ARIMA and ARFIMA models

Christopher F Baum

EC 823: Applied Econometrics

Boston College, Spring 2013

ARIMA and ARMAX models

The pure ARIMA model is an atheoretic linear univariate time series model which expresses that series in terms of three sets of parameters:

$$A(L)(1 - L)^d y_t = \alpha + B(L)\varepsilon_t$$

The first set of p parameters define the autoregressive polynomial in the lag operator L:

$$A(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p$$

The second set of q parameters define the moving average polynomial in the $i.i.d.$ disturbance process:

$$B(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$$

The third parameter, *d* above, expresses the integer order of differencing to be applied to the series before estimation to render it stationary. Thus, we speak of an *ARIMA*(p, d, q) model, with $p + q$ parameters to be estimated.

In order to be estimable, the *d*-differenced time series must be stationary, so that the AR polynomial in the lag operator may be inverted. Let y^* be the differenced time series:

$$y_t^* = A(L)^{-1} \left(\alpha + B(L)\varepsilon_t \right)$$

where the stability condition requires that the characteristic roots of the $A(L)$ polynomial lie strictly outside the unit circle. For an AR(1), that requires that $|\rho| < 1$. If the stability condition is satisfied, then an ARMA(p,q) model will have a MA(∞) representation.

We have presented the model to be a univariate autoregression with a moving-average disturbance process. However, it can also be cast in terms of an autoregression in the disturbances. For instance, the ARIMA(1,0,1) can be written as

$$y_t = \alpha + \rho y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

which is equivalent to the structural equation and ARMA(1,1) disturbance process:

$$
\begin{aligned}
y_t &= \gamma + \mu_t \\
\mu_t &= \rho \mu_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t
\end{aligned}
$$

This latter specification is more general, in that we can write the structural equation, replacing γ with $X\beta$, which defines a linear regression model with $ARMA(p, q)$ errors. This framework is sometimes termed ARMA-X or ARMAX, and generalizes the model often applied to regression with $AR(1)$ errors (e.g., `prais` in Stata).

Estimation of ARIMA models is performed by maximum likelihood using the Kalman filter, as any model containing a moving average component requires nonlinear estimation techniques. Convergence can be problematic for models with a large q.

The default VCE for ARIMA estimates is the outer product of gradients (OPG) estimator devised by Berndt, Hall, Hall and Hausman (BHHH), which has been shown to be more numerically stable for recursive computations such as the Kalman filter.

Once a time series has been rendered stationary by differencing, the choice of *p* and *q* may be made by examining two time-domain constructs: the autocorrelation function (ACF) and the partial autcorrelation function (PACF).

Use of these functions requires that the estimated model is both stationary and *invertible*: that is, that the model may be transformed by premultiplying by the inverse of the B(L) polynomial, rendering it as a AR(∞). For that representation to exist, the characteristic roots of the B(L) polynomial must lie outside the unit circle. In a MA(1), this condition requires that $|\theta| < 1$.

The principle of *parsimony* recommends that a model with fewer parameters is to be preferred, and information criteria such as the AIC and BIC penalize less parsimonious specifications.

Following estimation of an ARIMA(p,d,q) model, you should check to see that residuals are serially uncorrelated, via their own ACF and PACF and the Ljung–Box–Pierce Q statistic (`wntestq`). It may also be useful to fit the model over a subset of the available data and examine how well it performs on the full data set.

As the object of ARIMA modeling is often forecasting, you may want to apply a forecast accuracy criterion to compare the quality of forecasts of competing models. Diebold and Mariano (*JBES*, 1995) developed a test for that purpose, relaxing some of the assumptions of the earlier Granger–Newbold (*JRSS-B*, 1976) test. That routine is available from SSC as `dmariano`. It allows you to compare two *ex post* forecasts in terms of mean squared error, mean absolute error, and mean absolute prediction error.

Stata's capabilities to estimate ARIMA or 'Box–Jenkins' models are implemented by the `arima` command. These modeling tools include both the traditional $ARIMA(p, d, q)$ framework as well as multiplicative seasonal ARIMA components for a univariate time series model. The `arima` command also implements ARMAX models: that is, regression equations with ARMA errors.

In both the ARIMA and ARMAX contexts, the `arima` command implements dynamic forecasts, where successive forecasts are based on their own predecessors, rather than being one-step-ahead (static) forecasts.

To illustrate, we fit an ARIMA(p,d,q) model to the US consumer price index (CPI):

```
. use usmacro1
. arima cpi, arima(1, 1, 1) nolog
ARIMA regression
Sample:  1959q2 – 2010q3                          Number of obs      =        206
                                                  Wald chi2(2)       =   12657.64
Log likelihood =  -105.364                        Prob > chi2        =     0.0000
```

D.cpi	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
cpi						
_cons	.4711825	.0508081	9.27	0.000	.3716004	.5707646
ARMA						
ar						
L1.	-.3478959	.0590356	-5.89	0.000	-.4636036	-.2321882
ma						
L1.	.9775208	.0123013	79.46	0.000	.9534106	1.001631
/sigma	.4011922	.008254	48.61	0.000	.3850146	.4173697

```
. estimates store e42a
```

In this example, we use the `arima(p, d, q)` option to specify the model. The `ar( )` and `ma( )` options may also be used separately, in which case a *numlist* of lags to be included is specified. Differencing is then applied to the dependent variable using the `D.` operator. For example:

```
. use usmacro1
. arima D.cpi, ar(1 4) nolog
ARIMA regression
Sample:  1959q2 - 2010q3                        Number of obs    =        206
                                                Wald chi2(2)     =     105.12
Log likelihood = -112.7938                      Prob > chi2      =     0.0000
```

D.cpi	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
cpi						
_cons	.4578741	.1086742	4.21	0.000	.2448766	.6708716
ARMA						
ar						
L1.	.3035501	.0686132	4.42	0.000	.1690707	.4380295
L4.	.3342019	.0407126	8.21	0.000	.2544068	.413997
/sigma	.4177019	.0071104	58.75	0.000	.4037658	.4316381

Several prediction options are available after estimating an `arima` model. The default option, `xb`, predicts the actual dependent variable: so if `D.cpi` is the dependent variable, predictions are made for that variable. In contrast, the `y` option generates predictions of the original variable, in this case `cpi`.

The `mse` option calculates the mean squared error of predictions, while `yresiduals` are computed in terms of the original variable.

