Econ 2120: Section 5

Panel Data: Generalized Linear Predictor

Ashesh Rambachan

Fall 2018

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Reference

Classic textbook: Wooldridge – Econometric Analysis of Cross-Section and Panel Data

- Beautifully written and easy read.
- Will draw on it for portions of these notes.

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Motivating Story

Consider a population of individuals. Choose one at random.

- The outcome variable is Y_t – log of earnings.
- The vector of predictor variables is R_t – education, tenure, industry, etc. This is $K \times 1$.

There are T time periods so we observe (Y_t, R_t) for $t = 1, \ldots, T$.

- **NOTE**: Cross-walk to the notes: $T := M$, Example := families.

Wish to construct linear predictors of Y_t using R_t but imposing that the coefficients are the same across $t = 1, \ldots, T$.

Notation

Let

$$\underset{T\times 1}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_T \end{pmatrix}, \quad \underset{T\times K}{R} = \begin{pmatrix} R_{11} & \dots & R_{1K} \\ \vdots & & \vdots \\ R_{T1} & \dots & R_{TK} \end{pmatrix}.$$

Goal: Construct linear predictor of Y given R.

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Generalized Linear Predictor: Projection

Goal: Predict Y with R – use projection machinery to tackle this problem.

Predictor is

$$\underset{T\times K}{R}\,\underset{K\times 1}{\beta} = \begin{pmatrix} R^{(1)} & \ldots & R^{(K)} \end{pmatrix} \beta$$
$$= R^{(1)}\beta_1 + \ldots + R^{(K)}\beta_K,$$

where $R^{(j)} = (R_{1j}, \ldots, R_{Tj})'$ is the j-th column of R.

The prediction error is

$$\underset{T\times 1}{U} = Y - R\beta.$$

Generalized Linear Predictor: Projection

Recall: Best linear predictor – U scalar

Choose β to minimize $E[U^2]$.

Generalized Linear Predictor: U is $T \times 1$ vector – need to generalize mean-square error.

We'll consider a "weighted" mean sum of squares:

$$E[U'\Phi U] = \sum_{s=1}^{T} \sum_{t=1}^{T} \phi_{st} E[U_s U_t],$$

where Φ is a $T \times T$ symmetric, positive definite matrix.

Intuition: We need a way to "trade-off" errors across equations $U_t = Y_t - R_t'\beta$.

Simplest option: $\Phi = I_T$ and $E[U'\Phi U] = \sum_{t=1}^{T} E[U_t^2] \implies$ treat errors across equation symmetrically and view equations as "independent."

Generalized Linear Predictior: Projection

Recall: Best linear predictor

Want to minimize mean-square error, $E[U^2]$.

To do so, defined inner product $\langle X, Y \rangle = E[XY]$ and cast problem as minimum norm problem.

$$E[U^2] = \|U\|^2$$

Generalized Linear Predictor: Same idea!

Want to minimize, $E[U'\Phi U]$. To do so, we'll define an associated inner product.

Let U, V be $T \times 1$ vectors and Φ be a $T \times T$ non-random, positive definite, symmetric matrix. Define

$$\langle U, V \rangle_\Phi = E[U'\Phi V].$$

Can show this satisfies properties of an inner product.

Generalized Linear Predictor: Projection

So, we have that

$$E[U'\Phi U] = \|U\|_\Phi^2.$$

Generalized Linear Predictor Coefficients: β solves

$$\beta = \arg\min_c E[U'\Phi U] = \arg\min_c \|Y - R\beta\|_\Phi^2.$$

Denote this as

$$E_\Phi^*[Y|R] = R\beta.$$

In general, β will depend on Φ!

Generalized Linear Predictor: Orthogonality Conditions

The coefficients of the genearlized linear predictor β with weights Φ solves a minimum-norm problem:

$$\beta = \arg\min_{c} \| Y - R\beta \|_{\Phi}^{2}.$$

By the projection theorem, β is characterized by a set of orthogonality conditions. What are they?

We are projecting Y onto the linear subspace spanned by the columns of R

> Equivalently, projecting Y onto the space of all linear functions of the columns of R.

