THE JOURNAL OF ECONOMIC EDUCATION
2017, VOL. 48, NO. 3, 137–145
https://doi.org/10.1080/00220485.2017.1320610

Routledge
Taylor & Francis Group

CONTENT ARTICLES IN ECONOMICS

Time series econometrics for the 21st century

Bruce E. Hansen

Department of Economics, University of Wisconsin, Madison, WI, USA

ABSTRACT
What topics should be taught to undergraduate students in econometric time series?

KEYWORDS
Econometrics; regression; time series

JEL CODES
C10; C22

The field of econometrics largely started with time series analysis because many early datasets were time-series macroeconomic data. As the field developed, more cross-sectional and longitudinal datasets were collected, which today dominate the majority of academic empirical research. In nonacademic (private sector, central bank, and governmental) applications, time-series data is still important. Consequently, our undergraduate courses must still teach time-series econometric tools if we are to best serve our students. In this article, I review the key models that I believe are most critical for our students.

The core econometrics textbooks are Wooldridge (2013) and Stock and Watson (2015), both of which provide excellent treatments of econometric time series. My preference is the treatment in Stock and Watson, as it runs closer to the way I teach the material.

Two excellent texts on time-series econometrics and forecasting are Diebold (2015) and Gonzáles-Rivera (2012), which are appropriate for dedicated courses on economic forecasting. An excellent graduate-level forecasting textbook is Elliott and Timmermann (2016). The classic textbook for graduate-level time-series remains Hamilton (1994).

The data and Stata code used for the empirical results reported in this article are posted on the author's Web site: http://www.ssc.wisc.edu/~bhansen/.

Overview

Most undergraduate economic majors do not pursue PhD's in economics. Many go on to professional schools, some to government jobs, and others to the private sector. In many of these positions, it is quite common for our graduates to be exposed to economic data and analysis, including formal econometric (e.g., regression) analysis. Many of these applications are time series in nature. What tools can we give these students to help them succeed?

The core models that undergraduate students should learn are the Autoregressive (AR) model:

$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t, \tag{1}$$

the regression model:

$$y_t = \alpha + \delta_0 x_t + e_t, \tag{2}$$

CONTACT Bruce E. Hansen bruce.hansen@wisc.edu Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/vece.
Prepared for the 2017 AEA session on "Updating the Undergraduate Econometrics Curriculum." Research supported by the National Science Foundation.

the Distributed Lag (DL) model:

$$y_t = \alpha + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t, \tag{3}$$

and the Autoregressive-Distributed Lag (ADL) model:

$$y_t = \alpha + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + e_t. \tag{4}$$

In these equations, p is the number of lags of the dependent variable y_t, q is the number of lags of the explanatory variable x_t, and e_t is a mean-zero shock. In the ADL model, the contemporaneous regressor x_t is often omitted in contexts such as prediction.

With these core models, most concepts of single-equation time-series econometrics can be explained and taught. All are special cases of the ADL model, but it is useful to study them separately. These models allow us to teach the following core insights:

(1) The coefficients can be estimated by OLS just like a conventional regression.
(2) The appropriate standard error depends on whether or not the dynamics have been explicitly modeled. In AR and ADL models, the "robust" standard errors (e.g., the "r" option in Stata) are appropriate as long as the number of lags p is sufficiently large so that the errors are serially uncorrelated. In the regression and DL models, the equation error e_t will be serially correlated, so we should use HAC (heteroskedasticity-and-autocorrelation) standard errors, e.g., the "Newey" command in Stata.
(3) The AR model is useful to understand the serial correlation properties of the series y_t.
(4) The coefficients in the DL and ADL can be interpreted as multipliers. The coefficient δ_0 is the impact multiplier. The ratio $(\delta_0 + \cdots + \delta_q)/(1 - \phi_1 - \cdots - \phi_q)$ is the long-run multiplier.
(5) These multipliers have structural interpretations when the explanatory variable x_t is strictly exogeneous. This is a rather special situation.
(6) When the structural interpretation is taken, the ADL model can be used for counter-factual and policy analysis.
(7) The number of lags (p and q) in AR and ADL models may be selected by comparing models using the Akaike Information Criterion (AIC). Test statistics (t- and F-statistics) should not be used for model selection.
(8) In the regression and DL models, we should be greatly concerned about the possibility of a *spurious regression*: the setting where a regression with two unrelated time series has misleadingly large conventional t-statistics and R^2 values. Understanding the potential for spurious regression may be one of the greatest practical gifts we can teach our students.
(9) The parameters of time series models are likely to have changed over time. This requires care and attention.
(10) The ADL model without the contemporaneous regressor x_t is useful for prediction. The concept of Granger causality is taught within this model, namely that x_t does not Granger-cause y_t if $\delta_1 = \cdots = \delta_q = 0$.
(11) The ADL model without the contemporaneous regressor x_t may be used to produce one-step-ahead point forecasts for y_{n+1}.
(12) Point forecasts should be combined with interval forecasts, as the latter convey the degree of uncertainty about future outcomes. Teaching students to appreciate forecast uncertainty and interval forecasts is the most important lesson of prediction.
(13) To produce multi-step (h step) forecasts using a ADL model, we can use the multi-step version

