1 Proof of Theorem 1: General Class Size

Proof. The setup of the problem and the structure of the proof for the general class size case mimics the roommate case illustrated in Theorem 1. We continue to assume a homogeneous peer effect and consider the limiting case where

1. We observe students for at most two time periods.

2. Within each class there is at most one student that is observed for two periods. All other students are observed for only one time period.

Remark 1: Clearly if the estimator is consistent for $T = 2$, it is also consistent for $T > 2$. The second simplification is equivalent to allowing all of the individual effects in a class but one to vary over time. For example, suppose class size was fixed at $M + 1$ and there were $(M + 1)\mathcal{N}$ students observed for two periods, implying that $(M + 1)\mathcal{N}$ individual effects would be estimated. We could, however, allow the individual effects to vary over time for all students but one in each group, making sure to choose these students in such a way that they are matched with someone in both periods whose individual effect does not vary over time.[1] $(2M + 1)\mathcal{N}$ individual effects would then be estimated. Having M individuals whose effect varies over time is equivalent to estimating $2M$ individual effects—it is the same as having two sets of M individuals who are each observed once. If the estimator is consistent in this case, then it is also consistent under the restricted case when all of the individual effects are time invariant (fixed effects).

[1] To see how these assignments work, consider a two period model where the groups in period 1 are $\{A, B, C\}$ and $\{D, E, F\}$ and the groups in period 2 are $\{A, B, F\}$ and $\{D, E, C\}$. We could let the individual effects for $\{B, C, E, F\}$ vary over time. Each group in each time period will have one student observed twice and one student observed once. The number of individual effects would then increase from six to ten. More generally, with a common class size of $M + 1$, the most severe overlap that still allows variation in the peer group is to have M individuals in each class remain together in both periods. In this case, we could allow all individual effects to vary over time except for one of the individual effects of the M individuals in each class that stay together in both periods. Things become more complicated when class size is not constant, but allowing all individual effects to vary over time except for a set of individuals who never share a class will grow linearly in $\mathcal{N}$. Hence, while the asymptotic variance would be affected, identification, consistency, and asymptotic normality are unaffected.

Consider the set of students that are observed for two time periods. Each of these students has M peers in period one and M peers in period two. Denote a student block as one student observed for two periods plus his $2M$ peers. There are then $\mathcal{N}$ blocks of students, one block for each student observed twice. Denote the first student in each block as the student who is observed twice, where α_{1n} is the individual effect. For ease of exposition we will also write α_{1n} as α_{11n} or α_{12n}. The time subscripts are irrelevant here since time does not indicate a different individual. The individual effect for the ith classmate in block n at time period t is α_{itn}, where $i \geq 2$. For these individuals the time subscript is relevant for identifying each individual.

Remark 2: If the estimator is consistent when the class size is fixed at any value of M, then it will also be consistent when M varies as we could split the sample into those who had class sizes of M, estimating the parameters just on this group. Imposing the restriction that the peer effects parameter is the same across different class sizes will not affect whether the estimator is consistent.

The optimization problem is then:

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left[\left(y_{11n} - \alpha_{1n} - \frac{\gamma}{M} \sum_{j=2}^{M+1} \alpha_{j1n} \right)^2 + \left(y_{12n} - \alpha_{1n} - \frac{\gamma}{M} \sum_{j=2}^{M+1} \alpha_{j2n} \right)^2 \right.$$
$$\left. + \sum_{i=2}^{M+1} \left(y_{i1n} - \alpha_{i1n} - \frac{\gamma}{M} \sum_{j \neq i}^{M+1} \alpha_{j1n} \right)^2 + \sum_{i=2}^{M+1} \left(y_{i2n} - \alpha_{i2n} - \frac{\gamma}{M} \sum_{j \neq i}^{M+1} \alpha_{j2n} \right)^2 \right] \tag{1}$$

Within each block there are four terms, two residuals for the student observed twice, and peer residuals in time period one and two.

Again, conditional on γ, the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Hence, we are able to focus on individual blocks in isolation from one another when concentrating out the α's as a function of γ.

Our proof in the general class size case then consists of the following five lemmas, each of which is proven later in this appendix.

Lemma 1.G

The vector of unobserved student abilities, α, can be concentrated out of the least squares prob-

lem and written strictly as a function of γ and y.

Due to the complexity of these expressions we only provide them in the following proof.

We then show the form of the minimization problem when the α's are concentrated out.