Generalized Linear Predictor: Orthogonality Conditions

So, the orthogonality conditions are

$$\langle Y - R\beta, R^{(j)} \rangle_{\Phi} = 0$$

for $j = 1, \ldots, K$. Stacking these horizontally and subbing in the definition of the inner product, we have

$$E[\underset{1\times T}{(Y - R\beta)'} \underset{T\times T}{\Phi} \underset{T\times K}{R}] = \underset{1\times K}{0}$$

Can rewrite this:

$$E[(Y - R\beta)'\Phi R] = E[Y'\Phi R] - \beta' E[R'\Phi R] = 0$$

So, β is characterized by

$$E[R'\Phi R]\beta = E[R'\Phi Y].$$

Generalized Linear Predictor: Orthogonality Conditions

The projection $R\beta$ is unique.

If R has full column rank (with probability 1), then $E[R'\Phi R]$ is invertible and β is unique with

$$\beta = E[R'\Phi R]^{-1} E[R'\Phi Y].$$

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Understanding Φ

Φ provides tradeoffs across minimizing the mean-square error for different components of Y.

Simple example: Suppose that $T = 2$ with

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & Z_1 \\ 1 & Z_2 \end{pmatrix}.$$

Then,

$$E^*_\Phi[Y|R] = R\beta = \begin{pmatrix} \beta_1 + \beta_2 Z_1 \\ \beta_1 + \beta_2 Z_2 \end{pmatrix}.$$

Because

$$\beta_1 + \beta_2 Z_t \neq E^*[Y_t|1, Z_t]$$

in general, we need to trade-off the errors across $t = 1, 2$.

Different $\Phi \implies$ different trade-offs across errors $\implies$ different coefficients of the generalized linear predictor

Understanding Φ

There is a special case in which the generalized linear predictor does NOT depend on the choice of the weight matrix, Φ.

Known as **Seemingly Unrelated Regression (SUR)**.

Simple example: Suppose that

$$R = \begin{pmatrix} 1 & Z_1 & Z_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & Z_1 & Z_2 \end{pmatrix}$$

and so,

$$E^*_\Phi[Y|R] = R\beta = \begin{pmatrix} \beta_1 + \beta_2 Z_1 + \beta_3 Z_2 \\ \beta_4 + \beta_5 Z_1 + \beta_6 Z_3 \end{pmatrix}$$

Seemingly Unrelated Regression (SUR)

Define

$$E^*[Y_1|1, Z_1, Z_2] = \gamma_1 + \gamma_2 Z_1 + \gamma_3 Z_3$$
$$E^*[Y_2|1, Z_1, Z_2] = \delta_1 + \delta_2 Z_1 + \delta_3 Z_2.$$

It turns out that the solution to the minimum norm problem

$$\arg\min_c \|Y - Rc\|_\Phi$$

will be

$$(\beta_1, \beta_2, \beta_3) = (\gamma_1, \gamma_2, \gamma_3)$$
$$(\beta_4, \beta_5, \beta_6) = (\delta_1, \delta_2, \delta_3)$$

for any choice of Φ.

Seemingly Unrelated Regression (SUR)

General statement: Define

$$R = \begin{pmatrix} X' & 0 & \dots & 0 \\ 0 & X' & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X' \end{pmatrix} \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix}$$

where X is $q \times 1$ and β_t is $q \times 1$ for $t = 1, \dots, T$.

The generalized linear predictor with weight matrix Φ is

$$E^*_\Phi[Y|R] = \begin{pmatrix} X'\beta_1 \\ X'\beta_2 \\ \vdots \\ X'\beta_M \end{pmatrix}$$

and define the cross-sectional best linear predictors

$$E^*[Y_t|X] = X'\pi_t, \quad \pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_T \end{pmatrix}.$$

Seemingly Unrelated Regression (SUR): Claim

Claim 1: With this choice of R,

$$E^*_{\Phi}[Y|R] = R\pi$$

for any choice of Φ.

Seemingly Unrelated Regression (SUR): Proof

Write

$$Y - Rc = (Y - R\pi) + R(\pi - c).$$

Then, we can write

$$\begin{aligned} E[(Y - Rc)'\Phi(Y - Rc)] &= E[(Y - R\pi)'\Phi(Y - R\pi)] \\ &+ E[(Y - R\pi)'\Phi R](\pi - c) \\ &+ (\pi - c)'E[R'\Phi(Y - R\pi)] \\ &+ (\pi - c)'E[R'\Phi R](\pi - c). \end{aligned}$$