We recall the estimates from the first model fitted, and calculate predictions for the actual dependent variable, ΔCPI:

```
. estimates restore e42a
(results e42a are active now)

. predict double dcpihat, xb

. tsline dcpihat, ///
>     ti("ARIMA(1,1,1) model of {&Delta}US CPI") scheme(s2mono)
```

ARIMA(1,1,1) model of ΔUS CPI
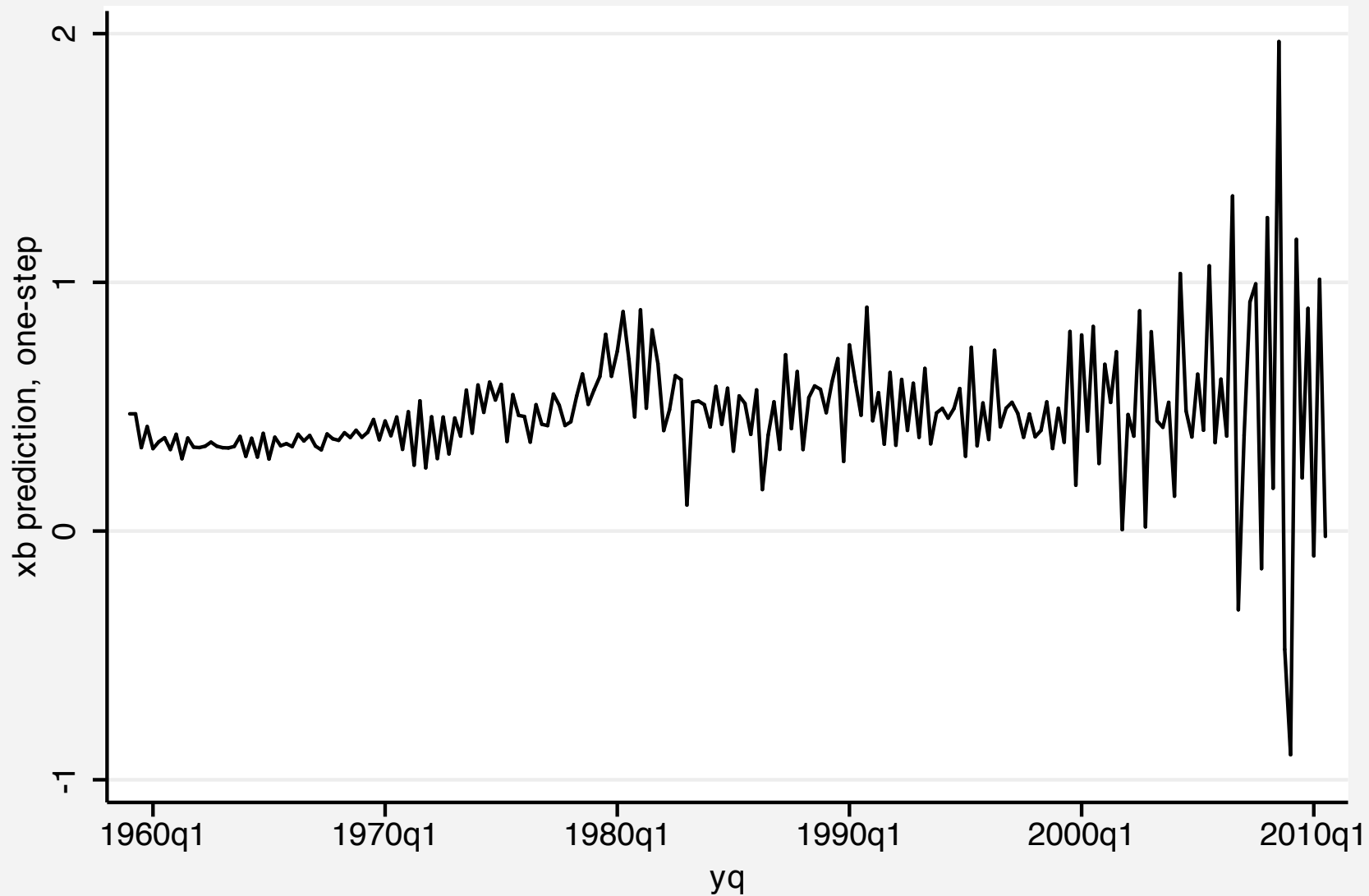
We can see that the predictions are becoming increasingly volatile in recent years.

We may also compute predicted values and residuals for the level of *CPI*:

```
. estimates restore e42a
(results e42a are active now)
. predict double cpihat, y
(1 missing value generated)
. predict double cpieps, yresiduals
(1 missing value generated)
. tw (tsline cpieps, yaxis(2)) (tsline cpihat), ///
>    ti("ARIMA(1,1,1) model of US CPI") scheme(s2mono)
```

