Math & Stat for MBA
Lecture 4
Basic Time Series Regression Analysis

Keith Lee

September 28, 2021

Differences between cross-sectional and time series data I

- Time series data come with a temporal ordering, usually from earliest to latest.
  - For many purposes, the ordering of the data is important.
- We cannot think of time series data as a random sample of units (individuals, firms, schools, and so on) from a large population.
  - Therefore, we cannot realistically impose random sampling (MLR.2) when using time series data.
  - In fact, time series data almost always exhibit correlation across time, sometimes very strongly.
    - To ensure standard inference applies, we will have to control the dependence from being too strong (persistence).
- A sequence of random variables indexed by time, $\{y_t\}_{t=1}^T$, is called a **stochastic process** or a **time series process**. A sample is one realized path out of many possible paths the process could take.

Differences between cross-sectional and time series data II

- For time series data at monthly or quarterly (or even weekly or daily) frequencies, seasonality can be an issue.
  - Examples: Christmas effect on expenditures, effectiveness of fertilizer on production.
- It is fairly standard to include seasonal dummies (and interactions) to deal with this.

$$y_t = \alpha_0 + \beta x_t + \gamma_1 s_{1t} + \gamma_2 s_{2t} + \gamma_3 s_{3t} + \epsilon_t$$
$$\{s_{1t}, s_{2t}, s_{3t}\} \text{ are seasonal dummies}$$
$$e.g., \;\; s_{1t} = 1, \text{ if } t \text{ falls in 1st quarter, } = 0 \text{ otherwise}$$

  - This is equivalent to running a regression of the deseasonalized series y on the deseasonalized series x. ("**Seasonally adjusted**") This is due to the "Partialling out" interpretation of Multiple Regression.

Differences between cross-sectional and time series data III

- Many time series variables exhibit trends
  - Examples: Exam score per study hour, GDP, and etc.
- When running regression using time series variables that are trending, we should be careful not to confuse a common tendency to grow (or fall) with that of a causal relationship (spurious regression problem).
  - Including a time trend in our model may allow us to prevent this, e.g., (linear trend assumed)

$$y_t = \alpha_0 + \beta x_t + \gamma t + \epsilon_t$$

  - Adding a time trend in our model is the same as working with detrended series.
    - Recall: "Partialling Out" interpretation of multiple regression.

Spurious regression problem (example)

- Yule (1926) observed a strong correlation between the proportion of UK marriages in church (ChurchMarr) and the UK mortality rate (Mortality) using data for the years 1866-1911. The regression

$$Mortality_t = \alpha_0 + \alpha_1 ChurchMarr_t + u_t$$

suggests a significant (positive) α_1. ($r = 0.9512$)

  - Obviously, it is very hard to explain how the proportion of marriages in church can possibly effect the mortality rate ("Non-sense Correlations in time-series").
  - **The high correlation, and significance of α_1 is purely a result of the common trending nature in both variables.** (see graph next page)
  - When we run the regression

$$Mortality_t = \alpha_0 + \alpha_1 ChurchMarr_t + \alpha_2 t + u_t$$

we expect the significance of α_1 to disappear.

Spurious regression problem (example)

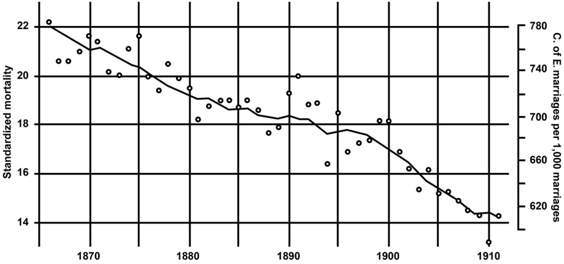


Figure: Correlation between standardized mortality per 1,000 persons in England and Wales (circles), and the propertion of Church of England marriages per 1,000 of all marriages (line), 1866-1911. $r = +0.9512$.