Seemingly Unrelated Regression (SUR): Proof

Next, note that

$$E[(Y - R\pi)'\Phi R] = E[((Y_1 - X'\pi_1)\dots(Y_T - X'\pi_T))\Phi R]$$

We can write

$$((Y_\mathbf{1} - X'\pi_\mathbf{1})\dots(Y_T - X'\pi_T))\Phi = \Big(\sum_{t=\mathbf{1}}^{T}\phi_{t\mathbf{1}}(Y_t - X'\pi_t)\dots\sum_{t=\mathbf{1}}^{T}\phi_{tT}(Y_t - X'\pi_t)\Big)$$

Then,

$$((Y_\mathbf{1} - X'\pi_\mathbf{1})\dots(Y_T - X'\pi_T))\Phi R = ((Y_\mathbf{1} - X'\pi_\mathbf{1})\dots(Y_T - X'\pi_T))\Phi \begin{pmatrix} X' & 0 & \dots & 0 \\ 0 & X' & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X' \end{pmatrix}$$

Seemingly Unrelated Regression (SUR): Proof

So, combining these last two equations, we get that

$$E[(Y - R\pi)'\Phi R] = \Big(\sum_{t=1}^{T}\phi_{t1}E[(Y_t - X'\pi_t)X'] \ldots \sum_{t=1}^{T}\phi_{tT}E[(Y_t - X'\pi_t)X']\Big) = 0$$

where each element by the orthogonality conditions of the cross-sectional best linear predictor.

So, we have that

$$\begin{aligned} E[(Y - Rc)'\Phi(Y - Rc)] = {} & E[(Y - R\pi)'\Phi(Y - R\pi)] \\ & + (\pi - c)'E[R'\Phi R](\pi - c). \end{aligned}$$

and clearly, this norm is minimized at $c = \pi$ because $E[R'\Phi R]$ is positive definite if R has full column rank – which we typically assume. $\square$

Seemingly Unrelated Regression (SUR) – What's the point?

In this case, the generalized linear predictor is equivalent to the equation-by-equation best linear predictor.

So, the equations across t are "seemingly unrelated."

However, inference may need to account for correlation in the error terms across equations!

Seemingly Unrelated Regression (SUR) – Example

Example: Demand Estimation (Wooldridge 2002; Ch. 7)

The system may be a set of demand functions across goods for the population of families.

$$\begin{aligned}
\text{housing} &= \beta_{10} + \beta_{11}\text{house–prc} + \beta_{12}\text{food–prc} + \\
&\quad \beta_{13}\text{clothing–prc} + \beta_{14}\text{inc} + u_1 \\
\text{food} &= \beta_{20} + \beta_{21}\text{house–prc} + \beta_{22}\text{food–prc} + \\
&\quad \beta_{23}\text{clothing–prc} + \beta_{24}\text{inc} + u_2 \\
\text{clothing} &= \beta_{30} + \beta_{31}\text{house–prc} + \beta_{32}\text{food–prc} + \\
&\quad \beta_{33}\text{clothing–prc} + \beta_{34}\text{inc} + u_3.
\end{aligned}$$

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Another interpretation of the GLP coefficients

In the SUR case, GLP equivalent to equation-by-equation BLP.

Suppose we are not in the SUR case, is there a connection between GLP and equation-by-equation BLP?

Yes – GLP coefficients are the “best-approximation” to the equation-by-equation BLP coefficients in a particular sense.

Best Approximation – Set Up

Begin with the $T \times K$ matrix R. Define a vector $q \times 1$ vector X, whose elements span the vector space generated by linear combinations of the elements of R.

- So, $q \leq M \times K$
- Each $R_{ij} = a'X$ for some $a \in \mathbb{R}^q$.
- Fancy (rigorous) way to say: $X = Vec(R)$.

The best linear predictor of Y_t using all elements in R is

$$\hat{Y}_t = E^*[Y_t|X] = X'\pi_t.$$

Best Approximation - Set Up

We can stack them up and write:

$$\hat{Y} = \begin{pmatrix} \hat{Y}_1 \\ \vdots \\ \hat{Y}_T \end{pmatrix} = \begin{pmatrix} X' & 0 & \dots & 0 \\ 0 & X' & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X' \end{pmatrix} \pi$$
$$= (I \otimes X')\pi$$

, where $\pi = (\pi_1, \dots, \pi_T)'$.

$\otimes$ is the **Kroenecker product**.