Lemma 2.G

Concentrating the α's out of the original least squares problem results in an optimization problem over γ that takes the following form:

$$\min_{\gamma} \sum_{n=1}^{\mathcal{N}} \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jtn} y_{jtn}\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jtn}^2}$$

where

$$\begin{aligned} W_{11n} &= M + \gamma(M-1) = -W_{12n} \\ W_{j1n} &= -\gamma = -W_{j2n} \ \forall \ j > 1 \end{aligned} \tag{2}$$

Our nonlinear least squares problem has only one parameter, γ. We are now in a position to investigate the properties of our estimator of γ_o. For ease of notation, define $q(w,\gamma)$ as:

$$q(w,\gamma) = \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt} y_{jt}\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}$$

where $w \equiv y$. We let $\mathcal{W}$ denote the subset of $\mathbb{R}^{2+2M}$ representing the possible values of w.

Our key result is then Lemma 3.G, which establishes identification.

Lemma 3.G

$$E[q(w,\gamma_o)] < E[q(w,\gamma)], \quad \forall\ \gamma \in \Gamma, \;\; \gamma \neq \gamma_o$$

Theorem 12.2 of Wooldridge (2002) establishes that sufficient conditions for consistency are identification and uniform convergence. Having already established identification, Lemma 4 shows uniform convergence.

Lemma 4.G

$$\max_{\gamma\in\Gamma} \left| \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} q(w_n,\gamma) - E[q(w,\gamma)] \right| \xrightarrow{p} 0$$

Consistency then follows from Theorem 12.2 of Wooldridge: $\hat{\gamma} \xrightarrow{p} \gamma_o$.

Finally, we establish asymptotic normality of $\hat{\gamma}$. Denote $s(w,\gamma_o)$ and $H(w,\gamma_o)$ as the first and second derivative of $q(w,\gamma)$ evaluated at γ_o. Then, Lemma 5 completes the proof.

Lemma 5.G

$$\sqrt{\mathcal{N}}(\hat{\gamma} - \gamma_o) \xrightarrow{d} N(0, A_o^{-1} B_o A_o^{-1})$$

where

$$A_o \equiv E[H(w,\gamma_o)]$$

and

$$B_o \equiv E[s(w,\gamma_o)^2] = Var[s(w,\gamma_o)]$$

QED.

Proof of Lemma 1.G

Our objective is to show, using matrix algebra, that the system of equations obtained by differentiating Equation (1) with respect to α can be expressed as a series of equations in terms of γ, M, and y. First, differentiate the original least squares problem with respect to each α_i. The first-order condition for any α_{1n} (first student in each block who is observed in both time periods) is given by:

$$\begin{aligned} 0 &= -2\left[\left(y_{11n} - \alpha_{1n} - \frac{\gamma}{M}\sum_{j=2}^{M+1}\alpha_{j1n}\right) + \left(y_{12n} - \alpha_{1n} - \frac{\gamma}{M}\sum_{j=2}^{M+1}\alpha_{j2n}\right)\right] \\ &\quad -\frac{2\gamma}{M}\sum_{t=1}^{2}\sum_{i=2}^{M+1}\left(y_{itn} - \alpha_{itn} - \frac{\gamma}{M}\sum_{j\neq i}^{M+1}\alpha_{jtn}\right) \end{aligned}$$

while the first-order condition for any α_{itn} (applicable to all students except the first in each class) is given by:

$$\begin{aligned} 0 &= -\frac{2\gamma}{M}\left[\left(y_{1tn} - \alpha_{1n} - \frac{\gamma}{M}\sum_{j=2}^{M+1}\alpha_{jtn}\right) + \sum_{j=2\,,\,j\neq i}^{M+1}\left(y_{jtn} - \alpha_{jtn} - \frac{\gamma}{M}\sum_{k\neq j}^{M+1}\alpha_{ktn}\right)\right] \\ &\quad -2\left(y_{itn} - \alpha_{itn} - \frac{\gamma}{M}\sum_{j\neq i}^{M+1}\alpha_{jtn}\right) \end{aligned}$$

This yields a system of $\mathcal{N}(2M+1)$ equations and $\mathcal{N}(2M+1)$ unknown abilities.

We can re-arrange the above first-order conditions such that all the parameters to be estimated (α's and γ) are on the left and all the observed grades (y) are on the right. Doing this for the first-order conditions derived for α_{1n} and α_{itn} yields the following two equations

$$\Big(2 + \frac{2\gamma^2}{M}\Big)\alpha_{1n} + \Big(\frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}\Big)\Big(\sum_{t=1}^{2}\sum_{j=2}^{M+1}\alpha_{jtn}\Big) = y_{11n} + y_{12n} + \frac{\gamma}{M}\sum_{t=1}^{2}\sum_{j=2}^{M+1}y_{jtn}$$

and

$$\Big(1 + \frac{\gamma^2}{M}\Big)\alpha_{itn} + \Big(\frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}\Big)\Big(\alpha_{1n} + \sum_{j=2,\,j\neq i}^{M+1}\alpha_{jtn}\Big) = y_{itn} + \frac{\gamma}{M}(y_{1tn} + \sum_{j=2,\,j\neq i}^{M+1}y_{jtn})$$

