). The idea that low frequency components dominate the spectrum of \( x \) has no implications for the magnitude of this integral.

Secondly, observe that if the employment of band filtering were to reduce the magnitude of regressor covariance with residuals, in the sense that

\[ (\int_{-\pi}^{\pi} f_{\varepsilon X}^*(\omega) d\omega)_i < (\int_{-\pi}^{\pi} f_{\varepsilon X}^*(\omega) d\omega)_i \]

one still cannot argue that in any sense the bias of a particular coefficient
\[
(f^\pi_{-\pi} f_{\ast} (\omega) d\omega)_{i} \leq (f^\pi_{-\pi} f_{\ast} (\omega) d\omega)_{i} \|i
\]

one still cannot argue that in any sense the bias of a particular coefficient will be reduced, since

\[
\int_{-\pi}^{\pi} f_{x} f_{x} (\omega) d\omega \neq \int_{-\pi}^{\pi} x_{x} (\omega) d\omega.
\]

A simple example of how a reduction in regressor error covariance may lead to an increase in bias occurs if \(k = 1\). The magnitude of \(\hat{b} - b\) is determined by the ratios

\[
\frac{\int_{-\pi}^{\pi} f_{x} (\omega) d\omega}{\int_{-\pi}^{\pi} f_{x} (\omega) d\omega} \quad \text{and} \quad \frac{\int_{-\pi}^{\pi} f_{x} x_{\epsilon} (\omega) d\omega}{\int_{-\pi}^{\pi} f_{x} (\omega) d\omega}
\]

If filtering the regressor increases the left hand side of the expression more than the right, filtering will exacerbate the asymptotic inconsistency of \(\hat{b}\). This is equivalent to reducing the variance of the signal relative to the variance of the noise, which of course exacerbates the errors-in-variables problem.

In closing this section, it should be observed that the difficulty in employing band spectrum regression with lagged dependent variables and independent exogenous variables is very clearly discussed in Engle (1980)
where consistent corrections are obtained. These corrections are straightforward since the cross spectrum between the lagged dependent variables and errors is implicitly determined by the model. Consistent estimation is substantially more difficult for general predetermined regressors, since it is difficult to determine appropriate instruments for filtered data.

In fact, without the introduction of some prior knowledge of components of the cross spectral density of $f_{\epsilon \epsilon}(\omega)$, it is impossible to construct band limited coefficients which permit hypothesis testing of coefficient stability for any band.

Proposition 2.2

There does not exist any set of frequencies $S$ such that for any

$$f_{\epsilon \epsilon}(\omega) \text{ a cross spectral density}$$

$$\hat{\beta}_S - \hat{\beta}_{OLS} \Rightarrow 0$$

if $y_t = x_t \beta + \epsilon_t$ is the correct model, but fails to converge to zero when the relationship between $y$ and $x$ depends upon frequency, i.e.

$$y_t = \int_{-\pi}^{\pi} e^{i\omega} \beta(\omega)dz(\omega) + \epsilon_t$$

pf. Since $\hat{\beta}_{OLS} \Rightarrow \beta$ under the null, the first part of the proposition will only hold if there exists a band $S$ such that

$$\int_S f_{\epsilon \epsilon}(\omega) = 0 \quad \forall \ f_{\epsilon \epsilon}(\omega)$$
Since \( f_{x_t}(\omega) = \sum_{j=-\infty}^{\infty} \sigma_{x_r}(j)e^{ij\omega} \) it is clear that \( \int_{-\infty}^{\infty} e^{ij\omega} d\omega \) must equal zero for all \( j \neq 0 \). Otherwise, given \( S \), one could construct another series \( \sigma_{x_r}(j) \) to exploit the failure of all such integrals to equal zero. Such an \( S \), however, generates \( \hat{\beta}_S = \hat{\beta}_{OLS} \) regardless of whether the null or alternative is true, \( \beta_S = \beta_{OLS} \) for all such \( S \), contradicting the second condition.

Furthermore, the standard regression requirement of predeterminedness, i.e. \( \int_{-\pi}^{\pi} e^{-ij\omega} f_{x_t}(\omega) d\omega \) \( j > 0 \), does not by itself generate any candidates for instrumental variables among the lagged \( x \)'s. To see this, consider the set of moments generated by the band limited spectral density \( f_{x_t}(\omega) \) for the case \( y_t = \beta x_{t-1} + \epsilon_t \)

\[
E(x_{t-1}^* x_{t-i}) = \int_S e^{-ij\omega} f_{x_t}(\omega) d\omega
\]

\[
E(y_{t-i}^* x_{t-1}) = \beta \int_S e^{-ij\omega} f_{x_t}(\omega) d\omega + \int_S e^{-ij\omega} f_{x_{t-i}}(\omega) d\omega
\]

\[
E(y_{t-i}^* y_{t-j}) = \beta^2 \int_S e^{-ij\omega} f_{x_t}(\omega) d\omega + 2\beta \int_S e^{-ij\omega} f_{x_{t-i}}(\omega) + \int_S e^{-ij\omega} f_{\epsilon_t}(\omega) d\omega
\]