- Yule, G.U. (1926): "Why do we somethimes get non-sense correlations between time-series?" Journal of the Royal Statistical Society, 89, 1-63

Static Models I

- The simplest model with time series is **static model**, where y_t (dependent variable at time t) is explained by regressors at the same time t.
- With one regressor z_t , we have

$$y_t = \beta_0 + \beta_1 z_t + u_t \quad for \;\; t = 1, ..., T$$

Example: Static Phillips Curve

Can consider the static Phillips curve:

$$inf_t = \beta_0 + \beta_1 unem_t + u_t \quad for \;\; t = 1, ..., T$$

where inf_t is, say, the annual rate of inflation during year t, and $unem_t$ is annual unemployment rate during year t.
β_1 attempts to measure the trade-off between inflation & unemployment.

Static Models II

- Static models are generally used when we are interested in a contemporaneous relationship.
  - They cannot capture effects that take place with a lag.
- In general, static models are not good for forecasting.
  - To forecast y_{T+1} at time T, we have to know z_{T+1} at time T. In example: to forecast inflation in time $T+1$, we would need to know what unemployment is in time $T+1$

$$inf_t = \beta_0 + \beta_1 unem_t + u_t \quad for \;\; t = 1, ..., T$$

  - Moreover, it ignores the fact that usually past outcomes of y help predict future values of y.

Finite Distributed Lag (FDL) Models I

- Suppose change in z today can affect y up to two periods into the future. This calls for a **FDL model** of order 2 (also known as **AR(2)**)

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

  - FDL model is good for estimating lagged effects of z. Recognizes that people often react with a lag to policy changes.

Example: Personal Exemption and Fertility.

The effect of making it monetarily more attractive to have children - by increasing the value of the personal exemption (pe) - is unlikely instantaneous (*biology*). Allowing for a two-year effect, we may model the general fertility rate (gfr) as:

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t$$

Finite Distributed Lag (FDL) Models II

Example: Minimum Wage and Employment.

Suppose we have monthly data on employment and the minimum wage. We may expect that the effect of a change in the minimum wage will take several months to have its full effect on employment.

- An FDL model of order q, FDL(q), or AR(q), is

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + ... + \delta_q z_{t-q} + u_t$$

  - As a practical matter, the choice of q can be hard. Often dictated by frequency of data.
    - With annual data, q is usually small. With monthly data, q is often chosen as 12 or 24 or even higher, depending on how many months of data we have.
  - Under some assumptions we can use an F test to see if additional lags are jointly significant.

Finite Distributed Lag (FDL) Models III

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + ... + \delta_q z_{t-q} + u_t$$

- δ_0, the coefficient on the contemporaneous z, is called the

$$\textbf{Short-Term Propensity}(\text{STP}) = \delta_0$$

  - It tells us the immediate change in y when z increases by one unit.
  - Is the change permanent?
- The sum of all lag coefficients is called the

$$\textbf{Long-Term Propensity}(\text{LTP}) = \delta_0 + \delta_1 + ... + \delta_q$$

  - LTP answers following thought experiment: Suppose z increases permanently today (e.g., minimum wage increases by $1.00/hour permanently): LTP is (ceteris paribus) change in y after change in z has passed through all q time periods.

Finite Distributed Lag (FDL) Models IV

- We can have more than one variable appear with multiple lags.

For example, a simple equation to explain how the Federal Reserve Bank in the U.S. changes the Federal Funds Rate is

$$\begin{aligned} ffrate_t = &\alpha_0 + \delta_0 inf_t + \delta_1 inf_{t-1} + \delta_2 inf_{t-2} \\ &+ \gamma_0 gdpgap_t + \gamma_1 gdpgap_{t-1} + \gamma_2 gdpgap_{t-2} + u_t, \end{aligned}$$

where inf_t is the inflation rate and $gdpgap_t$ is the GDP gap (measured as a percent).

- FDLs are often more realistic than static models (they typically forecast better) because they account for some dynamic behavior.
  - Nevertheless, FDL is not the most preferred for forecasting because they do not allow lagged y's to affect current y.