We can write this system of equations in matrix form such that $\tilde{X}\alpha = \tilde{Y}$, where α is simply a $((2M\mathcal{N} + \mathcal{N})$ x $1)$ vector of all the individual student abilities. $\tilde{X}$ is given by the following

$$\tilde{\mathbf{X}}_{(N*(2M+1)\,\mathrm{x}\,N*(2M+1))} = \begin{bmatrix} A & B & 0 & 0 & 0 & 0 & \dots \\ C & D & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & A & B & 0 & 0 & \dots \\ 0 & 0 & C & D & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & A & B & \\ 0 & 0 & 0 & 0 & C & D & \\ \vdots & \vdots & \vdots & \vdots & & & \ddots \end{bmatrix}$$

The sub-components of $\tilde{X}$ are defined below:

$$\mathbf{A}_{(1\mathrm{x}1)} = 2 + \frac{2\gamma^2}{M}$$

$$\mathbf{B}_{(1\mathrm{x}2M)} = \left[\frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}, \ \dots \ , \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}\right]$$

$$\mathbf{C}_{(2M\mathrm{x}1)} = \left[\frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}, \ \dots \ , \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2}\right]'$$

$$\mathbf{D}_{(2M\mathrm{x}2M)} = \begin{bmatrix} 1+\frac{\gamma^2}{M} & \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2} & \dots & 0 & 0 & \dots \\ \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2} & 1+\frac{\gamma^2}{M} & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 0 & 0 & \dots & 1+\frac{\gamma^2}{M} & \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2} & \dots \\ 0 & 0 & \dots & \frac{2\gamma}{M} + \frac{(M-1)\gamma^2}{M^2} & 1+\frac{\gamma^2}{M} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \end{bmatrix}$$

These components stem from the coefficients on α from the re-arranged system of first-order conditions.

Notice that not only is $\tilde{X}$ block diagonal, the diagonal takes the exact same value for the nth block of students. The block diagonal structure results from the fact that each block of students is entirely independent. In order to solve this system of equations we are going to need to invert $\tilde{X}$. The key is that in order to do this we only need to invert the sub-matrix $[A\ B; C\ D]$. For simplicity, we now refer to this sub-matrix as X.

X is constructed such that the first row is the equation pertaining to the first order condition for the student who appears in both classes. The second row is the first order condition for the first single observation student in time period one, and so on. The sub-matrix Y for the nth block of students then takes the following form:

$$\mathbf{Y}_{((2M+1)\mathrm{x}1)} = \begin{bmatrix} y_{11n} + y_{12n} + \frac{\gamma}{M}\sum_{t=1}^{2}\sum_{j=2}^{M+1} y_{jtn} \\ y_{21n} + \frac{\gamma}{M} y_{11n} + \frac{\gamma}{M}\sum_{j=3}^{M+1} y_{j1n} \\ \vdots \\ y_{(M+1)1n} + \frac{\gamma}{M} y_{11n} + \frac{\gamma}{M}\sum_{j=2}^{M} y_{j1n} \\ y_{22n} + \frac{\gamma}{M} y_{12n} + \frac{\gamma}{M}\sum_{j=3}^{M+1} y_{j2n} \\ \vdots \\ y_{(M+1)2n} + \frac{\gamma}{M} y_{12n} + \frac{\gamma}{M}\sum_{j=2}^{M} y_{j2n} \end{bmatrix}$$

The form for Y again derives from the re-arranged system of first order-conditions. To derive the full matrix $\tilde{Y}$, simply stack the Y's for each block n, yielding a $\mathcal{N}(2M+1)$ x 1 matrix.

Now we need to solve for the inverse of $X = [A\ B; C\ D]$. Once we have this we can easily construct the inverse of $\tilde{X}$ and solve the system of first-order conditions for α. We know that X can be inverted blockwise according to

$$\mathbf{X^{-1}} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix} \tag{3}$$

as derived by Banachiewicz(1937). Since $(A - BD^{-1}C)^{-1}$ is just a scalar, the only difficult component of this formula is D^{-1}. However, notice that D is block diagonal where each block is MxM. Thus to get D^{-1} we just need to invert one of these MxM matrices. Depending on the size of M this may in itself be difficult. However, we can recursively apply the same blockwise formula to this MxM matrix until we finally get to the point where we only have to invert a two-by-two matrix.