Since \( f_{x_t}(\omega) \) is unknown, no methods of moment estimators may be constructed. This problem is identical to the classical errors in variables problem where both regressors and measurement error possess unknown time series structures. Additional instruments, such as \( x_{t-i} \) or \( y_{t-i} \), keep introducing nuisance parameters which are functions of unobserved cross...
spectral densities, elements of the form \( \int_{S} e^{-i\omega} \cdot \chi_{\varepsilon}(\omega) d\omega \).

Predeterminedness, which requires such integrals to equal zero for \( S = [-\pi, \pi] \) and \( j > 0 \), places no restrictions on the value of the integral over any other \( S \). The inability of predeterminedness to permit construction of band specific instruments.

**Proposition 2.3**

Knowledge that \( \int_{-\pi}^{\pi} \chi_{\varepsilon}(\omega) d\omega = 0 \) does not identify any polynomial \( g(L \neq 0) \) such that \( \int_{S} g(e^{-i\omega}) \cdot \chi_{\varepsilon}(\omega) d\omega = 0 \) for a prespecified \( S \).

**Section 3. Temporal Dependence of Filtered Data**

In many empirical contexts, the researcher does not possess a fixed model in advance of data analysis. An obvious example is Granger-Sims causality testing where lag lengths are not specified a priori. Another example occurs in classical regression contexts, with attempts to control for residual autocorrelation which implicitly introduce lagged dependent variables into the analysis.

This section argues that band limited data is unsuitable for free form multiple regression contexts, in particular, those contexts which introduce lagged dependent variables as the researcher adjusts the model in response to the data fit. The basic problem is that a band limited time series is deterministic, i.e., the series may be perfectly predicted from its own past. As a result, moving average representations will not exist. Numerous difficulties are then introduced. First, statistical procedures which presuppose the existence of MA representations will be invalid. Second, since
the series is deterministic, a univariate AR model will always provide asymptotically perfect fit, making it difficult to detect cross variable relations. The impact of filtering on the ARMA structure of a time series is verified in:

**Proposition 3.1.**

Any frequency filtered time series will be rendered deterministic if a set of frequencies of positive Lebesgue measure are removed.

This proposition follows immediately from Szegö's condition for a time series to be deterministic. Szegö's condition states that a time series will be indeterministic if and only if

$$\int_{-\infty}^{\infty} (\omega^2 + 1)^{-1} \log f(\omega) d\omega > -\infty$$

where $f(\omega)$ is the spectral density of the absolutely continuous part of the spectral distribution function of the time series. Kolmogorov proved the refinement that if we only consider the interval $[-\pi, \pi]$ the spectral density of an indeterministic series must fulfill

$$\int_{-\pi}^{\pi} \log f(\omega) d\omega > -\infty$$

These conditions will naturally not hold for any series where a positive Lebesgue measure of frequencies are zero, since $\log(0) = -\infty$.

The formulation of an exact prediction formula for deterministic series is nevertheless quite difficult. No general procedures exist for producing exact formula for deterministic prediction based exclusively upon knowledge of
the spectral density. (See Dym and McKeen (1976) for discussion.) However, one can formulate some general expressions for time series which possess zero spectral densities over positive Lebesgue measures of frequencies. In particular,

**Proposition 3.2.**

For any time series $x_t$ with (1) an absolutely summable autocovariance function (2) spectral density such that $f_x(\omega) = 0$ for $\omega \in \mathbb{S}$, a closed interval, a perfect predictor of $x_t$ will possess the form

$$x_t = \pi(L)x_{t-1} = \gamma(L)x_{t-1}$$

such that

$$\pi_j = \int_{-\pi}^{\pi} \sin j \omega \, d\omega \quad \text{or} \quad \gamma_j = \int_{-\pi}^{\pi} \cos j \omega \, d\omega$$

(Length refers to Lebesgue measure)

pf. See appendix. The condition on the convergence rate of the autocovariance function is not very restrictive as the autocovariance function of any ARMA process will possess elements of order $O(k^T)$ and will encompass those long memory processes (defined as $AR(\infty)$ models with $\sigma(h) = 0(T^{2d-1})$) such that $d < 0$. The requirement for stationarity of the series is that $d < .5$, thus the proposition does exclude some processes.