ARIMA(1,1,1) model of US CPI
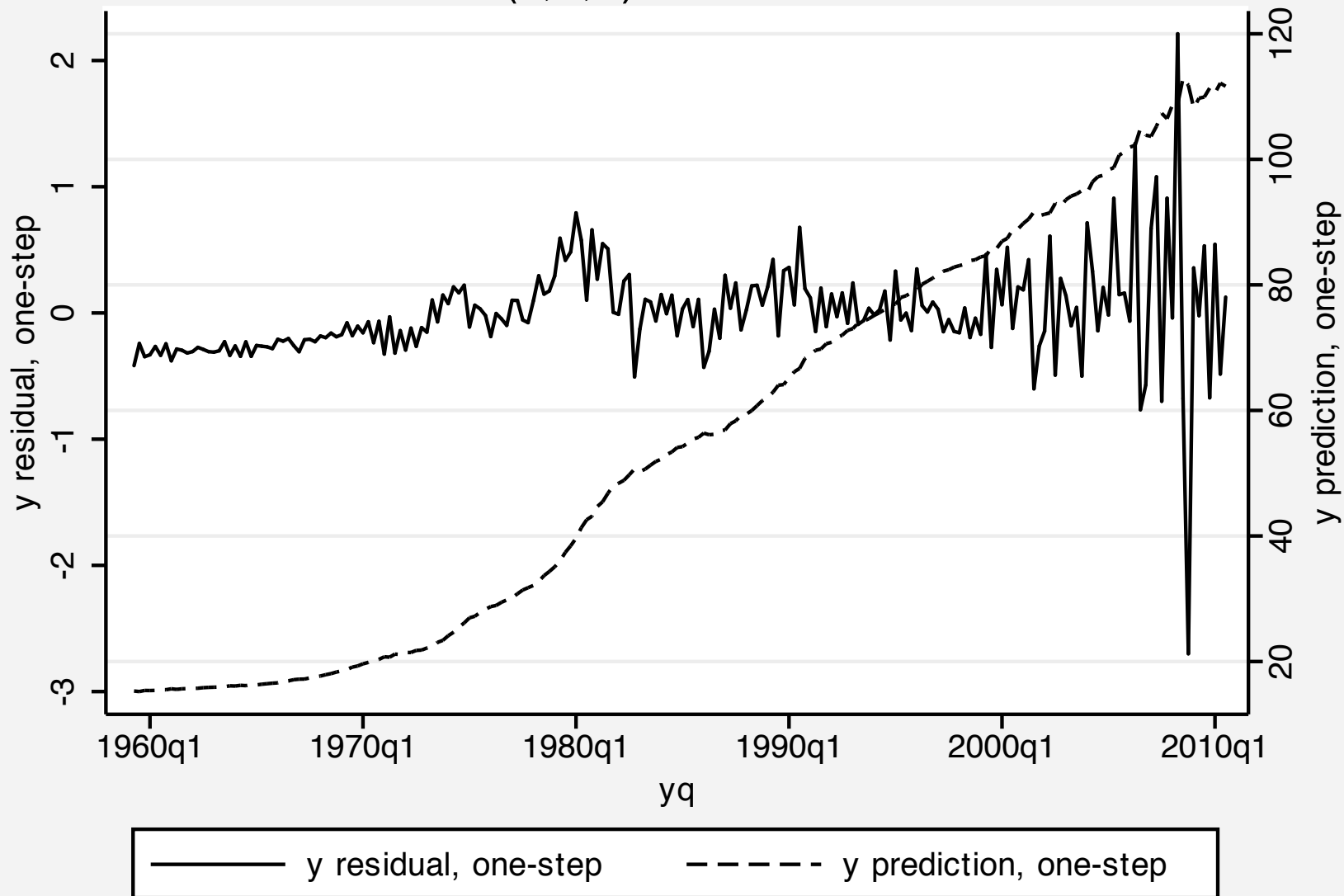
We now illustrate the estimation of an ARMAX model of Δcpi as a function of $\Delta oilprice$ with $ARMA(1, 1)$ errors. The estimation sample runs through 2008q4.

```
. arima d.cpi d.oilprice if tin(, 2008q4), ar(1) ma(1) nolog
ARIMA regression
Sample:  1959q2 - 2008q4                        Number of obs    =        199
                                                Wald chi2(3)     =    1829.64
Log likelihood = -27.08681                      Prob > chi2      =     0.0000
```

D.cpi	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf. Interval]	
cpi						
oilprice						
D1.	.0602003	.0021528	27.96	0.000	.0559808	.0644198
_cons	.4397912	.1833278	2.40	0.016	.0804753	.7991071
ARMA						
ar						
L1.	.9732011	.0296099	32.87	0.000	.9151667	1.031235
ma						
L1.	-.7867952	.0535747	-14.69	0.000	-.8917997	-.6817906
/sigma	.2765534	.0091383	30.26	0.000	.2586426	.2944642

We compute static (one-period-ahead) *ex ante* forecasts and dynamic (multi-period-ahead) *ex ante* forecasts for 2009q1–2010q3. In specifying the dynamic forecast, the `dynamic( )` option indicates the period in which references to *y* should first evaluate to the prediction of the model rather than historical values. In all prior periods, references to *y* are to the actual data.

```
. predict double cpihat_s if tin(2006q1,), y
(188 missing values generated)
. label var cpihat_s "static forecast"
. predict double cpihat_d if tin(2006q1,), dynamic(tq(2008q4)) y
(188 missing values generated)
. label var cpihat_d "dynamic forecast"
. tw (tsline cpihat_s cpihat_d if !mi(cpihat_s)) ///
>    (scatter cpi yq if !mi(cpihat_s), c(i)), scheme(s2mono) ///
>    ti("Static and dynamic ex ante forecasts of US CPI") ///
>    t2("Forecast horizon: 2009q1-2010q3") legend(rows(1))
```

Static and dynamic ex ante forecasts of US CPI
Forecast horizon: 2009q1-2010q3
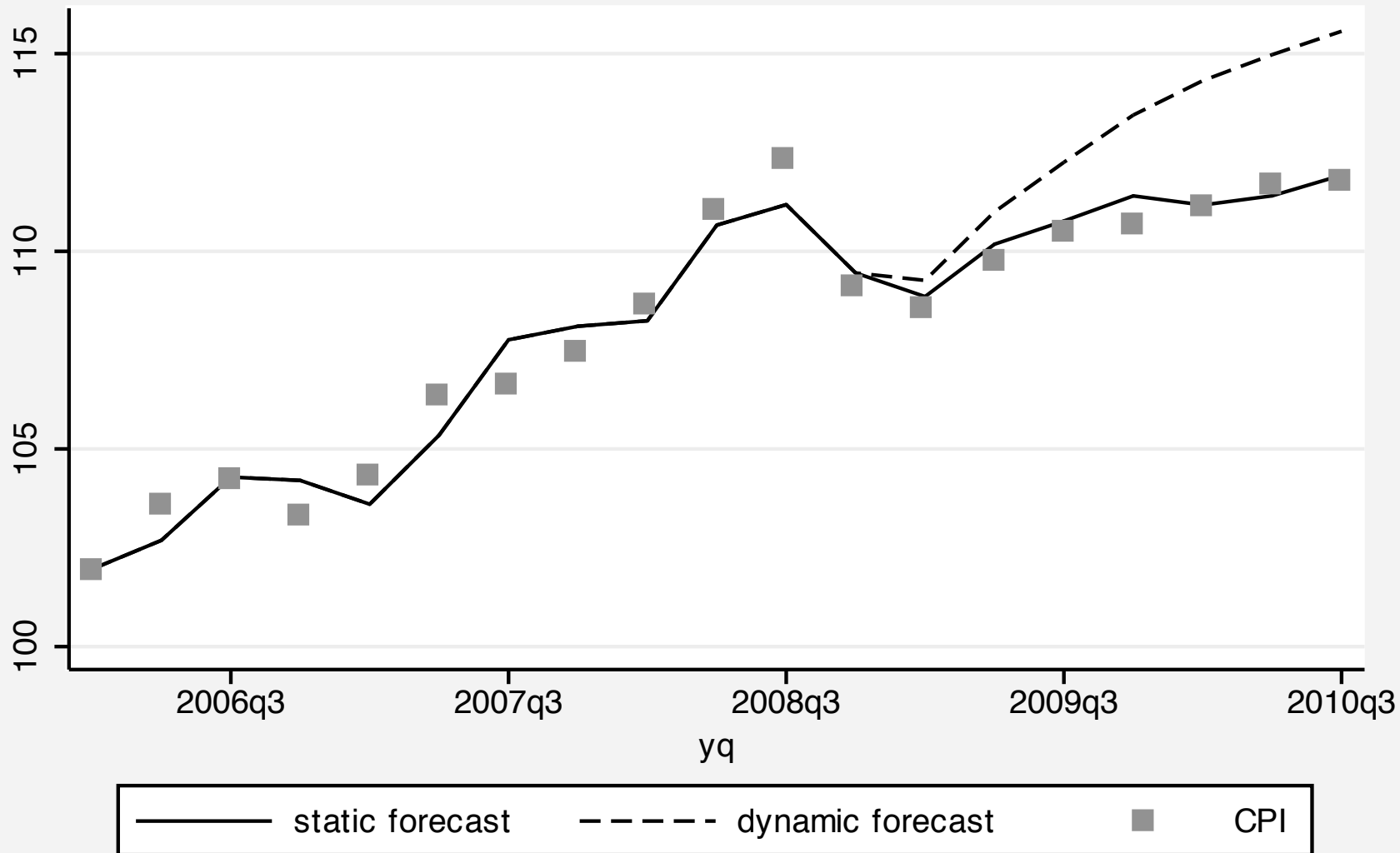
ARFIMA models