As an example, consider the case where $M = 2$. This is a relatively simple case, but it becomes clear where the recursive application of the blockwise inversion formula is used. For $M = 2$, the various components of X take the following form

$$\mathbf{A}_{(1\text{x}1)} = 2 + \gamma^2$$

$$\mathbf{B}_{(1\text{x}4)} = \frac{\gamma(4+\gamma)}{4}[1\ 1\ 1\ 1]$$

$$\mathbf{C}_{(4\text{x}1)} = \frac{\gamma(4+\gamma)}{4}[1\ 1\ 1\ 1]'$$

$$\mathbf{D}_{(4\text{x}4)} = \frac{1}{4}\begin{bmatrix} 2(2+\gamma^2) & \gamma(4+\gamma) & 0 & 0 \\ \gamma(4+\gamma) & 2(2+\gamma^2) & 0 & 0 \\ 0 & 0 & 2(2+\gamma^2) & \gamma(4+\gamma) \\ 0 & 0 & \gamma(4+\gamma) & 2(2+\gamma^2) \end{bmatrix}$$

The next step is to invert D. Because of its blockwise diagonal structure we just need to find the inverse of the two-by-two given by

$$\begin{bmatrix} 2(2+\gamma^2) & \gamma(4+\gamma) \\ \gamma(4+\gamma) & 2(2+\gamma^2) \end{bmatrix}$$

In this simple case, D^{-1} is given by

$$\mathbf{D}^{-1}_{(4\text{x}4)} = \frac{4}{16 + 3\gamma^4 - 8\gamma^3}\begin{bmatrix} 2(2+\gamma^2) & -\gamma(4+\gamma) & 0 & 0 \\ -\gamma(4+\gamma) & 2(2+\gamma^2) & 0 & 0 \\ 0 & 0 & 2(2+\gamma^2) & -\gamma(4+\gamma) \\ 0 & 0 & -\gamma(4+\gamma) & 2(2+\gamma^2) \end{bmatrix}$$

where $16 + 3\gamma^4 - 8\gamma^3$ is simply the determinant of the simple two-by-two outlined above.

Now that we have calculated D^{-1}, we want to start calculating the various components of the blockwise inversion formula for X^{-1} as given in Equation (3). One component that appears numerous times is $A - BD^{-1}C$. Given our definition of A, B, C, and D^{-1} we can easily calculate this.

$$BD^{-1} = \left(\frac{4}{16 + 3\gamma^4 - 8\gamma^3}\right)\left(\frac{\gamma(4+\gamma)}{4}\right)[1\ 1\ 1\ 1]\begin{bmatrix} 2(2+\gamma^2) & -\gamma(4+\gamma) & 0 & 0 \\ -\gamma(4+\gamma) & 2(2+\gamma^2) & 0 & 0 \\ 0 & 0 & 2(2+\gamma^2) & -\gamma(4+\gamma) \\ 0 & 0 & -\gamma(4+\gamma) & 2(2+\gamma^2) \end{bmatrix}$$

which simplifies to

$$BD^{-1} = \left(\frac{\gamma(4+\gamma)}{3\gamma^2 + 4\gamma + 4}\right)[1\ 1\ 1\ 1]$$

Then,

$$BD^{-1}C = \left(\frac{\gamma(4+\gamma)}{3\gamma^2 + 4\gamma + 4}\right)\left(\frac{\gamma(4+\gamma)}{4}\right)[1\ 1\ 1\ 1]\,[1\ 1\ 1\ 1]'$$

which simplifies nicely to

$$BD^{-1}C = \frac{\gamma^2(4+\gamma)^2}{3\gamma^2 + 4\gamma + 4}$$

Finally,

$$A - BD^{-1}C = 2 + \gamma^2 - \frac{\gamma^2(4+\gamma)^2}{3\gamma^2 + 4\gamma + 4}$$

which reduces to

$$A - BD^{-1}C = \frac{2(1+\gamma)^2(\gamma-2)^2}{3\gamma^2 + 4\gamma + 4}$$

We have just calculated the leading term in the matrix given by Equation (3). The rest of the terms take the following forms:

$$-(A - BD^{-1}C)^{-1}BD^{-1} = -\frac{\gamma(4+\gamma)}{2(1+\gamma)^2(\gamma-2)^2}[1\ 1\ 1\ 1]$$

$$-D^{-1}C(A - BD^{-1}C)^{-1} = -\frac{\gamma(4+\gamma)}{2(1+\gamma)^2(\gamma-2)^2}[1\ 1\ 1\ 1]'$$

$$D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} = \frac{1}{2(1+\gamma)^2(\gamma-2)^2(3\gamma^2 + 4\gamma + 4)}\begin{bmatrix} \tilde{c} & \tilde{d} & \tilde{e} & \tilde{e} \\ \tilde{d} & \tilde{c} & \tilde{e} & \tilde{e} \\ \tilde{e} & \tilde{e} & \tilde{c} & \tilde{d} \\ \tilde{e} & \tilde{e} & \tilde{d} & \tilde{c} \end{bmatrix}$$

where

$$\begin{aligned} \tilde{c} &= 16(2+\gamma^2)(1+\gamma)^2 + \gamma^2(4+\gamma)^2 \\ \tilde{d} &= (4+\gamma)(-8\gamma(1+\gamma)^2 + \gamma^2(4+\gamma)) \\ \tilde{e} &= \gamma^2(4+\gamma)^2 \end{aligned}$$

