

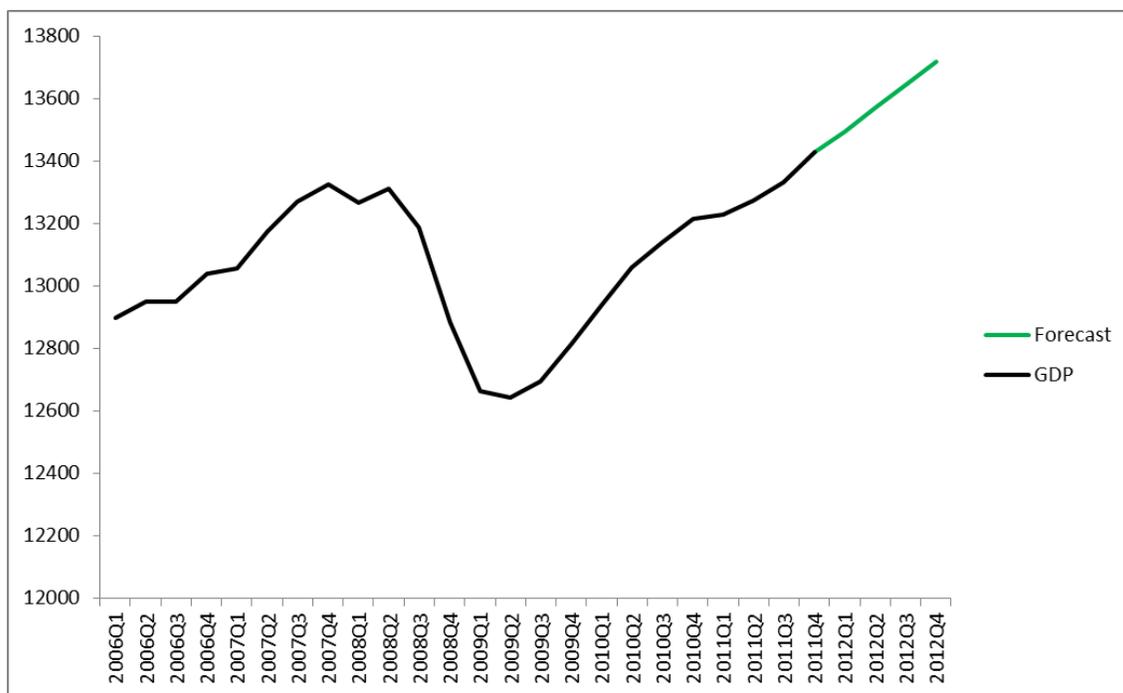
Linear Algebra Semester Project

Forecasting United States Real Gross Domestic Product

By:

Daniel Pellatt

Jared Kubly



$$G_t = 0.0022 + 0.2486 G_{t-1} + 0.002 C_{t-1} + 0.0007 R_{t-1} + 0.2154 G_{t-2} - 0.0014 C_{t-2} - 0.0003 R_{t-2}$$

Introduction

Gross Domestic Product (GDP) is a measure of the total output produced in an economy. It is a measure of the size of an economy in terms of economic activity. The nominal value of the United States economy is currently around 14 trillion dollars per year. The United States has the largest economy of any individual country in the world, although if the countries in the European Union were counted as a single entity it would have a larger GDP (and hence a larger economy) than the United States. Inflation increases the prices of goods and therefore GDP in later years after inflation has occurred can be exaggerated if one wants to compare the economic activity occurring within a country from year to year. Therefore, GDP is often measured in a way that adjusts for inflation. This is done by choosing a base year and expressing GDP for all years in terms of the value of a dollar in that base year rather than the actual nominal values. This new figure of GDP, which can be compared with other years, is called Real GDP. From now on when we say “GDP” we will be referring to Real GDP.

Our goal for this project is to do a simple forecast of United States GDP with a method that utilizes some of the linear algebra covered in class this term. The Federal Reserve Bank of the United States releases information about the United States economy on a regular basis. The first quarter of 2012 GDP data will be released April 27th 2012. We will make a forecast using a method that involves least squares approximations. Our forecast will include a forecast for the data to be released on the April 27th as well as some periods after that.