$$y_{t+h} = \alpha + \phi_1 y_t + \cdots + \phi_p y_{t-p+1} + \delta_1 x_t + \cdots + \delta_q x_{t-q+1} + e_t,$$

which can also be estimated by OLS.
(14) Multi-step point forecasts should also be accompanied by interval forecasts, and together can be combined into fan charts.

Other time-series issues that can be usefully discussed in an undergraduate course include the following:

(1) Trends (deterministic and stochastic),
(2) Unit Roots,
(3) Seasonality,
(4) Cointegration,
(5) ARCH,
(6) Vector autoregressions, and
(7) Out-of-sample and split-sample analysis.

Some traditional topics that can be safely omitted include the following:

(1) Durbin Watson statistic,
(2) Statistical properties under classical assumptions,
(3) GLS estimation, and
(4) Cochrane-Orcutt estimation.

Standard errors and *t*-statistics

A critical issue in time series (and econometrics in general) is which standard error to use. There are three popular standard error formulae for applied econometric time series: classical, heteroskedasticity-robust, and HAC.

So-called classical (or homoskedastic) standard errors may be useful pedagogically due to their simple form, but are not used in current applied econometric practice. Consequently, courses should quickly move beyond classical standard errors.

Robust standard errors (the "r" option in Stata) are the most commonly used in current applied practice. They are appropriate for time-series models that are dynamically well-modeled, including autoregressive and ADL models.

HAC standard errors (Newey-West, the "Newey" command in Stata) are appropriate for simple (non-dynamic) regressions and DL models. They should be used whenever the serial correlation in the error has not been modeled.

The appropriate way to report regression estimates is coefficient estimates plus standard errors, as the latter can be used to construct confidence intervals or test statistics. Inappropriate reporting methods include coefficient estimates plus *t*-ratios, as this emphasizes testing (which is not typically the goal of estimation), and coefficient estimates plus asterisks, as this emphasizes statistical significance rather than magnitudes. It is always better for students to focus on parameter estimates, their magnitudes, and interpretations, rather than simple statements about significance.

To illustrate some of these ideas, consider a simple regression of retail gasoline prices on crude oil prices. Let gas_t denote the weekly percentage change in U.S. retail gasoline prices, and let oil_t denote the weekly percentage change in the Brendt European spot price for crude oil, for 1991–2016. The estimates are:

$$\begin{array}{rccc} gas_t = & 0.029 & + & 0.269\ oil_t + \widehat{e}_t \\ & (0.046) & & (0.011) \\ & (0.046) & & (0.015) \\ & (0.073) & & (0.021) \end{array}$$

Here, we report the coefficient estimates plus three standard errors. The first set of standard errors are classical (homoskedastic), the second are heteroskedasticity-robust, and the third are Newey-West estimates with 12 lags. We can see that the choice of standard error formula matters greatly, with the Newey-West roughly twice the magnitude of the classical. Because this is a static regression, the Newey-West are the appropriate choice.

Furthermore, it is appropriate to report standard errors rather than *t*-statistics as the former convey the most important information about the regression. In the above regression, we learn that a 1 percent change in crude oil prices leads to a contemporaneous (within one week) 0.27 percent change in retail gasoline prices. This is an immediate but partial pass-through. The Newey-West standard error shows

that the degree of pass-through is relatively precisely estimated (a 95 percent confidence interval is 23 percent to 31 percent). In contrast, the t-statistic for the coefficient is 13, which simply tells us that the true coefficient is nonzero. This is somewhat informative, but much less so than the discussion about magnitudes.