We now have the form for X^{-1} in the case of $M = 2$. Remember that in order to solve the system of first order equations for α we need $\tilde{X}^{-1}$. This can be obtained according to the following: $\tilde{X}^{-1} = I_{\mathcal{N}} \otimes X^{-1}$. Finally, pre-multiply $\tilde{Y}$ with $\tilde{X}^{-1}$ to find a solution for α.

For $M > 2$ we can follow the same basic steps to calculate X^{-1}. The essential difference is that calculating D^{-1} is significantly more complicated. However, repeated application of Equation (3) on D will eventually yield a two-by-two matrix that can be inverted by hand. Doing so, one can show that the general formula for X^{-1} as a function of M is given by,

$$\mathbf{X}^{-1}_{((2M+1)x(2M+1))} = \begin{bmatrix} \tilde{A} & \tilde{B} & \tilde{B} & \dots & \tilde{B} & \tilde{B} & \tilde{B} & \dots \\ \tilde{B} & \tilde{C} & \tilde{D} & \dots & \tilde{E} & \tilde{E} & \tilde{E} & \dots \\ \tilde{B} & \tilde{D} & \tilde{C} & \dots & \tilde{E} & \tilde{E} & \tilde{E} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ \tilde{B} & \tilde{E} & \tilde{E} & \dots & \tilde{C} & \tilde{D} & \tilde{D} & \dots \\ \tilde{B} & \tilde{E} & \tilde{E} & \dots & \tilde{D} & \tilde{C} & \tilde{D} & \dots \\ \tilde{B} & \tilde{E} & \tilde{E} & \dots & \tilde{D} & \tilde{D} & \tilde{C} & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

where

$$\tilde{A} = \frac{\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2}{2(1+\gamma)^2(\gamma - M)^2}$$

$$\tilde{B} = \frac{\gamma^2 - \gamma(2+\gamma)M}{2(1+\gamma)^2(\gamma - M)^2}$$

$$\tilde{C} = \frac{\gamma^4 - 2\gamma^3(2+\gamma)M + \gamma^2(8+\gamma(12+5\gamma))M^2 - 4\gamma(1+\gamma)^2(2+\gamma)M^3 + 2(1+\gamma)^4 M^4}{2(1+\gamma)^2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2)}$$

$$\tilde{D} = \frac{\gamma^4 - 2\gamma^3(2+\gamma)M + \gamma^2(6+\gamma(8+3\gamma))M^2 - 2\gamma(1+\gamma)^2(2+\gamma)M^3}{2(1+\gamma)^2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2)}$$

$$\tilde{E} = \frac{\gamma^2(\gamma(M-1) + 2M)^2}{2(1+\gamma)^2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2)}$$

Again, X is just a sub-component of $\tilde{X}$. $\tilde{X}^{-1}$ is given by $I_{\mathcal{N}} \otimes X^{-1}$, where $I_{\mathcal{N}}$ is an ($\mathcal{N}$ x $\mathcal{N}$) identity matrix.

Using X^{-1} and the formula for Y we can solve for the α's within any particular block n as a function of γ, y, and M. As an example, the solution for α_{1n} can be obtained by multiplying Y by the first row of X^{-1}.

$$\alpha_{1n} = \tilde{A}\left(y_{11n} + y_{12n} + \frac{\gamma}{M}\sum_{t=1}^{2}\sum_{j=2}^{M+1} y_{jtn}\right) + \tilde{B}\sum_{t=1}^{2}\sum_{i=2}^{M+1}\left(y_{itn} + \frac{\gamma}{M}y_{1tn} + \frac{\gamma}{M}\sum_{j=2,j\neq i}^{M+1} y_{jtn}\right)$$

We can re-arrange this formula such that we group all the common y terms together. Doing so yields the solution for α_{1n} in terms of $\tilde{A}$ and $\tilde{B}$,

$$\alpha_{1n} = \left(\tilde{A} + \gamma\tilde{B}\right)(y_{11n} + y_{12n}) + \left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(M-1)+M}{M}\right)\sum_{t=1}^{2}\sum_{j=2}^{M+1} y_{jtn}$$

Several factors allow this expression to simplify quite nicely. First, row one of X^{-1} is relatively simple, containing only one $\tilde{A}$ and a string of $\tilde{B}$ terms. These two terms have the same denominator, making combinations relatively simply. Combine this with the fact that the outcomes for individual one enter Y symmetrically, and we are able to derive this relatively simple expression.

