Statistics 215B Mr. Freedman 15 April 2000

Hierarchical Linear Regression

The following hierarchical linear model was the basis for smoothing in the proposed 1990 census adjustment (Freedman et al, 1993):

(1a) $$Y = \gamma + \delta$$

(1b) $$\gamma = X\beta + \epsilon.$$

Here, X may be viewed as a fixed $n \times p$ matrix; δ and ϵ are assumed to be independent $n \times 1$ vectors; the means are 0; $\text{cov}(\delta) = K$ is a given positive definite matrix; $\text{cov}(\epsilon) = \sigma^2 I$, where σ^2 is given, being positive and finite; I is the $n \times n$ identity matrix. In the census context, Y would be the vector of "raw" adjustment factors and γ the "true" adjustment factors, with one component of the vector for each post stratum in a region. Theorem (3) below would not apply: σ^2 and K have to be estimated from the data, X is chosen by a data-based algorithm, etc. For more discussion of the model in the census context, see Freedman et al (1993).

Goldberger (1962) proposed the estimator $\hat{\gamma}$ for γ. His construction involves an auxiliary matrix Γ; to define the latter, let H be the OLS projection matrix, that is, $H = X(X'X)^{-1}X'$. Now

(2a) $$\Gamma^{-1} = K^{-1} + \sigma^{-2}(I - H)$$

(2b) $$\hat{\gamma} = \Gamma K^{-1} Y.$$

The estimator $\hat{\gamma}$ can be motivated in the normal case as Bayes with a diffuse prior on β; then Γ is the posterior covariance of γ given Y; see Lindley and Smith (JRSS, 1972). Being positive definite, $K^{-1} + \sigma^{-2}(I - H)$ is invertible. Hence, Γ is well-defined and positive definite. Moreover, Γ is symmetric because K and H are symmetric.

Notes. Being a covariance matrix, K is symmetric; so is Γ; but ΓK^{-1} is asymmetric. The estimator $\hat{\gamma}$ can also be represented as a generalized average of Y and gls; see (28). If U, V are $n \times 1$ and $m \times 1$, respectively, and $\text{E}(U) = \text{E}(V) = 0$, then $\text{cov}(U, V)$ is defined as $\text{E}(UV')$.

(3) Theorem (Goldberger).

(a) $\text{E}(\hat{\gamma} - \gamma) = 0$

(b) $\text{cov}(\hat{\gamma} - \gamma) = \Gamma$

(c) $\hat{\gamma}$ is the minimum-variance unbiased linear estimator of γ.

The object here is to prove (3). Claim (c) can be translated as follows. If M is an $n \times n$ matrix and $\text{E}(MY - \gamma) = 0$, then $\text{cov}(MY - \gamma) - \Gamma$ is non-negative definite, vanishing only if $M = \Gamma K^{-1}$.

Warning. $\text{cov}(\hat{\gamma}) \neq \Gamma$ for the objectivist, because γ is random.

(4) Lemma. $\Gamma K^{-1} X = X$ so $\Gamma K^{-1} H = H$.

Proof. Let x be in the column space of X. Then

$$\Gamma^{-1}x = K^{-1}x;$$

indeed, by (2a) the difference is $\sigma^{-2}(I - H)x = 0$. Pre-multiply the display by Γ. QED

Claim (a) in the theorem is immediate: $\mathrm{E}(\gamma) = X\beta = \mathrm{E}(\hat{\gamma})$, the last equality holding by (4).

(5) Lemma. If M is an $n \times n$ matrix and $\mathrm{E}(MY - \gamma) = 0$, then $MX = X$ and $MH = H$.

Proof. Clearly, $(MX - X)\beta = 0$ for all β. QED

(6) Lemma. If M is an $n \times n$ matrix and $MX = X$, then

$$\mathrm{cov}(MY - \gamma) = MKM' + \sigma^2(M - I)(M' - I).$$

Proof. Clearly,

$$MY - \gamma = M\delta + (M - I)\epsilon \tag{7}$$

because $MX = X$. Now use the assumptions on δ and ϵ. QED

Use (7) with $M = \Gamma K^{-1}$ to see

$$\hat{\gamma} - \gamma = \Gamma K^{-1}\delta + (\Gamma K^{-1} - I)\epsilon. \tag{8}$$

Claim (3b) follows from (6) applied to $M = \Gamma K^{-1}$ and the identity

$$\Gamma K^{-1}\Gamma + \sigma^2(\Gamma K^{-1} - I)(K^{-1}\Gamma - I) = \Gamma. \tag{9}$$

To prove (9), multiply from the right by Γ^{-1}; use definition (2a) to evaluate $K^{-1} - \Gamma^{-1}$ and (4) to simplify the results.