Autoregressive models

A good illustration of an autoregressive model is GDP growth. Let GDP_t denote quarterly real U.S. GDP percentage growth at annual rates, a series that is available for 1947Q2 through 2016Q3. An AR(3) is:

$$\underset{}{\mathrm{GDP}_t} = \underset{(0.32)}{1.93} + \underset{(0.06)}{0.34}\ \mathrm{GDP}_{t-1} + \underset{(0.06)}{0.13}\ \mathrm{GDP}_{t-2} - \underset{(0.06)}{0.09}\ \mathrm{GDP}_{t-3} + \widehat{e}_t$$

Here, we report heteroskedasticity-robust standard errors because this is a dynamic model.

The estimates show that GDP growth is positively autocorrelated, but mildly. This means that higher-than-average growth is followed in subsequent quarters by higher-than-average growth, but with quick mean reversion.

Another interesting example is stock price returns. Let $return_t$ denote the weekly percentage change in the S&P 500 index for 1950–2016. (For simplicity, we ignore dividends. Also, while daily observations are available, they have extra complications so it is easier to focus on weekly observations.) The estimates are:

$$\mathrm{return}_t = \underset{(0.04)}{0.16} - \underset{(0.029)}{0.032}\ \mathrm{return}_{t-1} + \underset{(0.025)}{0.037}\ \mathrm{return}_{t-2} + \widehat{e}_t$$

A simple form of the efficient market hypothesis suggests that stock returns are unpredictable, and thus the AR coefficients should be zero. Thus, the F-statistic for the AR coefficients is a simple test of efficient markets. In this example, the p value for the F-statistic is .12, so we fail to reject the efficient market hypothesis.

To repeat our message about the importance of using robust standard errors, if instead we had used the "old-fashioned" (homoskedastic) formula, the standard errors on the AR coefficients would be 0.17, the second lag would have a t-statistic of 2.2, and the F-statistic for the two AR coefficients would have a p value of .01, which would incorrectly suggest rejection of the efficient market hypothesis. Indeed, using the correct standard error formula makes a huge difference and alters inference.

Distributed lag models

Distributed lag models are useful when we want to estimate the impact of one variable upon another. As an example, consider the effect of crude oil prices upon retail gasoline prices, using the data from the earlier section on standard errors and t-statistics. A distributed lag model with a contemporaneous effect and six lags takes the form:

$$\begin{aligned}\mathrm{gas}_t = {} & \underset{(0.057)}{-0.009} + \underset{(0.016)}{0.243}\ \mathrm{oil}_t + \underset{(0.012)}{0.112}\ \mathrm{oil}_{t-1} + \underset{(0.011)}{0.063}\ \mathrm{oil}_{t-2} \\ & + \underset{(0.013)}{0.064}\ \mathrm{oil}_{t-3} + \underset{(0.010)}{0.030}\ \mathrm{oil}_{t-4} + \underset{(0.011)}{0.032}\ \mathrm{oil}_{t-5} + \underset{(0.012)}{0.018}\ \mathrm{oil}_{t-6} + \widehat{e}_t\end{aligned}$$

Here, the standard errors are computed using the Newey-West formula with 12 lags.

Under the assumption of strict exogeneity, the coefficients of a DL model are the effects of the regressor on the dependent variable. In this case, we see that a 1 percent change in crude oil prices leads to a

contemporaneous (within one week) change in retail gasoline prices of 0.24 percent, followed by further increases over the following six weeks.

The following equivalent regression can be used to estimate the cumulative multipliers:

$$\begin{aligned}\text{gas}_t = &\underset{(0.057)}{-0.009} + \underset{(0.016)}{0.243}\ \Delta\text{oil}_t + \underset{(0.024)}{0.355}\ \Delta\text{oil}_{t-1} + \underset{(0.028)}{0.418}\ \Delta\text{oil}_{t-2} \\ &+ \underset{(0.037)}{0.482}\ \Delta\text{oil}_{t-3} + \underset{(0.040)}{0.512}\ \Delta\text{oil}_{t-4} + \underset{(0.045)}{0.544}\ \Delta\text{oil}_{t-5} + \underset{(0.047)}{0.562}\ \text{oil}_{t-6} + \widehat{e}_t\end{aligned}$$

These results show that a 1 percent change in crude oil prices lead to a 0.56 percent cumulative change in retail gasoline prices after six weeks, with most of the change incorporated within the first four weeks. These estimates show how quickly changes in crude oil prices translate into retail prices.