The solution for any α in block n other than α_{1n} is significantly more complex since all the other rows of X^{-1} contain multiple terms. In addition, the outcomes for all the individuals observed only once do not enter Y in such a symmetric fashion. As an example, below is the formula for α_{21n}. To arrive at this formula simply multiply Y by the second row of X^{-1}.

$$\begin{aligned}\alpha_{21n} &= \tilde{B}\left(y_{11n} + y_{12n} + \frac{\gamma}{M}\sum_{t=1}^{2}\sum_{j=2}^{M+1} y_{jtn}\right) + \tilde{C}\left(y_{21n} + \frac{\gamma}{M}y_{11n} + \frac{\gamma}{M}\sum_{j=3}^{M+1} y_{j1n}\right)\\ &+ \tilde{D}\left(\sum_{i=3}^{M+1}\left(y_{i1n} + \frac{\gamma}{M}y_{11n} + \frac{\gamma}{M}\sum_{j=2,j\neq i}^{M+1} y_{j1n}\right)\right) + \tilde{E}\left(\sum_{i=2}^{M+1}\left(y_{i2n} + \frac{\gamma}{M}y_{12n} + \frac{\gamma}{M}\sum_{j=2,j\neq i}^{M+1} y_{j2n}\right)\right)\end{aligned}$$

Again, we can re-arrange the above, grouping on the y's

$$\begin{aligned}\alpha_{21n} &= \left(\tilde{B} + \tilde{C}\frac{\gamma}{M} + \tilde{D}\frac{(M-1)\gamma}{M}\right)y_{11n} + \left(\tilde{B} + \tilde{E}\gamma\right)y_{12n} + \left(\tilde{B}\frac{\gamma}{M} + \tilde{C} + \tilde{D}\frac{(M-1)\gamma}{M}\right)y_{21n}\\ &+ \left(\tilde{B}\frac{\gamma}{M} + \tilde{C}\frac{\gamma}{M} + \tilde{D}\frac{M+(M-2)\gamma}{M}\right)\sum_{j=3}^{M+1} y_{j1n} + \left(\tilde{B}\frac{\gamma}{M} + \tilde{E}\frac{M+(M-1)\gamma}{M}\right)\sum_{j=2}^{M+1} y_{j2n}\end{aligned}$$

The formula for α_{i1n} for $i > 2$ takes the same form as above, except that (1) y_{21n} becomes y_{i1n} and (2) the first summation on the second line will be over all $j \neq i$. The formula for α_{i2n}

for $i > 1$ also takes the same general form, except that all of the subscripts denoting period 1 need to be changed to denote period 2, and vice versa.

QED

Proof of Lemma 2.G

Lemma 1 provides a solution for α strictly as a function of y, γ, and M. We can substitute this solution back into the original optimization problem to derive the desired result.

First, re-write the original optimization problem as a function of the residuals for each student outcome,

$$\min_{\alpha,\gamma} \sum_{n=1}^{\mathcal{N}} (e_{11n}^2 + e_{12n}^2) + \sum_{n=1}^{\mathcal{N}} \sum_{t=1}^{2} \sum_{i=2}^{M+1} e_{itn}^2$$

where the residuals themselves are functions of α and γ. Next we show that the residuals simplify significantly once we substitute for α and re-arrange.

Consider the residual for individual 1 in block n at time period 1,

$$e_{11n} = y_{11n} - \alpha_{1n} - \frac{\gamma}{M} \sum_{j=2}^{M+1} \alpha_{j1n}$$

Substituting in the solutions for α_{1n} and α_{j1n} and combining like terms yields the following:

$$\begin{aligned} e_{11n} \quad = \quad & y_{11n} \left(1 - \tilde{A} - 2\gamma\tilde{B} - \tilde{C}\frac{\gamma^2}{M} - \tilde{D}\frac{\gamma^2(M-1)}{M}\right) - y_{12n}\left(\tilde{A} + 2\gamma\tilde{B} + \gamma^2\tilde{E}\right) \\ & - \left(\sum_{j=2}^{M+1} y_{j1n}\right)\left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma+M-1)+M}{M} + \tilde{C}\frac{\gamma M(1+\gamma)-\gamma^2}{M^2} + \tilde{D}\frac{\gamma(\gamma+M(M+M\gamma-2\gamma-1))}{M^2}\right) \\ & - \left(\sum_{j=2}^{M+1} y_{j2n}\right)\left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma+M-1)+M}{M} + \tilde{E}\frac{\gamma M(1+\gamma)-\gamma^2}{M}\right) \end{aligned} \tag{4}$$

Using the formulas for $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ we show that the coefficients on the outcomes (y) simplify quite nicely. First we illustrate how $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ are functionally related.

