An Introduction to Time Series Regression

Henry Thompson
Auburn University

An economic model suggests examining the effect of exogenous x_t on endogenous y_t with an exogenous control variable z_t. In general functional form $y_t = f(x_t, z_t)$ where x_t and z_t can refer to more than a single variable. This introduction focuses on estimating this single equation as the reduced form of an equilibrium condition.

Time series variables depend on their own history and estimating these underlying processes is the first step to estimating the relationship of interest. The best predictor of y_{t+1} may be its own past values that include the influence of all related variables while the estimated relationship is a hypothetical model and related variables are measured with error. Estimates of structural models isolate the effects of exogenous variables and may suggest ways to improve theory.

In theory, variables in an OLS regression have a normal distribution with a constant mean and each observation equal to the mean plus a white noise error. Distributions of non-stationary variables with trends have low peaks and fat tails. Further, time plays an explicit role. With a positive trend, early observations are below the mean and later ones above the mean. Standard errors assume constant means. Non-stationary variables understate standard errors, resulting in inflation significance and explanatory power.

Consider the OLS regression

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 z_t + \varepsilon_t$$

(1)

with variables y_t, x_t, and z_t in natural logs. The coefficients of log linear models are point estimates of elasticities. The explicit goal is to interpret theory in terms of the estimated coefficient $\alpha_1 = \partial y_t/\partial x_t$. Begin with a theoretical model and derive (1) to relate estimated coefficients to the theoretical model. Rely on theory and regression analysis to suggest the most relevant exogenous variables. The exogenous variable z_t can represent a vector. The residual ε_t should be white noise WN and has to pass critical tests for a zero mean, lack of residual correlation, and constant variance.

The ultimate form of the regression may not be as simple as (1) since OLS requires at least stationary variables. Time series variables may have trends, structural breaks, and heteroskedastic variance.

The typical problem in applied time series analysis is that variables are not stationary implying the standard errors in (1) are understated. If theory suggests x_t should affect y_t but both have trends, they will be correlated and coefficients will appear significant. Significance and explanatory power, however, are overstated.

The key concept in applied time series analysis is whether series are stationary. A stationary series has a long history and converges to a steady state. Stationary is weaker than normality but regressions with stationary variables typically produce reliable statistics. As the number of observations increases, reliability increases at an increasing rate.

The residual error ε_t in (1) should pass the tests of white noise WN. An OLS regression with non-stationary variables leads to residual correlation indicated by a significant correlation $\text{corr}(\varepsilon_t\ \varepsilon_{t-1})$. Residual correlation implies information resides in the residual. Something else must then affect y_t in a systematic way.

A spurious OLS regression has biased and underestimated variances, inflated t-statistics, and an inflated R^2. Coefficient estimates are unbiased, however, just as likely above as below the true value. Estimated coefficients are consistent, converging to the true value as the number of observations increases and the variance approaches zero. Coefficients are in fact super consistent with accelerating convergence as the number of observations increases. Nevertheless, the underestimated standard errors are a nagging problem in application.

A series that is not stationary may be a difference stationary. If so, regressions on differences of the variables will be reliable. Difference stationary random walks may also be cointegrated. An error correction process then adjusts the variable relative to the long run dynamic equilibrium in the economic model. From a practical perspective, a difference regression that is unsuccessful may conceal significant error correction process.

Solve economic models with more than a single dependent variable in reduced form with each dependent variable a function of the exogenous variables. The goal is to estimate this reduced form equation. For example, the market model determines endogenous price P and quantity Q = D = S from the demand function D = D(P,y) and the supply function S = S(P,w). The exogenous variables in the model are the demand shifter y and the supply shifter w. It would be appropriate to estimate Q or P as functions of y and w but inappropriate to estimate Q as a function of P, or vice versa. Theory is flexible in that various assumptions about endogeneity lead to different models and regressions. Be very specific in developing the regression that the right hand side variables are exogenous.

The ISLMBP macroeconomic model provides another example of flexibility, with national income Y as a function of exogenous government spending G, money supply M_s, the foreign interest rate r*, and foreign income Y*. The interest rate r is an endogenous variable and should not be exogenous in a regression on the dependent variable Y. It is possible to estimate a separate equation for r. A floating exchange rate e would be an endogenous variable with an exogenous balance of payments B since e adjusts if $B \neq 0$. If there were a fixed exchange rate, it would be exogenous and B would then be endogenous.

Time series processes will determine the form of variables in regression analysis. Lagged effects may be important. An increase in wage might lower labor input next year. The theoretical model and regression then includes lags of exogenous variables.

Regression options include transforming variables with logarithms, differences, inverses, and lags. The error correction model ECM includes the residual ε_t of the spurious equation (1) in a difference model that separates transitory adjustment from adjustment relative to the long term dynamic equilibrium.

Let theory be the guide to variable selection and endogenous variables. Regression results may suggest ways to refine theory. Empirical analysis leads to improved theory.

The rest of this Introduction is available. Sections headings are:

White noise
Stationary variables
Stationary with a structural break
Difference stationary variables
Unit root with a structural break
Difference models
Error correction model ECM
Lagged transformation models
Other econometric models
Conclusion

Contact thomph1@auburn.edu.