In estimating an ARIMA model, the researcher chooses the integer order of differencing d to ensure that the resulting series $(1 - L)^d y_t$ is a stationary process.

As unit root tests often lack the power to distinguish between a truly nonstationary ($I(1)$) series and a stationary series embodying a structural break or shift, time series are often first-differenced if they do not receive a clean bill of health from unit root testing.

Many time series exhibit too much long-range dependence to be classified as $I(0)$ but are not $I(1)$. The ARFIMA model is designed to represent these series.

This problem is exacerbated by reliance on Dickey–Fuller style tests, including the improved Elliott–Rothenberg–Stock (*Econometrica*, 1996, `dfgls`) test, which have $I(1)$ as the null hypothesis and $I(0)$ as the alternative. For that reason, it is a good idea to also employ a test with the alternative null hypothesis of stationarity ($I(0)$) such as the Kwiatkowski–Phillips–Schmidt–Shin (*J. Econometrics*, 1992, `kpss`) test to see if its verdict agrees with that of the Dickey–Fuller style test.

The KPSS test, with a null hypothesis of $I(0)$, is also useful in the context of the ARFIMA model we now consider. This model allows for the series to be *fractionally integrated*, generalizing the ARIMA model's integer order of integration to allow the d parameter to take on fractional values, $-0.5 < d < 0.5$.

The concept of fractional integration is often referred to as defining a time series with *long-range dependence*, or *long memory*. Any pure ARIMA stationary time series can be considered a *short memory* series. An *AR(p)* model has infinite memory, as all past values of ε_t are embedded in y_t, but the effect of past values of the disturbance process follows a geometric lag, damping off to near-zero values quickly. A *MA(q)* model has a memory of exactly *q* periods, so that the effect of the moving average component quickly dies off.

The ARFIMA model[1]

The model of an autoregressive fractionally integrated moving average process of a timeseries of order (p, d, q), denoted by ARFIMA (p, d, q), with mean μ, may be written using operator notation as

$$\Phi(L)(1 - L)^d (y_t - \mu) = \Theta(L)\epsilon_t, \ \epsilon_t \sim i.i.d.(0, \sigma_\epsilon^2)$$

where L is the backward-shift operator, $\Phi(L) = 1 - \phi_1 L - .. - \phi_p L^p$, $\Theta(L) = 1 + \vartheta_1 L + ... + \vartheta_q L^q$, and $(1 - L)^d$ is the fractional differencing operator defined by

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)L^k}{\Gamma(-d)\Gamma(k + 1)}$$

with $\Gamma(\cdot)$ denoting the gamma (generalized factorial) function. The parameter d is allowed to assume any real value.

---

[1]See Baum and Wiggins (*Stata Tech.Bull.*, 2000).

The arbitrary restriction of d to integer values gives rise to the standard autoregressive integrated moving average (ARIMA) model. The stochastic process y_t is both stationary and invertible if all roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle and $|d| < 0.5$. The process is nonstationary for $d \geq 0.5$, as it possesses infinite variance; see Granger and Joyeux (*JTSA*, 1980).

Assuming that $d \in [0, 0.5)$, Hosking (*Biometrika*, 1981) showed that the autocorrelation function, $\rho(\cdot)$, of an ARFIMA process is proportional to k^{2d-1} as $k \to \infty$. Consequently, the autocorrelations of the ARFIMA process decay *hyperbolically* to zero as $k \to \infty$ in contrast to the faster, geometric decay of a stationary ARMA process.

For $d \in (0, 0.5)$, $\sum_{j=-n}^{n} |\rho(j)|$ diverges as $n \to \infty$, and the ARFIMA process is said to exhibit long memory, or long-range positive dependence. The process is said to exhibit intermediate memory (anti-persistence), or long-range negative dependence, for $d \in (-0.5, 0)$.

The process exhibits short memory for $d = 0$, corresponding to stationary and invertible ARMA modeling. For $d \in [0.5, 1)$ the process is mean reverting, even though it is not covariance stationary, as there is no long-run impact of an innovation on future values of the process.

If a series exhibits long memory, it is neither stationary ($I(0)$) nor is it a unit root ($I(1)$) process; it is an $I(d)$ process, with d a real number.

A series exhibiting long memory, or persistence, has an autocorrelation function that damps hyperbolically, more slowly than the geometric damping exhibited by "short memory" (ARMA) processes. Thus, it may be predictable at long horizons. An excellent survey of long memory models—which originated in hydrology, and have been widely applied in economics and finance–is given by Baillie (*J. Econometrics*, 1996).

Approaches to estimation of the ARFIMA model

There are two approaches to the estimation of an ARFIMA (p, d, q) model: exact maximum likelihood estimation, as proposed by Sowell (1992), and semiparametric approaches. Sowell's approach requires specification of the p and q values, and estimation of the full ARFIMA model conditional on those choices. This involves the challenge of choosing an appropriate ARMA specification.

We first describe semiparametric methods, in which we assume that the "short memory" or ARMA components of the timeseries are relatively unimportant, so that the long memory parameter d may be estimated without fully specifying the data generating process.

The Lo Modified Rescaled Range estimator[2]

The Stata routine `lomodrs` performs Lo's (*Econometrica*, 1991) modified rescaled range (R/S, "range over standard deviation") test for long range dependence of a time series. The classical R/S statistic, devised by Hurst (1951) and Mandelbrot (*AESM*, 1972), is the range of the partial sums of deviations of a timeseries from its mean, rescaled by its standard deviation. For a sample of n values $\{x_1, x_2, \ldots x_n\}$,

$$Q_n = \frac{1}{s_n} \left[Max_{1 \leq k \leq n} \sum_{j=1}^{k} \left(x_j - \bar{x}_n\right) - Min_{1 \leq k \leq n} \sum_{j=1}^{k} \left(x_j - \bar{x}_n\right) \right]$$

where s_n is the maximum likelihood estimator of the standard deviation of x.