The Model and Method

The forecasting method we use is known as a vector autoregression (VAR). A VAR is a statistical model that can be used to forecast time series variables. A VAR generalizes a univariate autoregressive model. We will briefly explain this univariate model, then extend it to include multiple variables (becoming a VAR), and then present our VAR.

The univariate autoregressive model is an **approximation** of what is called the Wold representation. Wold's representation theorem says that if a time series variable extends infinitely into the past, has zero-mean, and is covariance stationary, then the correct “model” of the process that generates this variable is an infinite sum of the past values of its “innovations” (unpredictable departures from the expected value) multiplied by coefficients. Zero mean requires you to consider a variable as its departure from its average value, which is assumed to remain constant over time. Covariance stationary means that the covariances between the variable and specific lags of its past values remain constant. Covariance is a measure of the degree to which two variables change together. The lag of a time series variable is its value at

a specified previous time period. For example the lag 2 of y_t is y_{t-2} where t denotes the time period. These assumptions are not so farfetched for many time series data, especially if you consider variables in terms of percentage change.

The univariate autoregressive model approximates Wold's representation with what is called an autoregressive, or $AR(p)$, process. It says that a variable at time t is determined by the values it took at previous time periods and an error term. For example, here are $AR(1)$ and $AR(2)$ processes:

$$AR(1): \quad y_t = \beta * y_{t-1} + \varepsilon_t$$

$$AR(2): \quad y_t = \beta_1 * y_{t-1} + \beta_2 * y_{t-2} + \varepsilon_t$$

Where ε_t is the error term with expected value of zero and constant variance. As can be seen, the time series variable y at time t (denoted y_t) is dependent on its values at previous time periods and (hopefully small) error terms.

Our forecasting method uses a model that generalizes the univariate autoregressive model to include multiple variables. This is called a multivariate autoregression, or vector autoregression (VAR). Now we have several time series variables and each is determined by a specific number (p) of lags of the previous values of each of the time series variables in the model. Here is an example of a simple VAR model with two time series variables, y_1 and y_2 :

$$y_{1,t} = \beta_{11} * y_{1,t-1} + \beta_{12} * y_{2,t-1} + \varepsilon_{1,t}$$

$$y_{2,t} = \beta_{21} * y_{1,t-1} + \beta_{22} * y_{2,t-1} + \varepsilon_{2,t}$$

Now the model incorporates cross-variable dynamics as each time series variable at time t is determined by the lags of itself as well as the lags of the other time series variables in the model. The example is a model that incorporates one lag so it is a $VAR(1)$ model. The right-hand-side of the model equations have the same independent variables but different coefficients attached to them. It is also possible to include constant terms, time trends, etc. in the model as long as they are included in each equation of the VAR.

Each of the time series variables within a VAR model is called an endogenous variable. Variables like constants or time trend variables are called exogenous variables. There is one equation in the VAR for each endogenous variable. Each such equation explains the progression of its endogenous variable in terms of that endogenous variable's own past values as well as the past values of the other endogenous variables in the model.

We include three time series variables in our VAR: real GDP, total US capacity utilization, and the treasury yield (interest rate) on three month treasury bills. These are the endogenous variables. There is one equation in the VAR for each of these three variables.

That equation explains the value of its variable at a certain time as a function of previous values of all three endogenous variables and one exogenous constant term:

Define:

G:= GDP (note: G_t =GDP at time t, GDP_{t-1} =GDP at time t-1, and so on)

C:= Capacity Utilization

R:= Treasury Bills (3 month)