These representations are not unique, since any $x_{t-1}$ can be replaced with $\pi(L)x_{t-1}$ due to the deterministic nature of the entire $x_t$ sequence. Knowledge of the optimal predictors of time series with band limited
spectral densities will provide information about time series with spectral
densities which are smaller than the band limited time series almost
everywhere. We state the following corollary

**Corollary 3.1.**

If \( x_t = \pi(L)x_t \) and there exists a time series \( y_t \) such that

\[
y_y(\omega) < f_x(\omega) \quad \text{a.e., then}
\]

\[
y_t = \pi(L)y_t.
\]

**pf:** See appendix.

In addition, one can parameterize the class of polynomial operators
which provide perfect predictors.

**Corollary 3.2.**

A necessary and sufficient condition for a square summable polynomial
\( \pi(L) \) to fulfill the equality \( x_t = \pi(L)x_t \) in a mean square sense is that
\( \pi(e^{-i\omega}) \) equal the Fourier expansion of some function \( g(\omega) \) which is an
element of the Hardy Space \( \mathcal{H}^2 \) where \( g(\omega) = 1 \) almost everywhere on \( S^c \).

**pf.** See Appendix.

As a result of the deterministic nature of the filtered time series, it
becomes quite difficult to perform many standard multiple regression
exercises. Granger causality tests are likely to support (at least
asymptotically) the hypothesis that no series Granger causes the filtered
series under study. Poor Durbin-Watson statistics are similarly quite likely.
A simple example of the problem occurs when one considers that the time series \( X^0_t \), corresponding to data purged of all but the zero frequency, must possess identical components for all \( t \), since the same \( dZ(0) \) occurs in each \( X_t \). Procedures which strip out all frequencies but zero must generate this property since \( E(x_t, x_{t-i}) = \sigma^2 \) for all \( i \) if \( dZ(0) \) is the only component of the time series left.

The above discussion, while hopefully suggestive, does not necessarily mean that "free form" regression modelling, where the regressor set is adjusted by the researcher in response to data fit, will automatically lead the researcher to accept (asymptotically) the specification of a univariate AR representation. This ambiguity in the impact of using deterministic series arises because band limited filtering of time series places very few restrictions on the off diagonal elements of the spectral density matrix. These nondiagonal elements, of course, are critical in determining the properties of estimators and test statistics.


In order to concretely illustrate the problems of inference we shall concentrate on two examples. The Granger-Sims framework presupposes the existence of a unique decomposition of a random vector into the Hilbert Space generated by its past or equivalently, the existence of a moving average decomposition of the vector - a condition of course violated by frequency filtered data. One would therefore have a strong prior that the causality testing methodology would generate erroneous results.

Consider the bivariate regression
\[ x_t^* = \alpha(L)x_t^* + \beta(L)y_t^* + \epsilon_t \]

where the unfiltered processes follow

\[ x_t = \alpha(L)x_t + \beta(L)y_t + \epsilon_t \]

for nonnull polynomials \( \alpha(L) \) and \( \beta(L) \).

Since \( x_t \) is deterministic, the bivariate regression would appear to be a candidate for spurious acceptance of noncausality. In fact band filtering of data may or may not preserve causal relations between time series, depending upon the cross spectral density. This may be verified through a simple example. Define series

\[ z_t = \text{time series such that } f_z(\omega) = 0 \quad \omega \notin S \]

\[ \eta_t = \text{time series such that } f_\eta(\omega) = 0 \quad \omega \notin S \]

\[ \xi_t = \text{time series such that } f_\xi(\omega) = 0 \quad \omega \notin S \]

\[ x_t = z_t + \eta_t \]

\[ y_t = z_{t+1} + \xi_t \]

\( z, \eta, \xi \) independent.

It is easy to verify that \( y_t \) Granger causes \( x_t \). Further, if frequencies \( S \) are removed \( x_t \equiv y_{t-1} \). The sequence of \( F \) statistics will possess divergent elements.

Conversely, let the spectral densities of \( \eta \) and \( \xi \) equal zero outside \( S \), whereas the spectral density of \( z \) equals zero inside \( S \). In this case, removal of \( S \) will render \( x \) and \( y \) independent.
This ambiguity does not detract from the central message that free modelling of regressions can lead to highly spurious inferences if the underlying time series are deterministic.

In fact, the counterexample to the asymptotic acceptance is rather contrived, in that the band filter generated an exact dependence between the two series. In general, if the filtering does not generate an exact relation, non-causality will likely occur asymptotically in the sense that the test statistic for the test is asymptotically bounded. We conclude this section by verifying the existence of broad classes of timed series, there exist sequences of causality tests where the deterministic nature of the time series determines the hypothesis testing results regardless of the cross spectral density. The results are somewhat ad hoc in that they depend upon the way in which the regressor set is allowed to expand. However, no theory or standardized methodology exists which can be used to circumscribe the appropriate assumptions.