For claim (3c), let $\zeta = MY - \hat{\gamma}$, so that

$$MY - \gamma = (\hat{\gamma} - \gamma) + \zeta. \tag{10}$$

It suffices to show that $MX = X$ implies

$$\mathrm{cov}(\zeta, \hat{\gamma} - \gamma) = 0 \tag{11}$$

because then $\mathrm{cov}(MY - \gamma) = \mathrm{cov}(\hat{\gamma} - \gamma) + \mathrm{cov}(\zeta)$. Clearly,

$$\zeta = (M - \Gamma K^{-1})(\delta + \epsilon). \tag{12}$$

By (8) and (12),

$$\text{(13)} \qquad \mathrm{cov}(\zeta, \hat{\gamma} - \gamma) = \mathrm{E}\{\zeta(\hat{\gamma} - \gamma)'\} = (M - \Gamma K^{-1})\Gamma + (M - \Gamma K^{-1})(K^{-1}\Gamma - I)\sigma^2.$$

The right side of (13) is

$$\text{(14)} \qquad (M - \Gamma K^{-1})\Big[\Gamma + \sigma^2(K^{-1}\Gamma - I)\Big].$$

By the definition (2a) of Γ,

$$\Big[K^{-1} + \sigma^{-2}(I - H)\Big]\Gamma = I,$$

so

$$K^{-1}\Gamma = I - \sigma^{-2}(I - H)\Gamma$$

and

$$\text{(15)} \qquad \sigma^2(K^{-1}\Gamma - I) = -(I - H)\Gamma.$$

The identity (15) can be used to evaluate the expression (14) as

$$(M - \Gamma K^{-1})(\Gamma - \Gamma + H\Gamma) = \Big[(M - \Gamma K^{-1})H\Big]\Gamma = 0$$

by (4) and (5). This proves (11), hence, (3). QED

Notes. (i) The proof of (9) could be rearranged slightly to use (15).

(ii) For uniqueness, $\mathrm{cov}(\zeta) = 0$ means that $MY = \hat{\gamma}$ almost surely, and then $M = \Gamma K^{-1}$ rather easily.

Discussion

Goldberger's estimate is in the shrinkage—empirical Bayes style; Y is shrunk toward the column space C of X. This works fine if C is low-dimensional and γ is almost in there, i.e., σ^2 is small. The amount of shrinking and the directions are controlled by σ^2 and K. If you get these wrong, or some of the modeling assumptions break down, shrinking can actually make the errors bigger rather than smaller. Also, the benefits of shrinking—as gauged by Γ—depend rather critically on the assumptions about the error terms δ and ϵ. To sum up: If the model is wrong, the benefits of shrinking can be over-stated, and shrinking can be counter-productive. For some empirical evidence, see (Freedman and Navidi, 1986) or (Freedman et al, 1993).

Conditional Normal Distributions

Let U and V be jointly normal vectors, $n \times 1$ and $m \times 1$, respectively; both have mean 0. Suppose V is of full rank. By definition, $\mathrm{cov}(U, V) = \mathrm{E}(UV')$ and $\mathrm{cov}(V) = \mathrm{cov}(V, V)$. Let $M = \mathrm{cov}(U, V)\mathrm{cov}(V)^{-1}$; this is an $n \times m$ matrix, well-defined because V has full rank.

(16) Proposition.

(a) $E\{U|V\} = MV$.

(b) $\text{cov}\{U|V\} = \text{cov}(U) - \text{cov}(U, V)\text{cov}(V)^{-1}\text{cov}(V, U)$.

(c) The conditional distribution of U is normal.

Proof. Let $\zeta = U - MV$. Clearly, U, V, and ζ are jointly normal with mean 0; furthermore, $\text{cov}(\zeta, V) = 0$ and $\text{cov}(\zeta)$ is given by the right hand side of (b). Thus, ζ and V are independent; given $V = v$, U is distributed as $Mv + \zeta$. QED

Notes. (i) It is normality that converts $\text{cov}(\zeta, V) = 0$ to independence. To argue the independence, you need to write down the density and factor it.