Then our VAR has three equations, one for each of these variables:

$$\begin{aligned} G_t &= \alpha_g + \beta_{gg1} G_{t-1} + \beta_{gc1} C_{t-1} + \beta_{gr1} R_{t-1} + \beta_{gg2} G_{t-2} + \beta_{gc2} C_{t-2} + \beta_{gr2} R_{t-2} + \epsilon_g \\ C_t &= \alpha_c + \beta_{cg1} G_{t-1} + \beta_{cc1} C_{t-1} + \beta_{cr1} R_{t-1} + \beta_{cg2} G_{t-2} + \beta_{cc2} C_{t-2} + \beta_{cr2} R_{t-2} + \epsilon_c \\ R_t &= \alpha_r + \beta_{rg1} G_{t-1} + \beta_{rc1} C_{t-1} + \beta_{rr1} R_{t-1} + \beta_{rg2} G_{t-2} + \beta_{rc2} C_{t-2} + \beta_{rr2} R_{t-2} + \epsilon_r \end{aligned}$$

This could also be written:

$$\begin{pmatrix} G_t \\ C_t \\ R_t \end{pmatrix} = \begin{pmatrix} \alpha_g \\ \alpha_c \\ \alpha_r \end{pmatrix} + \begin{pmatrix} \beta_{gg1} & \beta_{gc1} & \beta_{gr1} \\ \beta_{cg1} & \beta_{cc1} & \beta_{cr1} \\ \beta_{rg1} & \beta_{rc1} & \beta_{rr1} \end{pmatrix} \begin{pmatrix} G_{t-1} \\ C_{t-1} \\ R_{t-1} \end{pmatrix} + \begin{pmatrix} \beta_{gg2} & \beta_{gc2} & \beta_{gr2} \\ \beta_{cg2} & \beta_{cc2} & \beta_{cr2} \\ \beta_{rg2} & \beta_{rc2} & \beta_{rr2} \end{pmatrix} \begin{pmatrix} G_{t-2} \\ C_{t-2} \\ R_{t-2} \end{pmatrix} + \begin{pmatrix} \epsilon_g \\ \epsilon_c \\ \epsilon_r \end{pmatrix}$$

In the VAR system here, the bold β 's denote the coefficients attached to each of the lags of the endogenous variables and the α 's denote constant terms. The β 's, α 's, and ϵ 's are the unknowns. The lags are historic data that is already known (these are denoted with subscript t-1 for the first lag and t-2 for the second lag). As can be seen, the VAR models of each of the endogenous variables as a function of two past values of all the endogenous variables (and a constant and error term). For example GDP at time t is a function of a constant, the value of GDP at time t-1, the value of GDP at time t-2, the value of capacity utilization at time t-1, the value of capacity utilization at time t-2, the value of the treasury yield at time t-1, the value of the treasury yield at time t-2 and an error term.

The idea of forecasting with a VAR model is as follows: determine the coefficients in the VAR's system of equations and then forecast the future values of the variables using that system of equations. Because the values of the endogenous variables in the VAR are functions of their *previous* values and *constants*, we can forecast their values at time t+1 once we know the values of the coefficients. This is done by plugging in the appropriate past values of the endogenous variables. This gives us a one step ahead forecast. We then assume that the forecasted values at t+1 are "correct" and then those values can be used to forecast time t+2 (at which time their values would be inserted in to the appropriate places for which they are now

time t-1 values). Therefore, all that is needed to forecast the variables in the VAR (GDP is the one we are interested in here) is their known values during each previous time period and the values of the coefficients in the system of equations.

The values of the coefficients in the VAR are solved for with linear algebra. This is done via least squares approximations. The least squares approximation can be done one equation at a time for the equations in the VAR (three equations in our model). To see how this is done we will examine the least squares approximation for the GDP at time t equation in the VAR. We considered GDP from 1992 up until the most recent known value of GDP (its 2011 quarter 4 value). We must set up a matrix equation that we can solve for the coefficients that have the “best fit” for the model. Our matrix interpretation of the GDP equation has one equation for each observation of GDP considered. Each equation has the value of GDP at a point in time for which the historic lags are available (one for each quarter that GDP was released from 1992 to 2011):