Aside: Kroenecker Product

The **Kroenecker Product** between matrices A, B is

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1K}B \\ \vdots & & \vdots \\ a_{J1}B & \dots & a_{JK}B \end{pmatrix}$$

So, if A is $J \times K$ and B is $L \times M$, $A \otimes B$ is $K \cdot L \times K \cdot M$.

Properties:

(1): $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

(2): If A, B nonsingular, then

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

(3): $(A \otimes B)' = A' \otimes B'$.

Best Approximation - Set Up

The generalized linear predictor imposes that

$$\tilde{Y}_t = R_{t1}\beta_1 + \ldots + R_{tK}\beta_K.$$

Since each element of R can be written as a linear combination of elements in X ($R_{tj} = X'a_{tj}$), we can write

$$\begin{aligned}\tilde{Y}_t &= X'\sum_{j=1}^{K} a_{tj}\beta_j \\ &= X'A_t\beta,\end{aligned}$$

where we simply collected the a's into A_t

$$\underset{q\times K}{A_t} = (a_{t1} \ldots a_{tK})$$

Best Approximation – Set Up

We can write

$$\begin{aligned} E^*_\Phi[Y|R] &= \begin{pmatrix} \tilde{Y}_1 \\ \vdots \\ Y_T \end{pmatrix} \\ &= (I \otimes X')A\beta \end{aligned}$$

where $(I \otimes X')A = R$ and

$$\underset{M \cdot q \times K}{A} = \begin{pmatrix} A_1 \\ \vdots \\ A_T \end{pmatrix}$$

Best Approximation – Claim

Claim: Let $E^*_\Phi[Y|R] = R\beta$. Then,

$$\beta = \arg\min_{c\in\mathbb{R}^K} (\pi - Ac)'\Big(\Phi \otimes E[XX']\Big)(\pi - Ac).$$

Proof: Same strategy as before. First, write

$$Y - Rc = [Y - (I \otimes X')\pi] + (I \otimes X')(\pi - Ac).$$

and substitute into $E[(Y - Rc)'\Phi(Y - Rc)]$. Show that the cross-terms cancel by an orthogonality argument. $\square$

Best Approximation – Looking ahead

Claim: Let $E^*_\Phi[Y|R] = R\beta$. Then,

$$\beta = \arg\min_{c \in \mathbb{R}^K} (\pi - Ac)' \Big(\Phi \otimes E[XX'] \Big) (\pi - Ac).$$

The generalized linear predictor coefficients is a type of "minimum-distance estimator" for the unrestricted, equation-by-equation BLP coefficients.

> A best K-dimensional approximation to the $K \cdot q$-dimensional equation by equation coefficients.

Outline

Set-Up

Generalized Linear Predictor
- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Consistent Estimation

Observations: Realizations of random variables $D_i = (Y_i, R_i)$ for $i = 1, \ldots, n$. Assume D_i i.i.d from some joint distribution.

The generalized linear predictor is

$$E^*_\Phi[Y_i|R_i] = R_i\beta, \quad \beta = E[R_i'\Phi R_i]^{-1}E[R_i'\Phi Y_i].$$

Sample counterpart: **(Feasible) Generalized Least Squares**

$$\hat{\beta} = \Big(\frac{1}{n}\sum_{i=1}^{n} R_i'\hat{\Phi}R_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} R_i'\hat{\Phi}Y_i\Big)$$

where $\hat{\Phi}$ is an estimator of Φ.

Consistent Estimation

Claim: If $\hat{\Phi} \xrightarrow{p} \Phi$, then

$$\hat{\beta} \xrightarrow{p} \beta$$

as $n \to \infty$.

Proof: See Lecture Note 5. Mechanical application LLN, Slutsky and CMT. Just need to be careful because $\hat{\Phi}$ also depends on the data in principle.

NOTE: This is **fixed-T, large-N** asymptotics!

> Suppose your panel is State $\times$ month. Does this sampling experiment make sense?

Outline

Set-Up

Generalized Linear Predictor

- Definition
- Understanding Φ
- Best Approximation
- Consistent Estimation
- Asymptotic Distribution

Asymptotic Distribution

Based on Wooldridge Ch. 7 here.

Simple Case: Assume Φ is known and will derivate asymptotic distribution.

> This is **generalized least squares**. Deriving the asymptotic distribution for case with $\hat{\Phi}$ is straightforward but just requires additional care.
>
> See Wooldridge Ch. 7 for details.