(ii) The right hand side in (b) is "total variance − explained variance."

(17) Lemma. Let A and B be $n \times n$ matrices; suppose $A + B$ is invertible.

(a) $A(A + B)^{-1} = I_n - B(A + B)^{-1}$.

(b) $(A + B)^{-1}A = I_n - (A + B)^{-1}B$.

(c) $A(A + B)^{-1}B = B(A + B)^{-1}A$.

(d) Suppose A is invertible and $A^{-1}B = BA^{-1}$. Then

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

Proof. For claim (a), $I_n = (A+B)(A+B)^{-1} = A(A+B)^{-1} + B(A+B)^{-1}$, and (b) is the same. For (c), use (a) and (b):

$$\begin{aligned} A(A + B)^{-1}B &= A\Big[(A + B)^{-1}B\Big] \\ &= A\Big[I_n - (A + B)^{-1}A\Big] \\ &= A - A(A + B)^{-1}A \\ &= \Big[I_n - A(A + B)^{-1}\Big]A \\ &= B(A + B)^{-1}A. \end{aligned}$$

For (d),

$$A(A + B)^{-1} = \left(I_n + \frac{B}{A}\right)^{-1} = (A + B)^{-1}A. \quad \text{QED}$$

(18) Proposition. Let ζ and η be independent mean 0 normal $n \times 1$ vectors with respective covariance matrices C and D. Then

(a) $E\{\zeta|\zeta + \eta\} = C(C + D)^{-1}(\zeta + \eta)$

(b) $\text{cov}\{\zeta|\zeta + \eta\} = C(C + D)^{-1}D$.

Proof. This follows from (16). Indeed, $\text{cov}(\zeta, \zeta + \eta) = \text{cov}(\zeta)$ so $M = C(C + D)^{-1}$. And $\text{cov}\{\zeta | \zeta + \eta\}$ is given by (17b) as

$$C - C(C + D)^{-1}C = C\Big(I_n - (C + D)^{-1}C\Big) = C(C + D)^{-1}D. \quad \text{QED}$$

Bayesian Least Squares

Consider the regression model

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I_n) \tag{19}$$

where Y is $n \times 1$, X is $n \times p$ of full rank, β is $p \times 1$, and ϵ is $n \times 1$. For the moment, β is unknown but σ is known. Take X to be constant (non-random). For the Bayesian, β has a prior distribution—and a posterior given the data Y.

(20) Proposition. With the OLS regression model (19), and $\beta \sim N(0, \tau^2 I_p)$ independent of ϵ, the posterior distribution of β given Y is normal, with conditional mean

$$\hat{\beta}_{\text{Bayes}} = \Big(I_p - \frac{\sigma^2}{\tau^2 X'X + \sigma^2 I_p}\Big)\hat{\beta}_{\text{ols}}$$

and conditional variance

$$\text{cov}\{\beta | Y\} = \frac{\tau^2\sigma^2}{\tau^2 X'X + \sigma^2 I_p}.$$

Proof. Since $X'Y$ is sufficient, it is enough to compute the conditional law of β given $X'Y$. This can be done using (16) with $U = \beta$ and $V = X'Y$. For the Bayesian, $\text{cov}(X'Y) = \tau^2(X'X)^2 + \sigma^2 X'X$ while $\text{cov}(\beta, X'Y) = \tau^2(X'X)$. So

$$\begin{aligned} M &= \frac{\tau^2 X'X}{\tau^2(X'X)^2 + \sigma^2 X'X} \\ &= \frac{\tau^2 X'X}{\tau^2 X'X + \sigma^2 I_p}(X'X)^{-1} \\ &= \Big(I_p - \frac{\sigma^2}{\tau^2 X'X + \sigma^2 I_p}\Big)(X'X)^{-1} \end{aligned}$$

and

$$MX'Y = \Big(I_p - \frac{\sigma^2}{\tau^2 X'X + \sigma^2 I_p}\Big)\hat{\beta}_{\text{ols}}.$$

The matrices commute, so the informal notation is unambiguous. The assertion about the conditional mean is now proved, and the conditional variance follows from (16b) by a straightforward calculation. QED

Notes. (i) With lots of data (and a little luck), $X'X$ is large. Then $\hat{\beta}_{\text{Bayes}} \approx \hat{\beta}_{\text{ols}} = \hat{\beta}_{\text{mle}}$ and $\text{cov}\{\beta|Y\} \approx \sigma^2(X'X)^{-1}$, the frequentist covariance of $\hat{\beta}_{\text{mle}}$. This is a special case of the Bernstein-von Mises theorem: under suitable regularity conditions, asymptotically, the posterior distribution of $\beta - \hat{\beta}_{\text{mle}}$ is close to the frequentist distribution of $\hat{\beta}_{\text{mle}} - \beta_{\text{true}}$.