Autoregressive distributed lag models

By combining the autoregressive and distributed lag models, we obtain a dynamically sound model useful for both policy analysis and forecasting. As an example, consider a short-run quarterly Phillips curve that models the change in U.S. CPI inflation as a function of the U.S. unemployment rate:

$$\begin{aligned}\Delta\text{Inf}_t = &\underset{(0.42)}{0.44} - \underset{(0.11)}{0.34}\ \Delta\text{Inf}_{t-1} - \underset{(0.09)}{0.39}\ \Delta\text{Inf}_{t-2} - \underset{(0.11)}{0.02}\ \Delta\text{Inf}_{t-3} - \underset{(0.07)}{0.17}\ \Delta\text{Inf}_{t-4} \\ &- \underset{(0.56)}{1.53}\ \text{Unemp}_{t-1} + \underset{(1.06)}{1.58}\ \text{Unemp}_{t-2} + \underset{(1.03)}{0.11}\ \text{Unemp}_{t-3} - \underset{(0.47)}{0.23}\ \text{Unemp}_{t-4} + \widehat{e}_t\end{aligned}$$

We estimate this equation without the contemporaneous unemployment rate so that it can be used as a forecasting equation.

The coefficients on lagged changes in inflation are large and negative, indicating that quarterly changes in inflation tend to be followed by a reversal over the next 6 months. The impact effect of the unemployment rate is negative, indicating that when the unemployment rate is higher than normal, the inflation rate tends to fall. Conversely, when the unemployment rate is lower than normal, the inflation rate tends to rise. This is the classic Phillips curve relationship. The sum of all four coefficients, however, is close to zero, indicating that the long-run impact of the unemployment rate on inflation rate changes is zero. This means that when the unemployment rate has been higher than normal and roughly constant for multiple periods (or lower than normal and roughly constant), then there is no effect on the inflation rate.

To assess if the overall effect is statistically significant, we can perform a joint statistical test on all four lags of the unemployment rate. This F-statistic has a p value of .03, so it is marginally significant. This is known as a Granger Causality test, and the evidence here suggests that the U.S. unemployment rate "Granger causes" U.S. inflation rates. This means that the unemployment rate helps to forecast inflation, not that it is causal in a strict sense.

Model selection

Economic theory does not provide guidance regarding the number of lags in an AR, DL, or ADL model. Rather, it is a matter of statistical fit. The more that lags are included, the smaller the bias is, yet the larger the estimation variance is. Lag selection is inherently a bias-variance tradeoff.

Consequently, statistical tests (Durbin-Watson, t- and F-tests) are not appropriate for lag selection. There is no economic hypothesis to test. There is no null hypothesis. Testing is the incorrect lens through which to view the model.

Instead, information criteria are appropriate. For undergraduate courses, the AIC is a good criterion as it is simple and readily available. AR and ADL models can be easily compared via the AIC. This is a simple and effective method to select the number of lags in a dynamic model.

As one example, take the GDP autoregressive model presented previously. Comparing AR(0) through AR(6) models, the AIC criterion selects the AR(3) model as presented.

As a second example, take the Phillips curve example from the section on autoregressive distributed lag models. Comparing models with one through six lags of the change in the inflation rate and one through four lags of the unemployment rate, the AIC selects the model with four lags of the change in the inflation rate and two lags of the unemployment rate. We will use this selected model for a forecast application in the final section.

Spurious regression

One of the dangers of estimating simple regressions or distributed lag models is that conventional standard errors can be off by several orders of magnitude. This danger is routinely ignored in nonacademic (e.g., journalistic) writing, and also can be seen in some academic (e.g., macroeconomic) articles. This is a serious practical issue facing economists working in industry.