Proposition 3.3.

If lagged, current, and future \( x_t^* \) cannot perfectly predict \( y_t^* \), then any sequence of two sided filters of the form

\[
y_t^* = \alpha(L^{-1}) x_t^* + \beta(L) x_t^* + v_t
\]

will generate asymptotic acceptance of the hypothesis of no causality leading from \( y \) to \( x \), for any infinite polynomial \( \beta(L) \) whose order is at least \( O(t^{1/2}) \) and any polynomial \( \alpha(L^{-1}) \) of lower order than \( \beta(L) \).

pf. See appendix.
The conditions of the proposition are unrestricitive. The first part of the proposition implying that \( x^*_t \) is not an exact linear combination of \( y_t \) ensures that the problem in the example cannot occur, and essentially avoids multicollinearity in the Granger form of the causality test. The requirements on \( \alpha(L^{-1}) \) and \( \beta(L) \) means that as \( T \to \infty \), there exist representations where \( \beta(L) \) dominates the regression. In the Granger framework, the lagged \( x_t \)'s must be given an opportunity to provide all explanatory power.

If a restriction is placed upon the relationship between \( y_t \) and \( x_t \) in the sense that the projection of \( y_t \) onto future \( x \)'s generates coefficients which delay in magnitude at a sufficient rate to render them square summable, then both \( \alpha(L^{-1}) \) and \( \beta(L) \) may be increased without bound, with no requirements that \( \alpha(L^{-1}) \) increase at a faster rate than \( \beta(L) \).

**Proposition 3.4.**

If the one sided projection of \( y_t \) onto \( \mathcal{H}_{x_+}(t, \infty) \) possesses square summable coefficients and if \( k_1(T) = k_2(T) \) are both \( O_p(T^{1/2+\delta}) \) for some \( \delta > 0 \), then the Sims causality test statistic will converge to zero.

pf. See Appendix.

Again, these last three propositions are ad hoc in the sense that the assumptions employed to generate certain acceptance of noncausality are sufficient, not necessary. However the assumptions are not restrictive in the sense that it is easy to construct examples where the conditions hold.

**Example 2.** Durbin-Watson Diagnostics.
Similar considerations exist for the behavior of the Durbin-Watson statistic. For any model with a frequency filtered regressand, there exists a danger of convergence to a univariate AR through autocorrelation corrections.

**Proposition 3.5.** In the frequency filtered regression \( y_t^* = x_t^* b + \epsilon_t^* \), the Durbin-Watson statistic converges to a value not equal to 2 for any formulation with imperfect prediction.

pf. \( y_t^* - x_t^* b = \epsilon_t^* \), will possess a zero spectrum at those frequencies where both the original series possess zero frequencies. Therefore, for all \( T \), \( y_t^* - x_t^* b_T \) will fulfill the conditions of Proposition 3.2. The conditions of the original series ensure that the sample moments \( \frac{\sum \epsilon_t^2}{T} \) and \( \frac{\sum \epsilon_t^2 \epsilon_{t-1}}{T} \) are both convergent. The deterministic nature of \( \epsilon_t^* \) will cause \( \frac{\sum \epsilon_t^2 \epsilon_{t-1}}{T} \to C \neq 0 \). The Durbin-Watson will therefore not converge to 2.

As a result, classical diagnostic procedures will lead researchers to the inappropriate structural model. It is easy to develop examples where the autocorrelation correction will adversely affect references concerning the \( b \) vector. Frequency filtering would therefore appear to be appropriate at best only after the applied worker has chosen a model specification.

**Section 4.** Long Run Relationships, Cointegrated Processes, and Low Frequency Filters

In the zero frequency analyses of Section 2, the appropriateness of low frequency analysis in the independent error, stationary regressor case stems from the lack of a natural decomposition of a stationary, indeterministic time series into trend and cyclical components. A recent
literature on integrated processes has argued that a number of macroeconomic
time series possess stochastic trends. Following decomposition terminology
suggested by Beveridge and Nelson (1981) and Watson (1986), an ARIMA time
series $x_t$ may be composed into trend, $x_{TL}$, and cycle $x_{CT}$ components

$$x_t = x_{Tt} + x_{Ct}$$

where

$$x_{Tt} = x_{Tt-1} + \zeta_t$$

(The decomposition is not unique, but has no effect on our results.)

For time series with integrated trend components, the notion of trend
versus cycle features to the data is substantially more appealing. Low
frequency filtering would appear to represent an appropriate way to capture
long term relationships as opposed to cyclical relationships. Following
Granger (1981), linear dependence among the stochastic trend components of
integrated time series means they are cointegrated.