$$\begin{pmatrix} G_{2011Q4} \\ G_{2011Q3} \\ \dots \\ G_{1992Q4} \end{pmatrix} = \begin{pmatrix} 1 & G_{2011Q3} & C_{2011Q3} & R_{2011Q3} & G_{2011Q2} & C_{2011Q2} & R_{2011Q2} \\ 1 & G_{2011Q2} & C_{2011Q2} & R_{2011Q2} & G_{2011Q1} & C_{2011Q1} & R_{2011Q1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & G_{1992Q3} & C_{1992Q3} & R_{1992Q3} & G_{1992Q2} & C_{1992Q2} & R_{1992Q2} \end{pmatrix} \begin{pmatrix} \alpha_g \\ \beta_{gg1} \\ \beta_{gc1} \\ \beta_{gr1} \\ \beta_{gg2} \\ \beta_{gc2} \\ \beta_{gr2} \end{pmatrix}$$

Our goal is to determine the coefficients $(\alpha_g, \beta_{gg1}, \dots, \beta_{gr6})$. Thus we have $Ax=b$. b is the column vector on the left-hand-side of the matrix equation above where component one is GDP at time 2011 quarter 4, component 2 is GDP at time 2011 quarter 3, and so on until finally the last component is GDP at the first observation during 1992. A is the matrix consisting of a column of ones for the constant and the (known) lag values of the endogenous variables that corresponding to the GDP component in column vector b . For example the second component of b is GDP at time 2011 Q3; then the corresponding row in matrix A has lags one and two of the endogenous variables where lag one is 2011Q2 and lag 2 is 2011Q1. Finally, x is the column vector of coefficients that we wish to determine. The first component of x is α_g , the next is β_{gg1} , and so on with the last component of x being β_{gr6} . Note: if A is an m by n matrix, column vector x has n components and column vector b has m components.

The error terms, the ε 's in the VAR equations, are not included in the $Ax=b$ equation we are considering. The error terms are the amount by which the value of GDP at time t differs from the value of the function to the right of GDP in the VAR system when the error term is not included. We find the "best fit" by determining the coefficients that minimize the error. This is where least squares approximation comes into play. Because b is not in the column space of A , $Ax=b$ is not solvable. Instead we solve $A^*(\hat{x})=p$, where p is the projection of b onto the column space of A . This " \hat{x} " is the coefficient vector that contains, as components, the coefficients that minimize the error and have the "best fit" for the model – " \hat{x} " contains the coefficients for the GDP equation in the VAR model. In maple, then we solve for this " \hat{x} " vector of coefficients. This is done with the usual least squares method. With the matrices and vectors described above, we have $(\hat{x})=(A^T A)^{-1} A^T b$. This is how the coefficients for the GDP equation in the VAR are solved for. The resulting error vector is in the left nullspace of A and is perpendicular to p in the column space; it is the smallest possible error vector. The same method is used to find the coefficients of the other equations in the VAR; a similar matrix equation was set up for each equation in the VAR and the coefficient vector was computed with least squares approximation.