Models with Lagged Dependent Variables I

- With time series data, there is the possibility of allowing past outcomes on y to affect current y. The simplest model is

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t,$$

  - Called an **autoregressive model of order 1**, or **AR(1)**.
  - This simple model typically does not have much economic or policy interest because we are just using lagged y to explain current y.
    - We can add even more lags of y to explain y_t .
    - An AR(p) process is given by

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + u_t$$

  - Autoregressive models can be remarkably good at forecasting, even compared with deep learning models (which is no more than a computer exercise of non-linear regression). Why?
  - AR order + coefficient - a joint problem

Models with Lagged Dependent Variables II

- It is easy to add other explanatory variables along with a lag:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 z_t + u_t$$

  - β_2 measures the effect of changing z_t on y_t , holding y_{t-1} fixed. It is a kind of **short-term effect of z on y**.
  - Controlling for y_{t-1} while estimating the effect of z_t can be effective for estimating the causal effect of z_t on y_t : it recognizes that the policy variable (z_t) may be correlated with y_{t-1}
    - Since y_{t-1} is likely a relevant regressor, its omission will cause OVB (more later) if z_t is related to y_{t-1}.
  - The **long-term effect of z on y** is given by $\beta_2/(1-\beta_1)$
    - Assumes $|\beta_1| < 1$
    - In the LT: $y = \beta_0 + \beta_1 y + \beta_2 z$, rewriting yields $y = \frac{\beta_0}{1-\beta_1} + \frac{\beta_2}{1-\beta_1} z$

Models with Lagged Dependent Variables III

Example:

Consider

$$gemp_t = \beta_0 + \beta_1 gemp_{t-1} + \beta_2 gminwage_t + u_t$$

where $gemp_t$ is, say, monthly employment growth in an economy, or sector of the economy, and $gminwage_t$ the percentage growth in the minimum wage.
β_2 measures the effect of changing minimum wage growth on employment growth this period, and we expect $\beta_2 \leq 0$.

- By controlling for $gemp_{t-1}$, we allow the possibility that $gminwage_t$ reacts to past employment growth.

Models with Lagged Dependent Variables IV

- Statistically, models with lagged dependent variables are more difficult to study:
  - OLS estimators are no longer unbiased (finite sample property) under any assumption
    - Therefore, use of large-sample analysis becomes very important.
  - Unfortunately, large-sample analysis is trickier with time series data because of correlation (dependence) across time.
    - Recall: With cross-sectional data, we relied on random sampling. Here we will need to restrict the dependence.

Time Series Models - Summary

- **Static Models**

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

  - Contemponeous (instantaneous) effect

- **Autoregressive Models - AR(2)**

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t$$

  - Capture effects that take place with a lag.

- **Models with Lagged Dependent Variables - ARMA(1,2)**

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 z_t + \beta_3 z_{t-1} + \beta_4 z_{t-2} + u_t$$

  - Allows lagged outcomes on y to directly affect current outcomes.
  - Also called Autoregressive and Moving Average

Finite-Sample Analysis of OLS for TS Data I

- Consider the time series process: $\{(y_t, x_{t1}, ..., x_{tk}) : \ t = 1, ..., T\}$ We often use T to denote the sample size in time series.
  - We explicitly use t to index time.
- OLS on time series data can be unbiased, but the assumptions needed are pretty restrictive
  - An important reason for this is that when using time series we cannot maintain the assumption of random sampling. Time series data are almost always correlated across time, sometimes very strongly.
- We review the assumptions used to derive the same kinds of finite-sample properties (unbiasedness, variance calculations, normality)

Finite-Sample Analysis of OLS for TS Data II

- Gauss-Markov (GM) assumptions for Time-Series (TS)
- **TS.1 No Perfect Collinearity**
  - Rules out perfect linear relations among the explanatory variables.
  - High correlation among x_{tj}'s (problem of **near multicollinearity**) does not violate this assumption but can yield imprecise parameter estimates.
    - This is particularly true in FDL models, such as

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

    - If $\{z_t\}$ is slowly moving over time, then z_t, z_{t-1}, and z_{t-2} can be highly correlated
    - Hence, DL coefficients δ_j estimated imprecisely.