In this section, we argue that if distinct trend components exist to
macroeconomic time series, standard regression techniques will provide long
run feedback estimates that are asymptotically equivalent to low frequency
filters. In other words, the cyclical high frequency components are
irrelevant to the estimated long run elasticities. This result complements
the work of McCallum discussed earlier.

In order to see how stochastic trend relationships dominate the
estimation of long run time domain elasticities, suppose that the time series \( x_t \) and \( y_t \) may be expressed as:

\[
x_t = x_{Tt} + x_{ct}
\]

\[
x_{Tt} = x_{Tt-1} + \eta_t \quad \eta_t \sim MA(K)
\]

and

\[
y_t = y_{Tt} + \zeta_t
\]

\[
= \beta(L)x_{Tt} + \zeta_t
\]

i.e., there exists a representation of \( y_t \) which renders the stochastic trend an exact function of the trend in \( x_t \). (One can verify that \( x_t \) and \( y_t \) will be cointegrated if such a specification exists.)

We claim the following.

**Proposition 4.1**

In the regression

\[
y_t = \gamma(L)x_t + \nu_t
\]

where \( \gamma_t \) denotes the least squares estimates of \( \gamma(L) \), then

\[
\hat{\gamma}_i \Rightarrow \beta_i
\]
where \( y_t = \beta(L) y_t + \xi \) is the underlying structural relation for any finite polynomials \( \gamma(L) \) and \( \beta(L) \) i.e., the long run regression relationship will converge to the long run trend relationship.

\[
\text{pf. } \frac{\sum (y_t - \gamma(L)x_t)^2}{T} = \frac{\sum ((\sum_{i} (\beta_i - \gamma_i)x_{Tt-k}) + (\beta(L) - \gamma(L)) \theta_t + \gamma(L) \eta_t + \xi_t)^2}{T}
\]

where \( \theta_t = \sum_{i=0}^{k} \gamma_{t-\lambda} \) and \( k = \) maximum of lengths of \( \beta(L) \) and \( \gamma(L) \). All terms in that quadratic form are stationary, except for \( x_{Tt-k} \). The leading term \( \frac{\sum x_{Tt-k}^2}{T} \) is \( O_p(T) \), as verified in Phillips and Durlauf (1986). Therefore, the problem will generate a finite minimum if and only if \( \sum \gamma_i = \sum \beta_i \).

Therefore, we see that if a meaningful distinction exists between trend and cycle, the high frequency terms will not prevent relevation of the trend relation. In addition, as the structure of proof suggests, the convergence rate is \( T^{-1} \).

Conversely, a low frequency filter will provide the exact same asymptotic value for the relationship between \( X \) and \( Y \). Consider the expression

\[
b_0 = \frac{\hat{f}_{xy}(0)}{\frac{\hat{f}_{x}(0)}{\hat{f}_{y}(0)}}
\]

The leading terms of the zero frequency filter will equal
\[ \frac{\hat{f}_x(x,t-1)}{f_x(x)} \text{ for some } \delta > 0 \]

\[ \Rightarrow \hat{f}_x(x,0) \]

as the terms \[ \frac{\hat{f}_x(x)}{f_x(x)} \] all converge to the same random variable regardless of \( i \).

The convergence rate \( \delta \) is a function of the way in which the zero frequency estimates are constructed. Since in the weakly stationary case,

\[ \frac{1}{2\pi} f(0) = \sigma_x(0) + 2 \sum_{i=1}^{\infty} \sigma_x(i) \]

Estimators take the form

\[ \frac{1}{2\pi} \hat{f}(0) = \hat{\sigma}_x(0) + 2 \sum_{i=1}^{d(T)} \hat{\sigma}_x(i) \]

for some function \( d(T) \) of the observations. Different choices of \( d(T) \) will generate different orders of magnitude of \( \hat{f}_x(0) \). Notice that if \( d(T) \) is \( O(T^{\delta}) \) for \( \delta < 1 \), then the zero frequency estimator will converge more slowly than the regression estimator. The standard estimators of the zero frequency employ a truncation lag of \( O(T^{-1/4}) \) or smaller, rendering the zero frequency filter a comparatively inefficient way of revealing the long run trend components.

Thus, we see that the McCallum irrelevance result still applies. The zero frequency estimates are asymptotically equivalent to the long run time domain feedback estimates.
5. Conclusions.

This paper has demonstrated that for a number of macroeconomic contexts, frequency filtering of data is an inappropriate technique. For models where regressors are predetermined rather than strictly exogenous, frequency filtering will generate inconsistent coefficient estimates. For models where theory does not specify the lag structure employed in estimation, the deterministic nature of filtered series may lead to spurious inferences. For models where stochastic trends are embedded in the data, the procedure generates values equivalent to the time domain results.