Stationary Flow Chart

WN residuals

(a) zero mean $-t < \mu_\varepsilon/SD < t$

(b) zero covariance $\rho(\varepsilon_t, \varepsilon_{t-1}) < \rho_c = \rho(n - 2)^{.5}/(1 - \rho^2)^{.5}$ LB DW Dubin-h

(c) constant variance ARCH1 $\beta_0 = \beta_1 = 0$ in (2) by F test and e_t in (2) WN by (a) & (b)

AR(1) **Model**

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t$ $(\alpha_1 + 2\sigma) < 1 \Rightarrow$ stationary y_t

↓ non-stationary, ε_t not WN by LB & ARCH

AR(2)

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \varepsilon_t$ $a = [\alpha_1 \pm (\alpha_1^2 + 4\alpha_2)^{.5}]/2$ y_t

↓ non-stationary, ε_t not WN by LB & ARCH unit roots $a_i + 2\sigma_i < 1$ (derive σ_i)

AR(1) with structural break

$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 D + \alpha_3 D y_{t-1} + \varepsilon_t$ $\alpha + 2\sigma < 1$ before & after break y_t, D

↓ non-stationary, ε_t not WN by LB & ARCH

DF

$\Delta y_t = \gamma y_{t-1} + \varepsilon_t$ $\gamma = 0 \Rightarrow$ random walk RW Δy_t ***Unit root**

↓ $\gamma \neq 0$, ε_t not WN by DW & ARCH τ stat

DFc

$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \varepsilon_t$ $\gamma = 0 \Rightarrow$ trend stationary ε_t WN

↓ $\gamma \neq 0$, ε_t not WN by DW & ARCH τ_μ stat

ϕ_1 test $\alpha_0 = \gamma = 0$ => RW Δy_t ***Unit root**

DFt

$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_1 t + \varepsilon_t$ $\gamma = 0$ ε_t WN

↓ $\gamma \neq 0$, ε_t not WN by DW& ARCH τ_T stat

ϕ_2 test $\alpha_0 = \gamma = \alpha_1 = 0$ Δy_t ***Unit root**

ADF

$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \alpha_1 \Delta y_{t-1} + \alpha_2 t + \varepsilon_t^{ADF}$ $\gamma = 0$ ε_t^{ADF} WN

↓ $\gamma \neq 0$, ε_t not WN by DW & ARCH τ_T stat

ϕ_3 test $\alpha_0 = \gamma = \alpha_1 = \alpha_2 = 0$ Δy_t ***Unit root**

ADF(2)

add $\alpha_1 \Delta y_{t-2}$

↓

Unit root with structural break

$y_t = a_0 + a_2 t + \mu_2 D + \varepsilon_t^P$

Perron $\varepsilon_t^P = a_1 \varepsilon_{t-1}^P + e_t$ $a_1 = 1 \Rightarrow \Delta\varepsilon_t^P = \Delta y_t$ Δy_t ***Unit root**

Models Spurious model $y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 z_t + \varepsilon_t^s$

ε_t^s stationary by EG $\Delta\varepsilon_t^s = a_1 \varepsilon_{t-1}^s + e_t$ with $a_1 < 0 \Rightarrow$ ECM

Difference $\Delta y_t = \alpha_0 + \alpha_1 \Delta x_t + \alpha_2 \Delta z_t + u_t$ or $\Delta y_t = \alpha_1 \Delta x_t + \alpha_2 \Delta z_t + u_t$ (spurious coefficients)

ECM $\Delta y_t = \beta_0 + \beta_1 \Delta x_t + \beta_2 \Delta z_t + \beta_y \varepsilon_{t-1}^s + u_t$ with lags $\beta_1 \Delta x_{t-1}$ etc

LTM $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 z_t + \alpha_4 x_{t-1} + \alpha_5 z_{t-1} + e_t$ if ε_t^s RW by EG test with $a_1 = 0$

Detrend $y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \varepsilon_t^{dt}$ ε_t^{dt} WN

Stationary Table

	AR(1)	DF	DFc	DFt	ADF	AR(2)	P_{SB}
y_t	α_1+2σ<1 LB*	t*	t F*	t F	t F	$a_i + 2\sigma_i$ > 1	t
ε_t				DW*	DW ARCH*		DW ARCH
x_t	α_1+2σ>1	t	t				
ε_t		DW*	DW ARCH				
z_t	α_1+2σ>1	t*	t F*	t F*	t F		
ε_t					DW ARCH		
τ_{DF} φ		τ	τ_μ ϕ_1	τ_T ϕ_2	τ_T ϕ_3		-3.76

Notes:

1. Variables are not AR(1) stationary, y_t due to residual correlation. Not necessary to report WN test if coefficient test fails. Other AR(p) or ARMA(p,q) models can be reported.

2. y_t has a unit root by Perron structural break test P_{SB} but not by the DF ($t > \tau$), DFc ($F > \phi_1$), DFt (DW), or ADF (ARCH). y_t is not AR(2) stationary.

2. x_t has unit root by DFc test but not DF test due to residual correlation.

3. z_t is RW with $F < \phi_3$ in ADF