Finite-Sample Analysis of OLS for TS Data III

- **TS.2 Linear in Parameters.**

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t, \quad t = 1, \ldots, T.$$

- Observe that we can write all of the previous examples as a time series regression model by appropriate choice of x_{tj}.
    - For example, if

    $$y_t = \alpha_0 + \delta_1 z_t + \delta_2 z_{t-1} + \delta_3 z_{t-2} + u_t$$

    then

    $$x_{t1} = z_t, \quad x_{t2} = z_{t-1}, \quad x_{t3} = z_{t-2}$$

    so there are $k = 3$ explanatory variables. The slopes $\beta_j(\delta_j)$ are the distributed lag parameters.

Finite-Sample Analysis of OLS for TS Data IV

- **TS.3 Zero Conditional Mean.**

$$\mathbb{E}(u_t|\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_t}, ..., \mathbf{x_T}) \equiv \mathbb{E}(u_t|\mathbf{X}) = 0 \text{ for each } t$$

where $x_t = (x_{t1}, ..., x_{tk})$ are the explanatory variables for y_t and $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, .., \mathbf{x_T})$.

  - The assumption ensures that the unobserved error is "uncorrelated" with the explanatory variables
  - Because we no longer assume independent observations, we need to explicitly rule out correlation between u_t and x_{sj} even when the time periods s and t do not match up.
  - In practice, we ask whether u_t **is uncorrelated with each** x_{sj} **for all** t **and** s**, including** $t = s$ **and all variables** $j = 1, ..., k$
  - TS.3 is often called **strict exogeneity** of $\{\mathbf{x_t} : t = 1, ..., T\}$.

Finite-Sample Analysis of OLS for TS Data V

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \ldots + \beta_k x_{tk} + u_t, \;\; t = 1, \ldots, T$$

Theorem

(**Unbiasedness of OLS for Time Series**): Under Assumptions TS.1, TS.2, and TS.3, the OLS estimators are unbiased

$$\mathbb{E}(\widehat{\beta}_j) = \beta_j, \;\; j = 0, \ldots, k.$$

- Note: regressors $\{x_{tj}\}$ are allowed to be correlated across time.
- Also errors, $\{u_t\}$ are allowed to be correlated across time.
- What we are ruling out with TS.3 is correlation between x_{tj} and u_s for any t and s.
- **Worry: Assumption TS.3 is often considered to be too strong.**

Finite-Sample Analysis of OLS for TS Data VI

- Strict exogeneity assumption rules out some practically important situations

1. Models with lagged dependent variables.
   - As soon as y_{t-1} is included among the regressors strict exogeneity fails.
2. Correlation of the error and future values of regressors.
   - Consider

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t$$

     We may argue that we have included enough lags in the FDL to ensure that u_t is uncorrelated with all past z.
   - It is possible (even likely) that z_{t+1} reacts to changes in the unobserved part of y_t (u_t).

Finite-Sample Analysis of OLS for TS Data VII

Example : Suppose we have an FDL relationship between inflation and the federal funds rate:

$$inf_t = \alpha_0 + \delta_0 ffrate_t + \delta_1 ffrate_{t-1} + \delta_2 ffrate_{t-2} + u_t$$

If we assume two lags of the FF rate suffice, we need not worry about correlation between ut and further lags of *ffrate*.
Nevertheless, we may be concerned about correlation between u_t and $ffrate_{t+1}$ if the Fed decides to respond to a positive shock to inflation at time t (i.e., $u_t > 0$) by increasing $ffrate_{t+1}$

- This would violate the assumption of strict exogeneity
    - It may be more appropriate in this example to make the weaker assumption (TS.3) of "contemporaneous" uncorrelatedness.