Frequency filtering thus should be employed with caution. The sorts of environments where the technique is informative are limited to those where the cross covariogram of regressors and errors is identically equal to zero and model specification is not dependent on the data. In absence of such a well defined framework, the use of frequency filtered data creates the potential for highly misleading inferences.
FOOTNOTES

1/ In a distinct critique Bennett McCallum demonstrated that low frequency band spectrum regression does not isolate long run as opposed to short run relationships for well specified time series models. In particular, McCallum showed that in standard regression contexts, with stationary regressors and independent errors, band spectrum regression employing only the zero frequency will generate long run relationships which are asymptotically equivalent to the long run relationships generated by vector autoregression (i.e., the sums of current and lagged variable coefficients). Our arguments differ from McCallum in that we address issues of estimator consistency rather than the interpretation of estimates in the context of particular economic models.

2/ An example of a time series where this formula is not applicable is \( x_t = \varepsilon \Psi_t \), where \( \varepsilon \sim (0,1) \). The spectral density will be divergent as it equals \( \sum_{t=-\infty}^{\infty} e^{it\omega} \).
BIBLIOGRAPHY


APPENDIX

Proof of Proposition 3.1.

This proof is based upon an idea of Whittle (1952) who noted that the
Kolmogorov condition may be thought of as a restriction upon the zero
eigenvalues of the variance/covariance matrix of $X_t$, since the spectral
density, evaluated at a discrete, dense subset of $[-\pi, \pi]$ will equal the
eigenvalues. Our proof will therefore construct a zero eigenvector of the
covariance matrix of the time series.

A standard result in time series analysis is the following:

Theorem A.1 Let $V_T$ = matrix such that $v_{ij} = E(x_i x_j)$ $c_j = 0...T$

$$\omega_i = \frac{2\pi}{T}$$

$$P_T = \begin{bmatrix} S_{0,T} \\ \vdots \\ S_{0,T,T} \end{bmatrix}$$ a square matrix

such that $S_{0,T} = T^{-1/2} [1,...,1]$ $S_{1,T} = T^{-1/2} [1, \sin \omega_1, \sin 2\omega_1,...]$ $i > 0$

and

$$D_T = \text{matrix such that } d_{ij} = f(\omega_i) \text{ if } i = j$$

$$0 \text{ otherwise}$$

Then $P_T V_T P_T - 2\pi D_T$ converges uniformly to zero as $T \to \infty$.

(See Brockwell and Davis (1986) for a proof.)

The elements of $P_T$ will therefore equal eigenvectors corresponding to zero
eigenvalues, when \( f(\omega_j) = 0 \). However, these are not satisfactory candidates for the perfect predictor of \( x_t \), since the elements of the vectors are each \( O(T^{-1/2}) \).

For the closed interval \( S \), we can choose those elements of \( S \) which correspond to elements in the sequence \( \frac{2\pi j}{T} \) \( 0 \leq j \leq T-1 \). There will be \( \text{Length}(S) \) \( \frac{2\pi}{T} \) such elements. We may further choose \( \sqrt{T} \) points within this set for \( T \) large, equally spaced over \( S \). Denote this set of elements \( V \).

It is easy to verify that the set of \( \frac{S_i}{\sqrt{T}} \) vectors evaluated at \( \omega \in V \) will fulfill the following

\[
(1) \quad \pi_T = \sum_{i=1}^{\sqrt{T}} S_i, T \Rightarrow (1, \int_S \sin \omega d\omega, \int_S \sin 2\omega d\omega, \ldots)
\]

where a sequence \( \mathbf{a}_T \) converges to \( \mathbf{b}_T \) if

\[
\sup_i \| \mathbf{a}_i - \mathbf{b}_i \| < \varepsilon \text{ for any } \varepsilon > 0 \text{ for all } i \text{ if } T \text{ large enough.}
\]

pf: Each element of \( \pi \) represents the sum of the form

\[
\pi_j = \frac{\sum_{\omega=1}^{\sqrt{T}} \sin \omega j}{\sqrt{T}}
\]

which will converge to the corresponding Lebesgue integral.

(2) \( \pi(L) \) will possess elements of magnitude

\[
\pi_i = O(1)
\]

pf. \( \int_V \sin j \omega d\omega \leq \frac{1}{j} \int_V \sin j \omega d\omega \leq \frac{\cos j \omega}{j} \leq \frac{2}{j} \)

(Note that this implies that \( \pi_i \to 0 \) as \( i \to \infty \), which is known as the Riemann-Lebesgue Lemma).