To illustrate, figure 1 displays a plot of two annual time series, labeled y_1 and y_2, for the period 1906–2015. The two appear to track each other fairly well. Their sample correlation is 0.73. Estimation of a linear regression of y_{1t} on y_{2t} (with conventional "robust" standard errors) is reported below. The coefficient for y_{2t} is highly significant (the t-statistic is about 13), and $R^2 = 0.54$. Clearly the relationship seems strong.

$$\underset{}{y_{1t}} = \underset{(0.52)}{-2.95} + \underset{(0.07)}{0.95}\ y_{2t} + \widehat{e}_t$$

The truth is that relationship is completely spurious. The two time series were generated using a random number generator, and are statistically independent. The two series are completely unrelated,

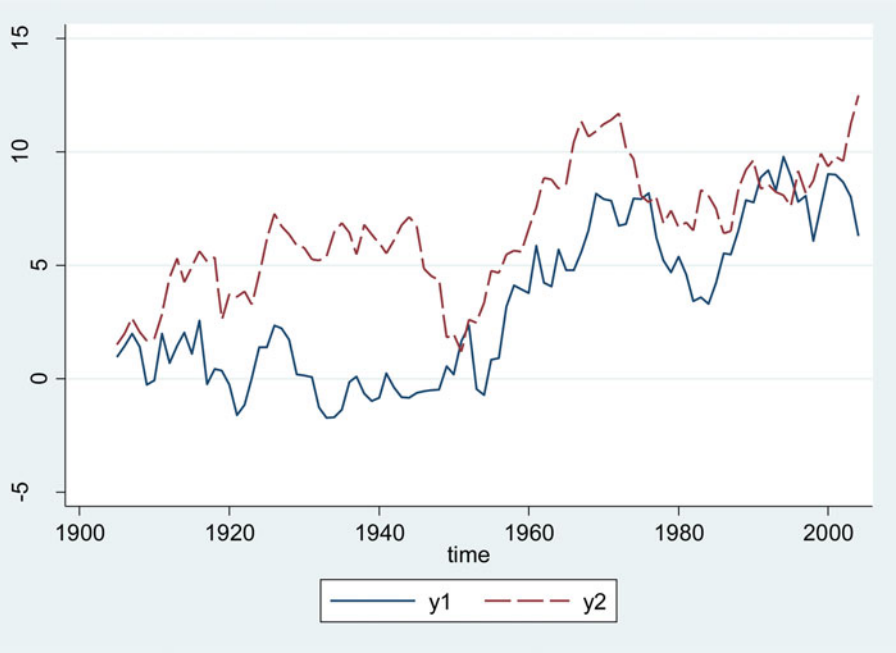

Figure 1. Plot of y_1 and y_2.

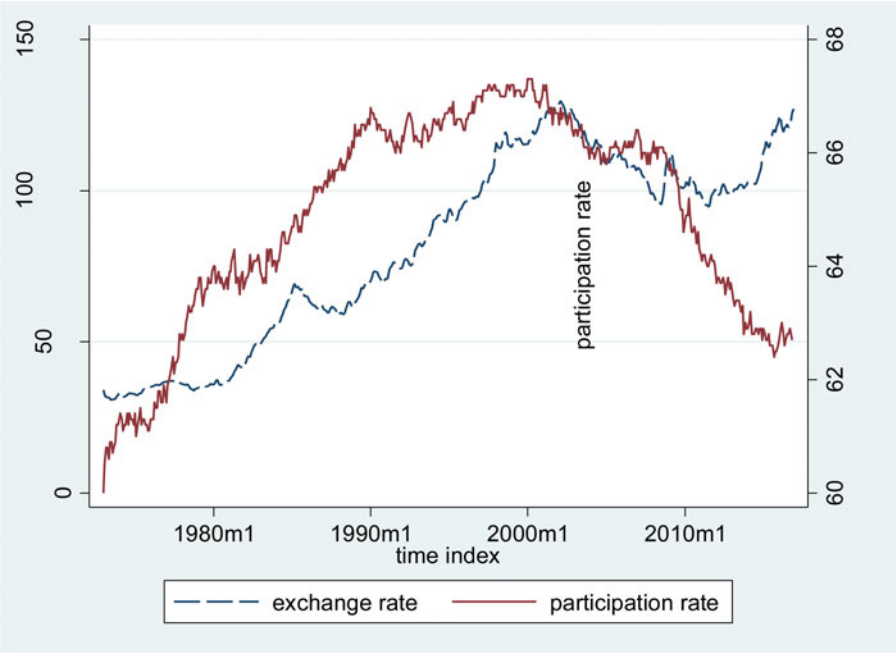

Figure 2. Plot of U.S. exchange rate and labor force participation rate.

yet have a high sample correlation, R^2, and t-statistics. The reason behind the deception is the strong serial correlation in the variables. In this example, they were generated as random walks. The seemingly large sample correlation, R^2, and t-statistic are all artifacts of unmodeled serial correlation.