Property 1

The components of X^{-1} are interrelated according to the following:

$$\tilde{A} = \tilde{B} + \frac{M^2}{2(\gamma-M)^2}, \quad \tilde{C} = \tilde{D} + \frac{M^2}{(\gamma-M)^2}, \quad \tilde{D} = \tilde{B} + \frac{V}{2}, \quad \tilde{E} = \tilde{B} - \frac{V}{2}$$

where

$$V = \frac{\gamma M^2(-\gamma + 2M + \gamma M)}{(\gamma-M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2)}$$

Proof of Property 1

Solving for $\tilde{A}$ as a function of $\tilde{B}$ is rather straightforward as they have the same denominator.

$$\begin{aligned}\tilde{A}-\tilde{B} &= \frac{\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2-(\gamma^2-\gamma(2+\gamma)M)}{2(1+\gamma)^2(\gamma-M)^2} \\ &= \frac{M^2}{2(\gamma-M)^2}\end{aligned}$$

In order to relate $\tilde{C}$ to $\tilde{B}$, we first show how $\tilde{C}$ is related to $\tilde{D}$ and then how $\tilde{D}$ is related to $\tilde{B}$. Below are the formulas for $\tilde{C}$ and $\tilde{D}$.

$$\tilde{C}=\frac{\gamma^4-2\gamma^3(2+\gamma)M+\gamma^2(8+\gamma(12+5\gamma))M^2-4\gamma(1+\gamma)^2(2+\gamma)M^3+2(1+\gamma)^4M^4}{2(1+\gamma)^2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}$$

$$\tilde{D}=\frac{\gamma^4-2\gamma^3(2+\gamma)M+\gamma^2(6+\gamma(8+3\gamma))M^2-2\gamma(1+\gamma)^2(2+\gamma)M^3}{2(1+\gamma)^2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}$$

Both terms share the same denominator, and in fact share the same first two terms in the numerator. Subtracting $\tilde{D}$ from $\tilde{C}$ and simplifying yields

$$\tilde{C}-\tilde{D}=\frac{M^2}{(\gamma-M)^2}$$

Next we want to find the difference between $\tilde{D}$ and $\tilde{B}$. This difference is more complicated than the first two since they do not share the same denominator. However we can easily get a common denominator since the denominator for $\tilde{B}$ is simply missing one term present in the denominator of $\tilde{D}$. Thus we can write the difference as

$$\begin{aligned}\tilde{D}-\tilde{B} &= \frac{\gamma^4-2\gamma^3(2+\gamma)M+\gamma^2(6+\gamma(8+3\gamma))M^2-2\gamma(1+\gamma)^2(2+\gamma)M^3}{2(1+\gamma)^2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)} \\ &- \frac{(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)(\gamma^2-\gamma(2+\gamma)M)}{2(1+\gamma)^2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}\end{aligned}$$

Combining terms and simplifying yields

$$\begin{aligned}\tilde{D}-\tilde{B} &= \frac{\gamma M^2(\gamma-2M-\gamma M)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)} \\ &= \frac{V}{2}\end{aligned}$$

The last piece is to relate $\tilde{E}$ to $\tilde{B}$. Just as with $\tilde{D}$ we need to find a common denominator.

$$\begin{aligned}\tilde{E}-\tilde{B} &= \frac{\gamma^2(\gamma(M-1)+2M)^2-(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)(\gamma^2-\gamma(2+\gamma)M)}{2(1+\gamma)^2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)} \\ &= \frac{\gamma M^2(-\gamma+2M+\gamma M)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)} \\ &= -\frac{V}{2}\end{aligned}$$

QED

Using Property 1, we now show that the coefficients on the observed grades in Equation (4) have other appealing properties. Then we use these properties to simplify Equation (4), in an effort to arrive at a simplified version of the least squares problem as a function of γ.

Property 2

Equation (4) describes the prediction error for the first outcome of the individual observed twice in block n. In Equation (4), the coefficient on y_{11n} is equal to the coefficient on y_{12n}, and the coefficient on $\sum_{j=2}^{M+1} y_{j1n}$ is equal in magnitude but of the opposite sign as the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$.