---

[2]See Baum and Röom (*Stata Tech. Bull.*, 2001).

The first bracketed term is the maximum of the partial sums of the first k deviations of x_j from the full-sample mean, which is nonnegative. The second bracketed term is the corresponding minimum, which is nonpositive. The difference of these two quantities is thus nonnegative, so that $Q_n > 0$. Empirical studies have demonstrated that the R/S statistic has the ability to detect long-range dependence in the data.

Like many other estimators of long-range dependence, though, the R/S statistic has been shown to be excessively sensitive to "short-range dependence," or short memory, features of the data. Lo (1991) shows that a sizable $AR(1)$ component in the data generating process will seriously bias the R/S statistic. He modifies the R/S statistic to account for the effect of short-range dependence by applying a "Newey–West" correction (using a Bartlett window) to derive a consistent estimate of the long-range variance of the timeseries.

For `maxlag` > 0, the denominator of the statistic is computed as the Newey–West estimate of the long run variance of the series. If `maxlag` is set to zero, the test performed is the classical Hurst–Mandelbrot rescaled-range statistic. Critical values for the test are taken from Lo, 1991, Table II.

Inference from the modified R/S test for long range dependence is complementary to that derived from that of other tests for long memory, or fractional integration in a timeseries, such as `kpss`, `gphudak`, `modlpr` and `roblpr`.

The Geweke–Porter-Hudak log periodogram regression (LPR) estimator

The Stata command `gphudak` (Baum and Wiggins, *Stata Tech. Bull.*, 2000) performs the Geweke and Porter-Hudak (*JTSA*, 1983) semiparametric log periodogram regression, often described as the "GPH test," for long memory (fractional integration) in a timeseries.

The GPH method uses nonparametric methods—a spectral regression estimator—to evaluate d without explicit specification of the "short memory" (ARMA) parameters of the series. The series is usually differenced so that the resulting d estimate will fall in the [-0.5, 0.5] interval.

Geweke and Porter-Hudak (1983) proposed a semiparametric procedure to obtain an estimate of the memory parameter d of a fractionally integrated process X_t in a model of the form

$$(1 - L)^d X_t = \epsilon_t,$$

where ϵ_t is stationary with zero mean and continuous spectral density $f_\epsilon(\lambda) > 0$.

The estimate $\hat{d}$ is obtained from the application of ordinary least squares to

$$\log\left(I_x\left(\lambda_s\right)\right) = \hat{c} - \hat{d}\log\left|1 - e^{i\lambda_s}\right|^2 + \textit{residual}$$

computed over the fundamental frequencies $\left\{\lambda_s = \frac{2\pi s}{n}, s = 1, ..., m, m < n\right\}$. We define $\omega_x\left(\lambda_s\right) = \frac{1}{\sqrt{2\pi n}}\sum_{t=1}^{n} X_t e^{it\lambda_s}$ as the discrete Fourier transform (dft) of the timeseries X_t, $I_x\left(\lambda_s\right) = \omega_x\left(\lambda_s\right)\omega_x\left(\lambda_s\right)^*$ as the periodogram, and $x_s = \log\left|1 - e^{i\lambda_s}\right|$. Ordinary least squares on (7) yields

$$\hat{d} = 0.5\frac{\sum_{s=1}^{m} x_s \log I_x\left(\lambda_s\right)}{\sum_{s=1}^{m} x_s^2}.$$

Various authors have proposed methods for the choice of *m*, the number of Fourier frequencies included in the regression. The regression slope estimate is an estimate of the slope of the series' power spectrum in the vicinity of the zero frequency; if too few ordinates are included, the slope is calculated from a small sample. If too many are included, medium and high-frequency components of the spectrum will contaminate the estimate. A choice of $\sqrt{T}$, or `power` = 0.5 is often employed.

To evaluate the robustness of the GPH estimate, a range of power values (from 0.40–0.75) is commonly calculated as well. Two estimates of the *d* coefficient's standard error are commonly employed: the regression standard error, giving rise to a standard *t*-test, and an asymptotic standard error, based upon the theoretical variance of the log periodogram of $\frac{\pi^2}{6}$. The statistic based upon that standard error has a standard normal distribution under the null.

The Phillips Modified GPH log periodogram regression estimator

The Stata routine `modlpr` (Baum and Wiggins, *Stata Tech. Bull.*, 2000) computes a modified form of the GPH estimate of the long memory parameter, d, of a timeseries, proposed by Phillips (Cowles, 1999a, 1999b). Phillips (1999a) points out that the prior literature on this semiparametric approach does not address the case of $d = 1$, or a unit root, in (6), despite the broad interest in determining whether a series exhibits unit-root behavior or long memory behavior, and his work showing that the $\hat{d}$ estimate of (7) is inconsistent when $d > 1$, with $\hat{d}$ exhibiting asymptotic bias toward unity.

This weakness of the GPH estimator is solved by Phillips' Modified Log Periodogram Regression estimator, in which the dependent variable is modified to reflect the distribution of d under the null hypothesis that $d = 1$. The estimator gives rise to a test statistic for $d = 1$ which is a standard normal variate under the null.

Phillips suggests that deterministic trends should be removed from the series before application of the estimator. Accordingly, the routine will automatically remove a linear trend from the series. This may be suppressed with the `notrend` option. The comments above regarding `power` apply equally to `modlpr`.

Phillips' (1999b) modification of the GPH estimator is based on an exact representation of the dft in the unit root case. The modification expresses

$$\omega_x(\lambda_s) = \frac{\omega_u(\lambda_s)}{1 - e^{i\lambda_s}} - \frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{X_n}{\sqrt{2\pi n}}$$

and the modified dft as

$$\upsilon_x(\lambda_s) = \omega_x(\lambda_s) + \frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{X_n}{\sqrt{2\pi n}}$$

with associated periodogram ordinates $I_v(\lambda_s) = \upsilon_x(\lambda_s)\upsilon_x(\lambda_s)^*$ (1999b, p.9). He notes that both $\upsilon_x(\lambda_s)$ and, thus, $I_v(\lambda_s)$ are observable functions of the data.