The secret to detecting spurious regressions of this form is to first examine the time-series plot. If strong serial correlation is present in the series, then static regressions and distributed-lag models need (at a minimum) HAC-type standard errors. If the serial correlation is sufficiently strong (as is the case for a random walk process such as in this example), then HAC standard errors will not be a sufficient correction either. Instead, a better check is to estimate an ADL model. Typically, it is sufficient to include a lagged dependent variable. We report such estimates here:

$$\underset{}{y_{1t}} = \underset{(0.35)}{-0.22} + \underset{(0.04)}{0.91}\ y_{1t-1} + \underset{(0.06)}{0.09}\ y_{2t} + \widehat{e}_t$$

The point estimate of the long-run multiplier is similar to the static regression, but the effect is no longer statistically significant. As a general rule, this is a simple method to break a spurious regression.

As an example using actual data, in figure 2 we plot monthly observations on the U.S. trade-weighted exchange rate, along with the U.S. labor force participation rate. The two seem to have a common relationship. The sample correlation is 0.59. The R^2 from a linear regression is 0.35, and the labor force participation rate has a highly significant t-statistic of 18.

$$\underset{t}{\text{Exchange Rate}} = \underset{(37)}{-578} + \underset{(0.56)}{10.2}\ \underset{t}{\text{Labor Participation}} + \widehat{e}_t$$

The relationship appears to be spurious once we account for the serial correlation. We show this by re-estimating the relationship to include four lags of the dependent variable. (Four lags were included to ensure that the serial correlation in the exchange rate was fully modeled.) In this ADL regression, the

Table 1. U.S. GDP growth rates.

	Mean	Standard Deviation	AR(1) coefficient
1947–1956	4.0	5.3	0.44
1957–1976	3.6	4.2	0.30
1977–1996	3.2	3.5	0.31
1997–2016	2.3	2.5	0.41

labor force participation rate has a t-statistic of just 0.4, so it is statistically insignificant. This is a more reasonable finding, as an economist would naturally be puzzled by a claim that the labor participation rate effects the exchange rate.

$$\begin{aligned}\text{Exchange Rate}_t = {} & \underset{(1.94)}{-0.58} + \underset{(0.05)}{1.43}\ \text{Exchange Rate}_{t-1} - \underset{(0.08)}{0.59}\ \text{Exchange Rate}_{t-2} \\ & + \underset{(0.09)}{0.18}\ \text{Exchange Rate}_{t-3} - \underset{(0.05)}{0.05}\ \text{Exchange Rate}_{t-4} \\ & - \underset{(0.032)}{0.013}\ \text{Labor Participation}_t + \widehat{e_t}\end{aligned}$$

Structural change

Time-series relationships often change. In practical applications, this is important to monitor. As a simple example, take U.S. GDP growth. There is considerable concern that growth rates may have slowed down in recent decades. To illustrate, we compute the average growth rate, standard deviation, and OLS AR(1) coefficient estimate over four historical periods.

Table 1 shows that average growth rates have decreased from an average near 4 percent from 1947 through 1976, to 2.3 percent over the past decade. Simultaneously, the standard deviation has decreased by half. This means that average growth has slowed and has become less volatile. The AR(1) coefficient estimates are all quite similar (between 0.3 and 0.4) and show no pattern, suggesting that there is no meaningful change in the serial correlation patterns.

These calculations, while simple, have important implications for economic analysis and forecasting. If the mean growth rate has changed, then using the full sample to compute forecasts will lead to systematic errors.

Forecasting

A common application for applied time series is forecasting. Given the ADL model (3) with the contemporaneous effect omitted, an estimated direct h-step forecasting equation is:

$$y_{t+h} = \hat{\alpha} + \hat{\phi}_1 y_t + \cdots + \hat{\phi}_p y_{t-p+1} + \hat{\delta}_1 x_t + \cdots + \hat{\delta}_q x_{t-q+1} + \hat{e}_t. \tag{5}$$

The key is that all variables on the right-hand side are dated at time t or before. Estimation is OLS using the sample from $t = 1, \ldots, n$. The point forecast for y_{n+h} is:

$$\hat{y}_{n+h} = \hat{\alpha} + \hat{\phi}_1 y_n + \cdots + \hat{\phi}_p y_{n-p+1} + \hat{\delta}_1 x_n + \cdots + \hat{\delta}_q x_{n-q+1}.$$

A simple $1-\alpha$ forecast interval is based on the normal approximation:

$$\hat{y}_{n+h} \pm \hat{s}_{n+h} z_{1-\alpha/2}$$

where z_α is the $1-\alpha/2$ normal quantile (e.g., $z_\alpha = 1.645$ for an 80% interval), and $\hat{s}_{n+h}$ is the standard error of the forecast. The latter is generated in Stata by the command predict *name*, stdf.