(ii) Suppose σ^2 is fixed and τ^2 is large, so the prior is "diffuse" or "uninformative" or an "ignorance prior." (An uninformative prior could be defined as Lebesgue measure on I_p.) Then the Bayes estimate is essentially the same as the OLS estimate.

(iii) Suppose τ^2 is fixed and σ^2 is large, so the data are "uninformative." Then the posterior is essentially the same as the prior. In the extreme, if $\tau^2 = 0$, you start by knowing that $\beta = 0$, that is how you end up.

(iv) The proof of (20) can be based on (18) with $\zeta = X'X\beta$ and $\eta = X'\epsilon$.

(21) Corollary. In (19), suppose $\epsilon \sim N(0, K)$ where K is $n \times n$ positive definite and given. With this GLS model, the posterior distribution of β given Y is normal, with conditional mean

$$\hat{\beta}_{\text{Bayes}} = \Big(I_p - (\tau^2 X'K^{-1}X + I_p)^{-1}\Big)\hat{\beta}_{\text{gls}}$$

and conditional variance

$$\text{cov}\{\beta|Y\} = \tau^2(\tau^2 X'K^{-1}X + I_p)^{-1}.$$

Proof. Multiplication from the left by $K^{-1/2}$ converts the GLS model to OLS, with $\sigma^2 = 1$ and design matrix $K^{-1/2}X$. QED

An Identity

Of course, with the GLS regression model, it is possible to compute the posterior mean directly from (16), as $\text{E}(\beta|Y) = \tau^2 X'(\tau^2 XX' + K)^{-1}$. Thus we have an indirect proof of the well-known identity

$$X'(\tau^2 XX' + K)^{-1} = (\tau^2 X'K^{-1}X + I_p)^{-1}X'K^{-1}. \tag{22}$$

This can be proved directly: multiply from the left by

$$\tau^2 X'K^{-1}X + I_p$$

and from the right by $\tau^2 XX' + K$; then clean up.

Note. XX' is singular, so the behavior of $(\tau^2 XX' + K)^{-1}$ on the left in (22) for large τ^2 is problematic. On the right, $X'K^{-1}X$ is invertible.

Bayes and Goldberger

Consider the model (1) from a Bayesian perspective: we assume K and σ^2 are known, and put a prior on the "hyper-parameters" β, namely, $\beta \sim N(0, \tau^2 I_p)$, independent of δ and ϵ.

(23) Theorem. Consider the model (1) with prior distribution $\beta \sim N(0, \tau^2 I_p)$ independent of δ and ϵ. Let $\tau \to \infty$. The Bayes estimate for γ converges to $\hat{\gamma}$, and the posterior variance converges to Γ.

The proof is a bit involved. As before, we can use (16) to compute the posterior law of γ given Y. This is best done in three steps:

Step 1. Compute the posterior law of β given Y. We have a GLS model with covariance matrix $\Sigma = K + \sigma^2 I_n$. The posterior distribution is by (16) multivariate normal with conditional mean

$$\hat{\beta}_{\text{Bayes}} = \Big(I_p - (\tau^2 X'\Sigma^{-1}X + I_p)^{-1}\Big)\hat{\beta}_{\text{gls}} \tag{24}$$

and conditional variance

$$\text{cov}\{\beta|Y\} = \tau^2(\tau^2 X'\Sigma^{-1}X + I_p)^{-1}. \tag{25}$$

Step 2. Compute the conditional law of γ given β and Y; this is done below, also using (16). Call the conditional density $f(\gamma|\beta, y)$. Of course, f is normal.

Step 3. Integrate out β in $f(\cdot|\beta, y)$ from Step 2 using the posterior law of β from Step 1. This too is done below.