(3) \( \pi \) is square summable.
pf: This follows immediately since \( \pi_j \leq \frac{2}{j} \Rightarrow \max \frac{\pi_j^2}{j^2} \text{ and } \frac{1}{j} \) is a square summable sequence. (One can also show that \( \pi_T \) is square summable for all \( T \) and \( \pi_T \) converges in the \( L^2 \) sense as well. These proofs are tedious and omitted. Heuristically, the latter holds because (1) the space of square summable sequences is complete in the \( L^2 \) norm, and (2) \( \pi_T \) is a Cauchy sequence in that norm with \( L^2 \) limit \( \pi \).

Further, each \( S_{i,T} \) is an eigenvector of \( C_T \), the circulant matrix

\[
\begin{bmatrix}
\sigma(0) & \sigma(1) & \sigma(2) & \ldots & \sigma(2) & \sigma(3) & \sigma(1) \\
\sigma(1) & \sigma(2) \\
\vdots & & & & & & \\
\sigma(2) & \sigma(1) & \sigma(2) & \sigma(3) & \sigma(1) & \sigma(0)
\end{bmatrix}
\]

which possesses the property that \( \sigma(r) = \sigma(T-r) \).

Since

\[
C_T S_{i,T} = \lambda_i S_{i,T},
\]

it is obvious that \( (E S_{i,T}) C_T E S_{i,T} = (E S_{i,T}) \lambda_i S_{i,T} \)

\[
C_T E S_{i,T} = E \lambda_i S_{i,T}
\]

i.e. \( \pi_T C_T \pi_T = \pi_T \lambda_i S_{i,T} \)

In addition, one can show that the difference between \( \lambda_i \) and \( f_x(\omega_i) \) is uniformly bounded when the autocovariance function is absolutely summable. Further, absolute summability ensures that \( \text{ess sup } |\lambda_i - f_i(\omega_i)| \to 0 \), since the \( L^\infty \) is coarser than the \( L^1 \) metric. We may therefore conclude that

\[
\pi_T C_T \pi_T = \Sigma (\lambda_i - f_i(\omega_i)) \pi_i^2 \leq \text{ess sup } |\lambda_i - f_i(\omega_i)| \pi_{i,T}^2 \to 0
\]

by Holder's inequality.

This last result ensures that if \( |\pi_T C_T \pi_T - \pi_T V_T \pi_T| \to 0 \), then

\[
|\pi_T V_T \pi_T| \to 0
\]

which implies that \( \pi_T \) is a square summable eigenvector with the desired properties.
But,

$$\left| \sum_{m=1}^{T-1} \frac{\sigma(m)}{T} \sum_{k=1}^{m} (\pi_{i,k,T-m+k} - \pi_{j,T-m+k}) \right|$$

$$\leq \frac{K}{T} \left( \sum_{m=1}^{T-1} m |\sigma(m)| + \sum_{m=1}^{T-1} m |\sigma(T-m)| \right)$$

for some $k$ since each $\pi_{i,T}$ element is $O(\frac{1}{T})$.

Mirroring Brockwell and Davis (1986) pg. 131, this final expression converges to zero, which is the required result. Finally, notice that if $S_{i,t}$ vectors were defined as:

$$T^{-\frac{1}{2}} [1, \cos \omega_j, \cos 2 \omega_j ...]$$

the argument would follow symmetrically, since each eigenvalue of the circulant matrix possesses these two real eigenvectors.

Proof of Corollary 3.1

If the Cramer representation of a random variable $x_t$ is

$$x_t = \int_{-\pi}^{\pi} e^{it \omega} dz(\omega)$$

then $C(L)x_t = \int_{-\pi}^{\pi} C(e^{-i \omega}) e^{it \omega} dz(\omega)$ and the variance of $x_t - C(L)x_t$ will therefore equal

$$\text{var}(x_t - C(L)x_t) = \int_{-\pi}^{\pi} |e^{it \omega} - C(e^{-i \omega}) e^{it \omega}|^2 f_x(\omega) d\omega$$
This expression leads to the formulation of the prediction problem in the
frequency domain as:

Choose a polynomial \( C(L) \) such that
\[
\int_{-\pi}^{\pi} |(e^{i\omega} - C(e^{-i\omega})e^{i\omega})|^2 f(\omega) d\omega
\]
is minimized.

Perfect prediction means that there exists a polynomial \( C(e^{-i\omega}) \) such that
this integral equals zero.

If such a \( C(\cdot) \) exists, then
\[
\int_{-\pi}^{\pi} |(e^{i\omega} - C(e^{-i\omega})e^{i\omega})|^2 \mu(\omega) d\omega = 0
\]
for any nonnegative \( \mu(\omega) \ll f(\omega) \) a.e., which is the corollary.