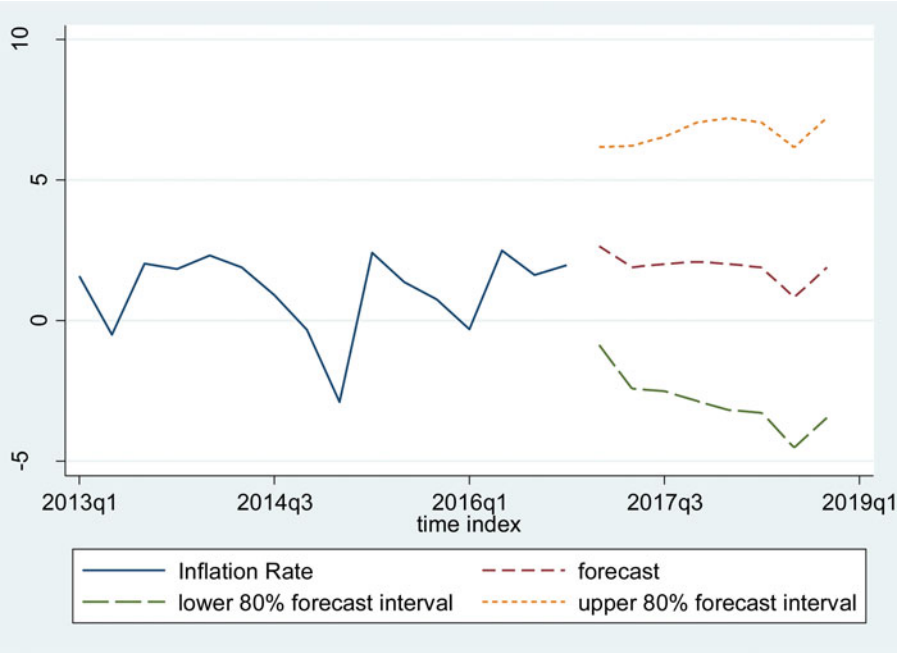

Figure 3. U.S. inflation rate forecasts for 2017–18.

Alternatively, a reasonable approximation is the standard deviation of the regression residual:

$$\hat{s}_{n+h} \approx \left(n^{-1} \sum_{t=1}^{n} \hat{e}_t^2 \right)^{1/2}.$$

For direct forecasts, a separate regression is estimated for each forecast horizon h. Thus, for multiple forecasts, we must estimate multiple OLS regressions.

To illustrate, let us take quarterly U.S. inflation, and use the ADL Phillips curve model selected by the AIC (four autoregressive lags and two lags of the unemployment rate) to make forecasts for 2017–18. This requires estimation of eight regressions of the form (5) for $h = 1, \ldots, 8$. The point forecasts and 80 percent forecast intervals are plotted in figure 3. These are forecasts for quarterly inflation at an annual rate. You can see that the Phillips curve equation predicts that the inflation rate will continue at its current 2 percent rate. The 80 percent forecast intervals show that there is considerable uncertainty in the inflation realizations.

References

Diebold, F. X. 2015. Forecasting in economics, business, finance and beyond. http://www.ssc.upenn.edu/~fdiebold/
Elliott, G., and A. Timmermann. 2016. *Economic forecasting*. Princeton, NJ: Princeton University Press.
Gonzáles-Rivera, G. 2012. *Forecasting for economics and business*. New York: Routledge.
Hamilton, J. D. 1994. *Time series analysis*. Princeton, NJ: Princeton University Press.
Stock, J. H., and M. W. Watson. 2015. *Introduction to econometrics*. 5th ed. Boston: Pearson.
Wooldridge, J. M. 2013. *Introductory econometrics: A modern approach*. 5th ed. Mason, OH: South-Western.