The log-periodogram regression is now the regression of $\log I_v(\lambda_s)$ on $a_s = \log\left|1 - e^{i\lambda_s}\right|$. Defining $\bar{a} = m^{-1}\sum_{s=1}^m a_s$ and $x_s = a_s - \bar{a}$, the modified estimate of the long-memory parameter becomes

$$\tilde{d} = 0.5\frac{\sum_{s=1}^m x_s \log I_\nu(\lambda_s)}{\sum_{s=1}^m x_s^2}.$$

Phillips proves that, with appropriate assumptions on the distribution of ϵ_t, the distribution of $\tilde{d}$ follows

$$\sqrt{m}\left(\tilde{d} - d\right) \rightarrow_d N\left(0, \frac{\pi^2}{24}\right),$$

so that $\tilde{d}$ has the same limiting distribution at $d = 1$ as does the GPH estimator in the stationary case so that $\tilde{d}$ is consistent for values of d around unity. A semiparametric test statistic for a unit root against a fractional alternative is then based upon the statistic (1999a, p.10):

$$z_d = \frac{\sqrt{m}\left(\tilde{d} - 1\right)}{\pi/\sqrt{24}}$$

with critical values from the standard normal distribution. This test is consistent against both $d < 1$ and $d > 1$ fractional alternatives.

Robinson's Log Periodogram Regression estimator

The Stata routine `roblpr` (Baum and Wiggins, *Stata Tech. Bull.*, 2000) computes the Robinson (*Ann. Stat.*,1995) multivariate semiparametric estimate of the long memory (fractional integration) parameters, $d(g)$, of a set of G timeseries, $y(g)$, $g = 1$, G with $G \geq 1$. When applied to a set of timeseries, the $d(g)$ parameter for each series is estimated from a single log-periodogram regression which allows the intercept and slope to differ for each series.

One of the innovations of Robinson's estimator is that it is not restricted to using a small fraction of the ordinates of the empirical periodogram of the series: that is, the reasonable values of `power` need not exclude a sizable fraction of the original sample size. The estimator also allows for the removal of one or more initial ordinates, and for the averaging of the periodogram over adjacent frequencies. The rationales for using non-default values of either of these options are presented in Robinson (1995).

Robinson (1995) proposes an alternative log-periodogram regression estimator which he claims provides "modestly superior asymptotic efficiency to $\bar{d}\,(0)$" ($\bar{d}\,(0)$ being the Geweke and Porter-Hudak estimator) (1995, p.1052).

Importantly, Robinson's formulation of the log-periodogram regression also allows for the formulation of a multivariate model, providing justification for tests that different time series share a common differencing parameter. Normality of the underlying time series is assumed, but Robinson claims that other conditions underlying his derivation are milder than those conjectured by GPH.

We present here Robinson's multivariate formulation, which applies to a single time series as well. Let X_t represent a $G-$dimensional vector with g^{th} element $X_{gt}, g = 1, ..., G$. Assume that X_t has a spectral density matrix $\int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) \, d\lambda$, with (g, h) element denoted as $f_{gh}(\lambda)$. The gth diagonal element, $f_{gg}(\lambda)$, is the power spectral density of X_{gt}. For $0 < C_g < \infty$ and $-\frac{1}{2} < d_g < \frac{1}{2}$, assume that $f_{gg}(\lambda) \sim C_g \lambda^{-2d_g}$ as $\lambda \to 0+$ for $g = 1, ..., G$. The periodogram of X_{gt} is then denoted as

$$I_g(\lambda) = (2\pi n)^{-1} \left| \sum_{t=1}^{n} X_{gt} e^{it\lambda} \right|^2, g = 1, ... G$$

Without averaging the periodogram over adjacent frequencies nor omission of l initial frequencies from the regression, we may define $Y_{gk} = \log I_g(\lambda_k)$. The least squares estimates of $c = (c_1, ...c_G)'$ and $d = (d_1, ..., d_G)'$ are given by

$$\begin{bmatrix} \tilde{c} \\ \tilde{d} \end{bmatrix} = \text{vec}\left\{ Y'Z(Z'Z)^{-1} \right\},$$

where $Z = (Z_1, ...Z_m)'$, $Z_k = (1, -2\log \lambda_k)'$, $Y = (Y_1, ...Y_G)$, and $Y_g = (Y_{g,1}, ...Y_{g,m})'$ for m periodogram ordinates.

Standard errors for $\tilde{d}_g$ and for a test of the restriction that two or more of the d_g are equal may be derived from the estimated covariance matrix of the least squares coefficients. The standard errors for the estimated parameters are derived from a pooled estimate of the variance in the multivariate case, so that their interval estimates differ from those of their univariate counterparts. Modifications to this derivation when the frequency-averaging (j) or omission of initial frequencies (l) options are selected may be found in Robinson (1995).

Maximum likelihood estimators of ARFIMA models

The official Stata command `arfima` implements the full maximum likelihood estimation of the ARFIMA(p,d,q) model, as proposed by Sowell (*J. Econometrics*, 1992). The ARFIMA model has the *d* parameter to handle long-run dependence and ARMA parameters to handle short-run dependence. Sowell has argued that using different parameters for different types of dependence facilitates estimation and interpretation.

The ARFIMA model specifies that

$$y_t = (1 - L)^{-d} \left(\Phi(L) \right)^{-1} \Theta(L) \varepsilon_t$$

After estimation, the short-run effects are obtained by setting $d = 0$, and describe the behavior of the fractionally differenced process $(1 - L)^d y_t$. The long-run effects use the estimated value of d, and describe the behavior of the fractionally integrated y_t.

Granger and Joyeux (1980) motivate ARFIMA models by noting that their implied spectral densities for $d > 0$ are finite except at frequency 0, whereas stationary ARMA models have finite spectral densities at all frequencies. The ARFIMA model is able to capture the long-range dependence, which cannot be expressed by stationary ARMA models.

lomodrs and classical rescaled range estimators

Data from Terence Mills' *Econometric Analysis of Financial Time Series* on returns from the annual S&P 500 index of stock prices, 1871-1997, are analyzed.

```
. use http://fmwww.bc.edu/ec-p/data/Mills2d/SP500A.DTA, clear
. lomodrs sp500ar

Lo Modified R/S test for sp500ar

Critical values for H0: sp500ar is not long-range dependent

90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:      .781   (1 lags via Andrews criterion)  N = 124
```

```
. lomodrs sp500ar, max(0)

Hurst-Mandelbrot Classical R/S test for sp500ar

Critical values for H0: sp500ar is not long-range dependent

90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:        .799     N = 124
. lomodrs sp500ar if tin(1946,)

Lo Modified R/S test for sp500ar

Critical values for H0: sp500ar is not long-range dependent

90%: [ 0.861, 1.747 ]
95%: [ 0.809, 1.862 ]
99%: [ 0.721, 2.098 ]

Test statistic:      1.29  (0 lags via Andrews criterion)  N = 50
```

For the full sample, the null of stationarity may be rejected at 95% using either the Lo modified R/S statistic or the classic Hurst–Mandelbrot statistic. For the postwar data, the null may not be rejected at any level of significance. Long-range dependence, if present in this series, seems to be contributed by pre-World War II behavior of the stock price series.