Step 2 can be implemented as follows: β and Y contain the same information (i.e., span the same σ-field) as β and $\delta + \epsilon$. Recall that $\gamma = X\beta + \epsilon$. Let $\Sigma = \sigma^2 I_p + K$. By (18),

$$\begin{aligned} \text{E}\{\gamma|\beta, Y\} &= \text{E}\{\gamma|\beta, \delta + \epsilon\} \\ &= X\beta + \text{E}\{\epsilon|\beta, \delta + \epsilon\} \\ &= X\beta + \text{E}\{\epsilon|\delta + \epsilon\} \qquad \text{because } \beta \perp (\delta, \epsilon) \\ &= X\beta + \sigma^2\Sigma^{-1}(\delta + \epsilon) \\ &= X\beta + \sigma^2\Sigma^{-1}(Y - X\beta) \\ &= \sigma^2\Sigma^{-1}Y + K\Sigma^{-1}X\beta \qquad \text{by (17a)}. \end{aligned} \tag{26}$$

Likewise,

$$\text{cov}\{\gamma|\beta, Y\} = \sigma^2\Sigma^{-1}K. \tag{27}$$

Note. Step 2 is conditional on β and results do not involve τ^2.

Step 3 is done by the following computation:

$$\begin{aligned} \text{E}\{\gamma|Y\} &= \sigma^2\Sigma^{-1}Y + K\Sigma^{-1}X\text{E}\{\beta|Y\} \\ &= \sigma^2\Sigma^{-1}Y + K\Sigma^{-1}X\hat{\beta}_{\text{Bayes}} \\ &= \sigma^2\Sigma^{-1}Y + K\Sigma^{-1}X\hat{\beta}_{\text{gls}} - \Delta \\ &= \sigma^2\Sigma^{-1}Y + K\Sigma^{-1}\hat{Y}_{\text{gls}} - \Delta \end{aligned} \tag{28}$$

where

(29) $$\Delta = K\Sigma^{-1}X(\tau^2 X'\Sigma^{-1}X + I_p)^{-1}\hat{\beta}_{\text{gls}}.$$

For τ^2 large the term Δ is negligible, presenting $\hat{\gamma}$ as a mixture—with matrix weights—of Y and the GLS projection onto the column space of X.

Now $\text{cov}(\gamma|Y)$ may be computed by integrating out β but holding Y fixed:

(30) $$\text{E}\{\text{cov}(\gamma|\beta, Y) \bigm| Y\} + \text{cov}\{\text{E}(\gamma|\beta, Y) \bigm| Y\}.$$

In the first term of (30), $\text{cov}(\gamma|\beta, Y) = \sigma^2 K\Sigma^{-1}$ and is constant, see (27). In the second term, $\text{E}(\gamma|\beta, Y)$ is by (26) equal to

$$\sigma^2\Sigma^{-1}Y + K\Sigma^{-1}X\beta,$$

whose covariance given Y is

(31) $$K\Sigma^{-1}X\text{cov}\{\beta|Y\}X'\Sigma^{-1}K.$$

But $\text{cov}\{\beta|Y\}$ was computed in Step 1, see (25); and (31) is

$$K\Sigma^{-1}X\Big[\tau^2(\tau^2 X'\Sigma^{-1}X + I_p)^{-1}\Big]X'\Sigma^{-1}K \to K\Sigma^{-1}H_\Sigma K$$

as $\tau^2 \to \infty$. (By definition, $\Sigma = \sigma^2 I_n + K$, and $H_\Sigma = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$ is the GLS projection matrix relative to Σ.) Thus,

(32) $$\text{cov}\{\gamma|Y\} = \sigma^2\Sigma^{-1}K + K\Sigma^{-1}H_\Sigma K.$$

Note: We have to use $\text{cov}(\beta|Y)$ not $\text{cov}(\beta)$ in (31).

(33) Lemma. Let $\Sigma = \sigma^2 I_n + K$ and $H_\Sigma = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$.

(a) $X'\Sigma^{-1}(I_n - H_\Sigma) = 0$.

(b) $H\Sigma^{-1}(I_n - H_\Sigma) = 0$.

(c) $(I_n - H)H_\Sigma = 0$.

(d) $\Gamma K^{-1} = H_\Sigma + \sigma^2\Sigma^{-1}(I_n - H_\Sigma)$.

(e) $\Gamma K^{-1} = \sigma^2\Sigma^{-1} + K\Sigma^{-1}H_\Sigma$.

(f) $\Gamma = \sigma^2\Sigma^{-1}K + K\Sigma^{-1}H_\Sigma K$.