**Proof of Corollary 3.2.**

The variance of \( x_t - \pi(L)x_t \) will equal
\[
\int_{S^c} (g(\omega) - \pi(e^{-i\omega}))^2 f_x(\omega) \ll \int_{S^c} (g(\omega) - \pi(e^{-i\omega}))^2 d\omega \text{ (ess sup } |f_x(\omega)|) \]
by the Holder Inequality. The definition of a Hardy Function requires the
first integral on the RHS to converge to zero. By the Lebesgue convergence
theorem, \( g(x) \) may be replaced by a constant \( k \), which proves one direction of
the proposition.

Conversely, if
\[
\int_{S^c} (g(\omega) - \pi(e^{-i\omega}))^2 f_x(\omega) d\omega = 0
\]
with \( \pi(\cdot) \) possessing square summable coefficients, then \( \pi(\cdot) \) will always
Proof of Proposition 3.3

Define $RSS_0$ = residual sum of squares if $\alpha(L^{-1}) = 0$

$RSS_1$ = residual sum of squares if $\alpha(L^{-1})$ unconstrained

$k_2(T)$ = order of $\alpha(L^{-1})$, possibly dependent on $T$

$k_1(T)$ = order of $\beta(L)$

The $F$ statistic for the test $\alpha(L^{-1}) = 0$ is

$$\frac{RSS_0 - RSS_1}{RSS_0} T \frac{1}{k_2}$$

By assumption, $\frac{RSS_0}{T}$ is $O_p(1)$. (In fact, it will converge to a constant.) To analyze the numerator, define

$$H_{y-}(t, k_1) = \text{Hilbert space spanned by } y_t \ldots y_{t-k_1}$$

$$H_{y+}(t, k_2) = \text{Hilbert space spanned by } y_{t+1} \ldots y_{t+k_2}$$

$H_z(t, k_2) = \text{Hilbert space orthogonal to } H_{y-}$, which is implicitly defined by

$$H_{y-} \cap H_{y+} \equiv H_{y-} \cap H_z.$$

Since $y_t$ is deterministic, for fixed $k_2$, the residual of any element of

$H_{y+}$ after projection onto $H_{y-}$ is $O_p(1)$ as $k_1 \rightarrow \infty$. Any linear combination of elements of $H_{y-}$ which fulfill the square summability condition will therefore be $O_p(k_1^{-2})$ as well. The magnitude of $\text{var } (\xi(L)z_t)$ will equal the magnitude of $\frac{RSS_0 - RSS_1}{T}$, which means that the $F$ statistic is of magnitude

$$O_p(1) \frac{T-k}{k_2} \rightarrow 0$$
if \( k_1 \) is \( O_p(T^{1/2+\delta}) \) for some \( \delta < 0 \).

Notice that if \( k_2 \) is also a function of \( T \), the same result will hold provided that \( k_1 - k_2 \) is \( O_p(T^{1/2+\delta}) \) for some \( \delta < 0 \). This will hold, in turn, if \( \frac{k_1}{k_2} \) is \( O_p(T^{\delta}) \) for some \( \delta > 0 \). The proof is complete.

Note that \( \frac{k}{T} \rightarrow 0 \) is the standard condition for regressor coefficient consistency, as discussed in Huber (1982). Asymptotically infinite lag length choices which produce regressor coefficient consistency therefore exist which produce Granger noncausality.

Proof of Proposition 3.4

For the projection of \( y_t \) onto \( H_{x+}(t,k_1) \oplus H_{x-}(t,k_2) \), i.e.

\[ y_t = \alpha_k x_t + \beta_k (L)x_t + v_t \]

the reduction in residual variance which occurs from adding the \( x_{t+i} \) terms will be bounded from above by the value of the projection of \( \alpha_k (L^{-1})x_t \) onto \( H_{x-}(t,k_1) \). Using earlier arguments, for each \( k_1, k_2 \) the projection of the \( x_{t+i} \) variables will generate residuals \( \xi_{t+i} \) whose variances will be proportional to the sequence \( \frac{1}{(k_1 - k_2)^2} \ldots \frac{1}{k_1^2} \).

Finally, from the Cauchy-Schwartz inequality

\[ E[\Sigma_i \xi_{t+i}^2] < E(\Sigma_i^2 \xi_{t+i}^2) \]

For any \( \alpha > 1 \), the sum \( \Sigma_{t=1}^\infty \frac{1}{t^\alpha} \) converges to the Riemann zeta function \( \zeta(\alpha) \). \( \zeta(\alpha) - \sum_{t=1}^k t^{-\alpha} \) will be of order \( O(k^{-\alpha}) \). Therefore, if \( k_1(T) = k_2(T) \) are \( O(T^{1/2+\delta}) \) then \( \Sigma_{t+i}^2 \) is convergent and \( O_p(T^{-1-2\delta}) \). Square summability will bound the \( \Sigma_{t+i}^2 \) sequence. Following identical arguments as above, the \( F \) statistic will converge to zero.