GPH, Phillips modlpr, Robinson roblpr

Data from Terence Mills' *Econometric Analysis of Financial Time Series* on UK FTA All Share stock returns (ftaret) and dividends (ftadiv) are analyzed.

```
. use http://fmwww.bc.edu/ec-p/data/Mills2d/FTA.DTA, clear
. gphudak ftaret,power(0.5 0.6 0.7)
GPH estimate of fractional differencing parameter
------------------------------------------------------------------------
                                          Asy.
Power   Ords     Est d    StdErr   t(H0: d=0)  P>|t|    StdErr   z(H0: d=0)  P>|z|
------------------------------------------------------------------------
   .5     20    -.00204    .1603     -0.0127   0.990    .1875     -0.0109   0.991
   .6     35    .228244    .1459      1.5645   0.128    .1302      1.7529   0.080
   .7     64    .141861    .08992     1.5776   0.120    .09127     1.5544   0.120
------------------------------------------------------------------------
```

```
. modlpr ftaret, power(0.5 0.55:0.8)
Modified LPR estimate of fractional differencing parameter for ftaret
-------------------------------------------------------------------------------
Power    Ords      Est d     Std Err  t(H0: d=0)    P>|t|     z(H0: d=1)    P>|z|
-------------------------------------------------------------------------------
   .5      19    .0231191    .139872     0.1653      0.870      -6.6401      0.000
  .55      25    .2519889   .1629533     1.5464      0.135      -5.8322      0.000
   .6      34    .2450011   .1359888     1.8016      0.080      -6.8650      0.000
  .65      46    .1024504   .1071614     0.9560      0.344      -9.4928      0.000
   .7      63    .1601207   .0854082     1.8748      0.065     -10.3954      0.000
  .75      84    .1749659     .08113     2.1566      0.034     -11.7915      0.000
   .8     113    .0969439   .0676039     1.4340      0.154     -14.9696      0.000
-------------------------------------------------------------------------------


. roblpr ftaret
Robinson estimates of fractional differencing parameter for ftaret
--------------------------------------------------------------
Power    Ords       Est d     Std Err   t(H0: d=0)     P>|t|
--------------------------------------------------------------
   .9     205    .1253645    .0446745      2.8062      0.005
--------------------------------------------------------------
```

```
. roblpr ftap ftadiv
Robinson estimates of fractional differencing parameters
Power =   .9                                    Ords      =     205
------------------------------------------------------------------
Variable          |      Est d       Std Err       t       P>|t|
------------------+-----------------------------------------------
ftap              |   .8698092      .0163302  53.2640      0.000
ftadiv            |   .8717427      .0163302  53.3824      0.000
------------------------------------------------------------------
Test for equality of d coefficients:   F(1,406) =  .00701   Prob > F = 0.9333



. constraint define 1 ftap=ftadiv

. roblpr ftap ftadiv ftaret, c(1)
Robinson estimates of fractional differencing parameters
Power =   .9                                    Ords      =     205
------------------------------------------------------------------
Variable          |      Est d       Std Err       t       P>|t|
------------------+-----------------------------------------------
ftap              |   .8707759      .0205143  42.4473      0.000
ftadiv            |   .8707759      .0205143  42.4473      0.000
ftaret            |   .1253645      .0290116   4.3212      0.000
------------------------------------------------------------------
Test for equality of d coefficients:   F(1,610) =  440.11   Prob > F = 0.0000
```

The GPH test, applied to the stock returns series, generates estimates of the long memory parameter that cannot reject the null at the ten percent level using the t-test.

Phillips' modified LPR, applied to this series, finds that $d = 1$ can be rejected for all powers tested, while $d = 0$ (stationarity) may be rejected at the ten percent level for powers 0.6, 0.7, and 0.75. Robinson's estimate for the returns series alone is quite precise.

Robinson's multivariate test, applied to the price and dividends series, finds that each series has $d > 0$. The test that they share the same d cannot be rejected. Accordingly, the test is applied to all three series subject to the constraint that price and dividends series have a common d, yielding a more precise estimate of the difference in d parameters between those series and the stock returns series.

Sowell MLE ARFIMA

We model the log of the monthly level of *CO* above Mauna Loa, Hawaii, removing seasonal effects by using the twelfth seasonal difference (`S12.` in Stata parlance) of that series. We first consider an ARMA model with a first lag in the AR polynomial and the second lag in the MA polynomial.

```
. webuse mloa

. arima S12.log, ar(1) ma(2) vsquish nolog

ARIMA regression

Sample:  1960m1 – 1990m12                        Number of obs      =        372
                                                 Wald chi2(2)       =     500.41
Log likelihood =  2001.564                       Prob > chi2        =     0.0000
```

		OPG					
S12.log	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
log						
_cons	.0036754	.0002475	14.85	0.000	.0031903	.0041605
ARMA						
ar						
L1.	.7354346	.0357715	20.56	0.000	.6653237	.8055456
ma						
L2.	.1353086	.0513156	2.64	0.008	.0347319	.2358853
/sigma	.0011129	.0000401	27.77	0.000	.0010344	.0011914

```
Note: The test of the variance against zero is one sided, and the two-sided
    confidence interval is truncated at zero.

. psdensity d_arma omega1
```

All parameters are statistically significant, and indicate a high degree of dependence. This model is nested within the ARFIMA model:

```
. arfima S12.log, ar(1) ma(2) vsquish nolog
ARFIMA regression
Sample: 1960m1 - 1990m12                        Number of obs   =        372
                                                Wald chi2(3)    =     248.87
Log likelihood =  2006.0805                     Prob > chi2     =     0.0000
```

S12.log	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf. Interval]	
S12.log						
_cons	.003616	.0012968	2.79	0.005	.0010743	.0061578
ARFIMA						
ar						
L1.	.2160894	.1015596	2.13	0.033	.0170362	.4151426
ma						
L2.	.1633916	.051691	3.16	0.002	.062079	.2647041
d	.4042573	.080546	5.02	0.000	.2463899	.5621246
/sigma2	1.20e-06	8.84e-08	13.63	0.000	1.03e-06	1.38e-06

```
Note: The test of the variance against zero is one sided, and the two-sided
   confidence interval is truncated at zero.
```

Here, too, all parameters are significant at the five percent level. The estimate of d, 0.404, is far from zero, indicating the presence of long-range dependence. We can compare the models' ability to capture the dynamics of the series by computing their implied spectral densities over $(0, \pi)$.