Proof. Claim (a) is easy and (b) follows on multiplying from the left by $X(X'X)^{-1}$. Claim (c) is easy. For (d), by (2a),

(34) $$K\Gamma^{-1} = I_n + \sigma^{-2}K(I_n - H).$$

We have to prove

(35) $$\Big[I_n + \sigma^{-2}K(I_n - H)\Big]\Big[H_\Sigma + \sigma^2\Sigma^{-1}(I_n - H_\Sigma)\Big] = I_n.$$

By (c), the left side of (35) is

$$
\begin{aligned}
H_\Sigma + \sigma^2\Sigma^{-1}(I_n - H_\Sigma) + K\Sigma^{-1}(I_n - H_\Sigma) \\
&= H_\Sigma + (\sigma^2 I_n + K)\Sigma^{-1}(I_n - H_\Sigma) \\
&= H_\Sigma + \Sigma\Sigma^{-1}(I_n - H_\Sigma) = I_n.
\end{aligned}
$$

This proves (35) and hence (d). Claim (e) follows from (d): the difference between the two right hand sides is 0, as one sees by collecting terms. For (f), multiply (e) on the right by K. QED

(36) Corollary. Let $\tau \to \infty$.

(a) $\mathrm{E}\{\gamma|Y\} \to \sigma^2(\sigma^2 I_p + K)^{-1}Y + K(\sigma^2 I_p + K)^{-1}\hat{Y}_{\text{gls}} = \Gamma K^{-1}Y$.

(b) $\mathrm{cov}\{\gamma|Y\} \to \Gamma$.

Proof. Claim (a). Convergence follows from (28), because $\Delta \to 0$. Then use (33e). Claim (b) is immediate from (32) and (33f). QED

This completes our first proof of (23).

Bayes, Goldberger and the Identity

Here is a more direct approach, with the same model and prior as in the previous section. We use (18) with $\zeta = \gamma$ and $\eta = \delta$. For the Bayesian, these are independent normal vectors with mean 0; furthermore,

(37) $$C = \mathrm{cov}(\zeta) = \mathrm{cov}(\gamma) = \sigma^2 I_n + \tau^2 XX'$$

(38) $$D = \mathrm{cov}(\eta) = \mathrm{cov}(\delta) = K.$$

Thus, $\mathrm{E}\{\gamma|Y\} = MY$ where $M = C(C + D)^{-1}$ so

(39) $$M^{-1} = I_n + \sigma^{-2}K(I_n + \lambda XX')^{-1}$$

with $\lambda = \tau^2/\sigma^2$. We want $M^{-1} \to K\Gamma^{-1}$ as $\lambda \to \infty$, that is,

(40) $$(I_n + \lambda XX')^{-1} \to I - H.$$

Of course, if $x \perp X$, then $(I_n + \lambda XX')x = x$ so $(I_n + \lambda XX')^{-1}x = x = (I - H)x$. Suppose now that x is in the column space of X, that is, $x = Xc$ where c is $p \times 1$. We must show

(41) $$(I_n + \lambda XX')^{-1}x \to 0.$$

To prove (41), we use (22) twice, with $K = I_n$ and $\tau^2 = 1/\lambda$:

$$
\begin{aligned}
\|(I_n + \lambda XX')^{-1}Xc\|^2 &= c'X'(I_n + \lambda XX')^{-2}Xc \\
&= c'(I_p + \lambda X'X)^{-1}X'(I_n + \lambda XX')^{-1}Xc \\
&= c'(I_p + \lambda X'X)^{-2}X'Xc \\
&= \lambda^{-2}c'(\lambda^{-2}I_p + X'X)^{-2}X'Xc \to 0.
\end{aligned}
$$

The point of rigor: the function $A \to A^{-2}$ is continuous at $X'X$ not at XX'.

The covariance of γ given Y follows from (18b), indeed

$$\text{cov}\{\beta|Y\} = C(C + D)^{-1}D = MK;$$

But $M \to \Gamma K^{-1}$ as $\tau \to \infty$. This completes our second proof of (23).

References

D. Freedman and W. Navidi (1986). Regression models for adjusting the 1980 census. *Statistical Science* 1: 1–39 (with discussion).

D. Freedman, K. Wachter, D. Coster, R. Cutler and S. Klein (1993). Adjusting the census of 1990: the smoothing model. *Evaluation Review* 17: 371–444.

A. Goldberger (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57: 369–75.

D. V. Lindley and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society* 67: 1–19.