Finite-Sample Analysis of OLS for TS Data VIII

- **TS.3 Contemporaneous exogeneity**

$$\mathbb{E}(u_t|\mathbf{x_t}) = 0$$

  - This assumption implies uncorrelatedness between errors (u_t) and regressors (x_t).
    - **If** $x_t = (z_t, z_{t-1}, z_{t-2})$ **, the assumption requires** u_t **to be uncorrelated with** z_t, z_{t-1} **and** z_{t-2}.
    - **If** $x_t = (z_t, z_{t-1}, z_{t-2})$ **, the assumption requires** u_t **to be uncorrelated with** y_{t-1}, z_t, z_{t-1} **and** z_{t-2}.
  - Contemporaneous exogeneity does not restrict correlations between the error and explanatory variables across other time periods.
    - In the same example, there may be correlation between u_t and z_{t+1}.
- While this assumption is not enough to ensure unbiasedness of OLS, it will be enough for large-sample properties (e.g., consistency)
- Any insight on BigData?

Finite-Sample Analysis of OLS for TS Data IX

- Unbiasedness says nothing about how precise the OLS estimators are.
  - To obtain the familiar expressions for the variances of the OLS estimators, we add two more assumptions: homoskedasticity and no serial correlation.
- **TS.4-1 Homoskedasticity.**

$$Var(u_t|\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_t}, ..., \mathbf{x_T}) = \sigma^2 \text{ for each } t$$

  - Variance of u_t cannot depend on $\mathbf{x}s$ for any s or change over time for reasons we do not know.
  - Some form of homoskedasticity is needed for usual variance formulas and Gauss-Markov Theorem to hold.
  - This version of violation is called "Hetero-skedasticity".

- **TS.4-2 No Serial Correlation.**

$$Corr(u_t, u_s|\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_t}, ..., \mathbf{x_T}) = 0 \text{ for all } t \neq s$$

  - For the usual OLS analysis to be valid with TS data, we must rule out correlation in the errors over time (automatically satisified in cross-sectional setting by random sampling).
    - In practice, do not worry about the conditioning on $\mathbf{x_1}, ..., \mathbf{x_T}$. Just consider $Corr(u_t, u_s)$.
  - When TS.4-2 is violated, that is, when $Corr(u_t, u_s) \neq 0$ for some (t, s) pairs, the errors exhibit **serial correlation** or **autocorrelation**.
    - In static models TS.4-2 is most likely to be violated: the dynamics will then form part of the error term (our ignorance)

Finite-Sample Analysis of OLS for TS Data XI

- Confusion can arise because there are three are three very different kinds of "correlations" that may arise with time series regression.
  1. The explanatory variables, x_{tj}, might be correlated over time.
     - This is almost always true. For example, the correlation between $unem_t$ and $unem_{t-1}$ is 0.752.
     - We only ruled out perfect correlation in regressors by TS.2.
  2. Correlation between x_{sj} and u_t.
     - If the errors or any of the regressors are correlated, TS.3 is violated, and OLS is biased. So we are ruling out this kind of correlation.
  3. Correlation between u_s and u_t
     - This is the problem of serial correlation ruled out in TS.4-2.
     - The presence of serial correlation itself does not cause bias in the OLS estimators but it is imposed to obtain familiar variance formula.

Finite-Sample Analysis of OLS for TS Data XII

Theorem

(OLS Variance Formulas for TS): Under Assumptions TS.1 to TS.4, the usual OLS variance formulas are valid, i.e.,

$$Var(\widehat{\beta}_j|X) = \frac{\sigma^2}{SST_j(1-R_j^2)} \quad for \;\; j = 1, ..., k$$

- Assumptions GM.1 through GM.4 are the **Gauss Markov assumptions for time series data.** Efficiency!

Theorem

(Gauss-Markov Theorem for TS): Under TS.1 through TS.4, the OLS estimators are BLUE: the best, linear, unbiased estimators.