For a stationary time series, the spectral density describes the relative importance of components at different frequencies. The integral of the spectral density over $(-\pi, \pi)$ is the variance of the time series.

We can also compute the implied spectral density of the ARFIMA model, setting d to zero to compute the short-run estimates. The long-run estimates have infinite density at frequency zero.

All parameters are statistically significant, and indicate a high degree of dependence. This model is nested within the ARFIMA model:

```
. psdensity d_arfima omega2
. psdensity ds_arfima omega3, smemory
. line d_arma d_arfima omega1, name(lmem) scheme(s2mono) nodraw ylab(,angle(0))
. line d_arma ds_arfima omega1, name(smem) scheme(s2mono) nodraw ylab(,angle(0)
> )
. graph combine lmem smem, cols(1) xcommon  ///
> ti("ARMA and ARFIMA implied spectral densities")
. gr export 82308b.pdf, replace
(file /Users/cfbaum/Dropbox/baum/EC823 S2013/82308b.pdf written in PDF format)
```

ARMA and ARFIMA implied spectral densities
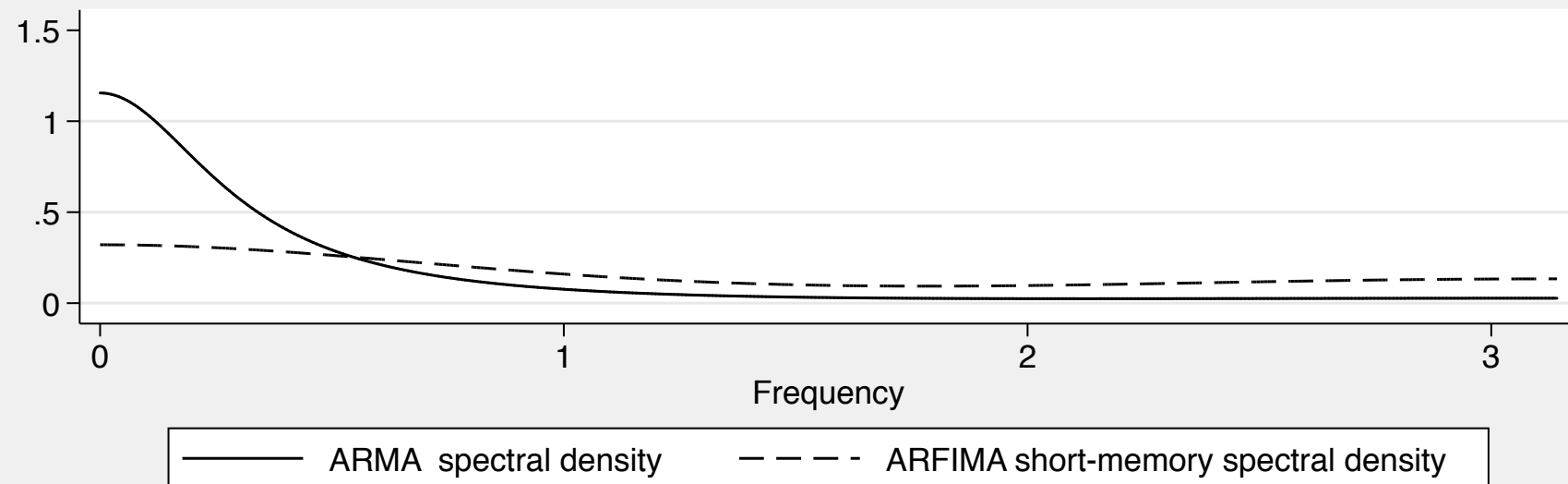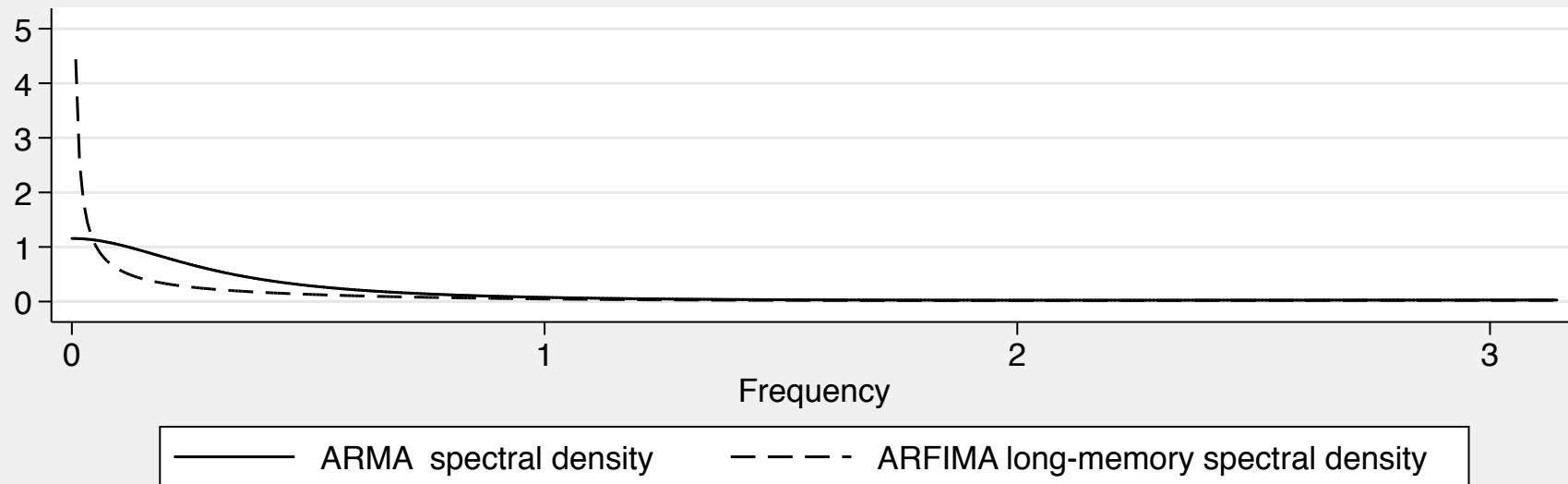
The two models imply different spectral densities for frequencies close to zero when $d > 0$. The spectral density of the ARMA model remains finite, but is pulled upward by the model's attempt to capture long-range dependence. The short-run ARFIMA parameters can capture both low-frequency and high-frequency components in the spectral density.

In contrast, the ARMA model confounds the long-run and short-run effects. The added flexibility of the ARFIMA model, with a separate parameter to capture long-run dependence, is evident in these estimates.

Although we have not illustrated it here, `arfima` may also fit 'ARFIMA-X' models with additional exogenous regressors.