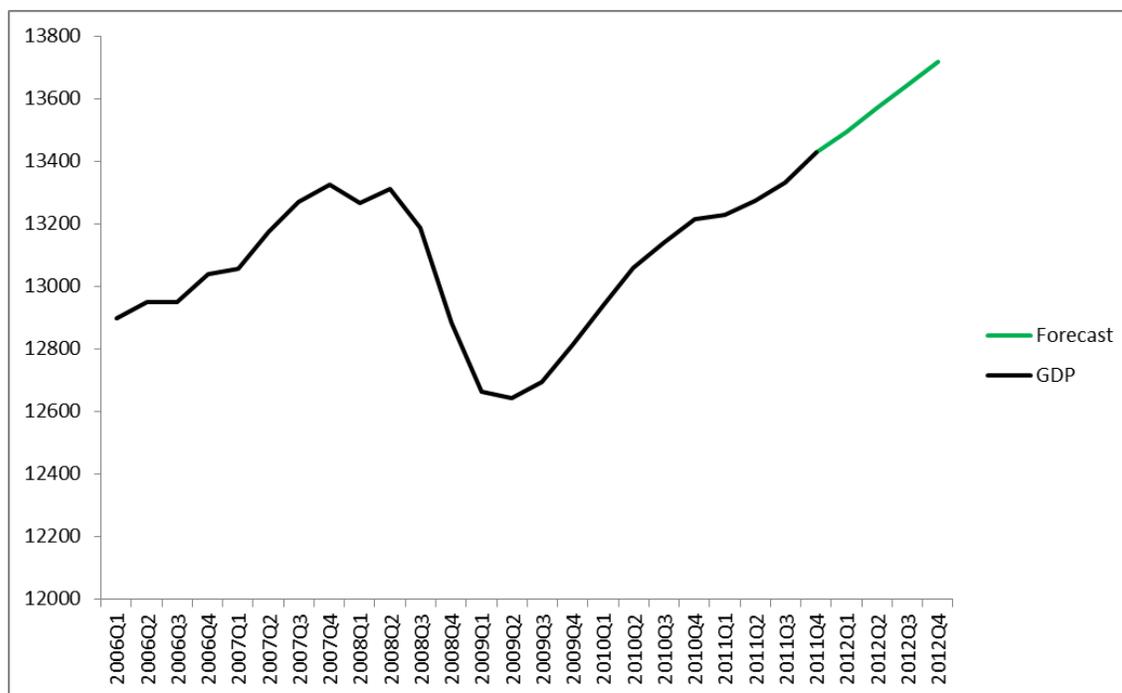
Results

The least squares approximation procedure described above was performed for each of the equations in the VAR to solve for the coefficients in the VAR model. The resulting system of equations from the VAR model, now with the coefficients calculated is as follows:

$$\begin{aligned}
 G_t &= 0.0022 + 0.2486 G_{t-1} + 0.002 C_{t-1} + 0.0007 R_{t-1} + 0.2154 G_{t-2} - 0.0014 C_{t-2} - 0.0003 R_{t-2} \\
 C_t &= 0.1897 + 35.85 G_{t-1} + 0.6678 C_{t-1} + 0.2028 R_{t-1} - 4.163 G_{t-2} - 0.199 C_{t-2} - 0.3255 R_{t-2} \\
 R_t &= 0.057 + 13.519 G_{t-1} + 0.0024 C_{t-1} + 1.537 R_{t-1} - 2.9259 G_{t-2} + 0.0024 C_{t-2} - 0.58 R_{t-2}
 \end{aligned}$$

Then the forecast was performed using the one step ahead method also described above. Each variable was therefore forecasted into the future with one step ahead forecasts. Our goal was to forecast GDP so we will only report the GDP forecast. It is found below:

		Forecasted Q/Q change	Forecasted Annualized Rate of Change
2012Q1	13493.92	64.02	1.92%
2012Q2	13569.67	75.75	2.26%
2012Q3	13642.21	72.54	2.16%
2012Q4	13718.25	76.04	2.25%



Note1: We used the first difference of logs for GDP and the first difference of capacity utilization. The GDP forecast was converted back to real GDP terms for the above forecast.

Note2: This is the general procedure for performing such a vector autoregression forecast. We did not detail some of the more complicated statistical details involved with performing such a forecast (such as making sure the model is stable by checking inverse unit roots). However, these details were paid attention to when carrying out the process of generating the model and coefficients. As we were more interested with carrying out the linear algebra behind the method, we did not spend a lot of time researching and selecting the endogenous variables in the VAR. Also we forecasted GDP as we felt this would be a more interesting topic for presentation. Needless to say, we are not expert econometricians. Our forecast may well be off; the actual quarter 1 2012 GDP data will be released on April 27th for comparison with our forecast.

Works Cited

Diebold, Francis X. *Elements of Forecasting*. Fourth Ed. Mason, OH: Thompson South-Western, 2007. Print.

Strang, Gilbert. *Introduction to Linear Algebra*. Fourth Ed. Wellesley, MA: Wellesley-Cambridge Press, 2009. Print.

"Vector Autoregressive Models for Multivariate Time Series." *University of Washington*. University of Washington, Web. 20 Mar. 2012. <<http://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf>>.