I will be extra explicit (probably repetitive) in stating our assumptions and how we build up to the asymptotic distribution.

Asymptotic Distribution of Generalized Least Squares

Suppose weight matrix is known. Assumptions we need:

(1): $E[U_i' \Phi R_i] = 0$ – Orthogonality condition.

(2): $E[R_i' \Phi R_i]$ is non-singular.

Under assumptions (1), (2), we can write GLP coefficients $E^*_\Phi[Y_i | R_i] = R_i \beta$ as

$$\beta = E[R_i' \Phi R_i]^{-1} E[R_i' \Phi Y_i].$$

Asymptotic Distribution of Generalized Least Squares

Consider the GLS estimator

$$\hat{\beta} = \left(n^{-1}\sum_{i=1}^{n} R_i'\Phi R_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} R_i'\Phi Y_i\right).$$

Claim: Under assumptions (1), (2):

$$\hat{\beta} \xrightarrow{p} \beta.$$

as $n \to \infty$. <u>Proof</u>: By assumptions, $Y_i = R_i\beta + U_i$. So,

$$\hat{\beta} = \beta + \left(n^{-1}\sum_{i=1}^{n} R_i'\Phi R_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} R_i'\Phi U_i\right)$$

where the second term $\xrightarrow{p} 0$ by assumption (1). $\square$

Asymptotic Distribution of Generalized Least Squares

Now consider

$$\sqrt{n}(\hat{\beta} - \beta) = \left(n^{-1} \sum_{i=1}^{n} R_i' \Phi R_i \right)^{-1} \left(\sqrt{n}^{-1} \sum_{i=1}^{n} R_i' \Phi U_i \right)$$

By assumption (1) plus some additional moment existence assumptions,

$$\sqrt{n}^{-1} \sum_{i=1}^{n} R_i' \Phi U_i \xrightarrow{d} N(0, \Omega)$$

as $n \to \infty$, where

$$\Omega = E[R_i' \Phi u_i u_i' \Phi R_i].$$

Asymptotic Distribution of Generalized Least Squares

Claim: Under assumptions (1), (2)

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \alpha\Omega\alpha'),$$

as $n \to \infty$, where $\alpha = E[R_i'\Phi R_i]^{-1}$ and $\Omega = E[R_i'\Phi U_i U_i'\Phi R_i]$.
Proof: Follows by CMT, Slutsky and CLT.

Estimating the Asymptotic Variance

Then, we have that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, E[R_i' R_i]^{-1} E[R_i' \Phi U_i U_i' \Phi R_i] E[R_i' R_i]^{-1})$$

So,

$$Avar(\hat{\beta}) = n^{-1} E[R_i' R_i]^{-1} E[R_i' \Phi U_i U_i' \Phi R_i] E[R_i' R_i]^{-1}$$

Estimating the Asymptotic Variance

Consistently estimating this is simple. Consider

$$\hat{\alpha} = \left(n^{-1}\sum_{i=1}^{n} R_i' R_i\right)^{-1}, \quad \hat{\Omega} = n^{-1}\sum_{i=1}^{n} R_i' \Phi \hat{U}_i \hat{U}_i' \Phi R_i$$

And form

$$A\hat{var}(\hat{\beta}) = n^{-1}\left(n^{-1}\sum_{i=1}^{n} R_i' R_i\right)^{-1}\left(n^{-1}\sum_{i=1}^{n} R_i' \Phi \hat{U}_i \hat{U}_i' \Phi R_i\right)\left(n^{-1}\sum_{i=1}^{n} R_i' R_i\right)^{-}$$

$$= \left(\sum_{i=1}^{n} R_i' R_i\right)^{-1}\left(\sum_{i=1}^{n} R_i' \Phi \hat{U}_i \hat{U}_i' \Phi R_i\right)\left(\sum_{i=1}^{n} R_i' R_i\right)^{-1}$$

This variance estimator is a **robust variance matrix estimator** – it allows for arbitrary cross-equation correlation. It is also heteroskedasticity robust.

AGAIN NOTE: The asymptotics are fixed-T, large-N.

> We are describing the behavior of this estimator as the number of cross-sectional units grows large for a fixed number of periods.
>
> There are plenty of cases where this may not be a good guide for the behavior of the estimator – Bertrand, Duflo, Mullainathan (2004) & Hansen (2007).