Finite-Sample Analysis of OLS for TS Data XIII

- The standard error of $\widehat{\beta}_j$ is the same as before:

$$SE(\widehat{\beta}_j) = \sqrt{\frac{\widehat{\sigma}^2}{SST_j(1 - R_j^2)}}$$

  - As with CS regressions, this is reported as the default standard error.
    - We use the unbiased estimator of $\sigma^2 = Var(u_t)$

$$\widehat{\sigma}^2 = (T - k - 1)^{-1} \sum_{t=1}^{T} \widehat{u}_t^2$$

  - **Heteroskedasticity or serial correlation invalidate this simple formula.**
    - We should use robust (HAC) standard errors if we are concerned about the validity of TS.4-1 or TS.4-1 (Optional for MBA)

Finite-Sample Analysis of OLS for TS Data XIV

- Finally, to enable us to perform exact inference, we add:
- **TS.5 Normality**. $\{u_t\}$ is independent of $\mathbf{x_1}, .., \mathbf{x_T}$ and

$$u_t \sim i.i.d.\ \ Normal(0, \sigma^2), \quad t = 1, 2, ..., T$$

Theorem

(Statistical Inference for TS): Under TS.1 to TS.5, all of the statistical inference procedures for the cross-sectional case carry over to time series.

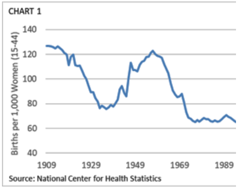
- Under the null, our t statistics have t_{T-k-1} distribution and F statistics have F distributions (usual confidence intervals hold).

Summary Finite-Sample Analysis of OLS for TS Data

1. Under TS.1-3, OLS is unbiased. Strict exogeneity of regressors (TS.3) is key but rules out some interesting cases.
2. If we add homoskedasticity (TS.4-1) and no serial correlation (TS.4-2) assumptions, OLS is BLUE and usual variance formula holds. Serial correlation (violation of TS.4-2) is often a problem.
3. If we add normality (TS.5), then exact inference is possible.

Finite-Sample Analysis TS Data: Example I

EXAMPLE: Effects of Tax Policy on the U.S. Fertility Rate



- *gfr* is the number of children born per 1,000 women 15-44
- *pe* is the real value of the personal tax exemption,
- *ww*2 and *pill* are dummy variables (WW II, avail. birth control pill)

With a sample data, we estimate the static equation

$$\widehat{gfr}_t = \underset{(3.21)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24} ww2_t - \underset{(4.08)}{31.59} pill_t, \quad T = 72,\ R^2 = .473$$

Finite-Sample Analysis TS Data: Example II

- $\widehat{gfr}_t = \underset{(3.21)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24} ww2_t - \underset{(4.08)}{31.59} pill_t, \quad T = 72,\ R^2 = .473$
  - Under TS.1-5, we can reject $H0 : \beta_{pe} = 0$ against the two-sided alternative at less than the 1% significance level.
  - The estimated effect is rather large: a \$100 increase in *pe* increases the estimated fertility rate by 8.3 children per thousand women.
  - Fertility rates were much lower, on average, during WW II and after the introduction of the birth control pill.

Finite-Sample Analysis TS Data: Example III

- A finite distributed lag model with two lags of *pe* gives

$$\widehat{gfr}_t = \underset{(3.28)}{95.87} + \underset{(.126)}{.073}\, pe_t - \underset{(.1557)}{.0058}\, pe_{t-1} - \underset{(.126)}{.034}\, pe_{t-2}$$
$$- \underset{(10.13)}{22.13}\, ww2_t - \underset{(3.98)}{31.30}\, pill_t, \quad T = 70,\ R^2 = .473$$

- The DL coefficients are very imprecisely estimated. None is statistically different from zero using the t statistics.
  - Joint F statistic for pe, pe_{t-1}, and pe_{t-2} gives p-value $= .012$, so they are jointly significant.
- The LTP equals $.073 - .0058 + .034 \approx .101$.
  - Gives the long-run increase in *gfr* if *pe* increases permanently by $1
  - Its t statistic (obtained using the lincom command) is 3.38. So we estimate a strong long-run effect (roughly 10 children per 1,000 women for an increase in pe of $100).
  - Its 95% CI is about [.041, .160].