Proof of Property 2

The coefficient on y_{11n} is given by

$$1-\tilde{A}-2\gamma\tilde{B}-\tilde{C}\frac{\gamma^2}{M}-\tilde{D}\frac{\gamma^2(M-1)}{M}$$

Substituting in for $\tilde{A}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ as a function of $\tilde{B}$ using Property 1 and simplifying yields the following

$$1-\tilde{A}-2\gamma\tilde{B}-\tilde{C}\frac{\gamma^2}{M}-\tilde{D}\frac{\gamma^2(M-1)}{M}=\frac{2\gamma^2-2\gamma(\gamma-2)M+M^2-V\gamma^2(\gamma-M)^2}{2(\gamma-M)^2}-\tilde{B}(1+\gamma)^2$$

The coefficient on y_{12n} is given by

$$\tilde{A}+2\gamma\tilde{B}+\gamma^2\tilde{E}$$

Again making the appropriate substitutions allowed by Property 1, we can re-write this expression as

$$\tilde{A}+2\gamma\tilde{B}+\gamma^2\tilde{E}=\frac{M^2-V\gamma^2(\gamma-M)^2}{2(\gamma-M)^2}+\tilde{B}(1+\gamma)^2$$

Finally, taking the difference between the coefficients on y_{11n} and y_{12n}, we find

$$\begin{aligned}
&\tilde{A} + 2\gamma\tilde{B} + \gamma^2\tilde{E} - (1 - \tilde{A} - 2\gamma\tilde{B} - \tilde{C}\frac{\gamma^2}{M} - \tilde{D}\frac{\gamma^2(M-1)}{M}) \\
&= 2\tilde{B}(1+\gamma)^2 + \frac{M^2 - V\gamma^2(\gamma - M)^2 - 2\gamma^2 + 2\gamma(\gamma-2)M - M^2 + V\gamma^2(\gamma-M)^2}{2(\gamma-M)^2} \\
&= 2\tilde{B}(1+\gamma)^2 + \frac{-2\gamma^2 + 2\gamma(\gamma-2)M}{2(\gamma-M)^2} \\
&= \frac{\gamma^2 - \gamma(2+\gamma)M}{(\gamma-M)^2} + \frac{-2\gamma^2 + 2\gamma(\gamma-2)M}{2(\gamma-M)^2} \\
&= 0
\end{aligned}$$

where the second to last line results from substituting in our formula for $\tilde{B}$.

Now we show that the coefficient on $\sum_{j=2}^{M+1} y_{j1n}$ is equal in magnitude but of the opposite sign as the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$. The coefficient on $\sum_{j=2}^{M+1} y_{j1n}$ is given by

$$\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma+M-1)+M}{M} + \tilde{C}\frac{\gamma M(1+\gamma) - \gamma^2}{M^2} + \tilde{D}\frac{\gamma(\gamma + M(M + M\gamma - 2\gamma - 1))}{M^2}$$

and the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$ is given by

$$\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma+M-1)+M}{M} + \tilde{E}\frac{\gamma M(1+\gamma) - \gamma^2}{M}$$

If we add these two coefficients together we arrive at the following expression

$$2\tilde{A}\frac{\gamma}{M} + 2\tilde{B}\frac{\gamma(\gamma+M-1)+M}{M} + \tilde{C}\frac{\gamma M(1+\gamma) - \gamma^2}{M^2} + \tilde{D}\frac{\gamma(\gamma + M(M + M\gamma - 2\gamma - 1))}{M^2} + \tilde{E}\frac{\gamma M(1+\gamma) - \gamma^2}{M}$$

Now, we substitute for $\tilde{A}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ as functions of $\tilde{B}$ from Property 1. After some manipulation, we can write the above expression in the following form

$$2\tilde{B}(1+\gamma)^2 + \frac{4\gamma M + 2\gamma^2 M - 2\gamma^2}{2(\gamma-M)^2}$$

Notice that this expression contains no V terms as they cancel out when substituting in for $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$. The last step is to substitute in for $\tilde{B}$.

$$\frac{\gamma^2 - \gamma(2+\gamma)M}{(\gamma-M)^2} + \frac{4\gamma M + 2\gamma^2 M - 2\gamma^2}{2(\gamma-M)^2}$$

All of the terms in the above expression cancel out, proving that the sum of the coefficients on $\sum_{j=2}^{M+1} y_{j1n}$ and $\sum_{j=2}^{M+1} y_{j2n}$ are equal in magnitude and of the opposite sign.

QED

Now we return to Equation (4), which describe the prediction error for the first observation of the student observed twice in block n. Using Properties 1 and 2 we will show how to simplify this expression, and in turn describe how the prediction errors for all of the other outcomes in block n can be similarly simplified. This will yield a simplified version of the original least squares problem, Equation (1), strictly as a function of γ, M, and y.

Property 2 indicates that

$$1-\tilde{A}-2\gamma\tilde{B}-\tilde{C}\frac{\gamma^2}{M}-\tilde{D}\frac{\gamma^2(M-1)}{M}=\tilde{A}+2\gamma\tilde{B}+\gamma^2\tilde{E}$$

and

$$\begin{aligned}&\tilde{A}\frac{\gamma}{M}+\tilde{B}\frac{\gamma(\gamma+M-1)+M}{M}+\tilde{C}\frac{\gamma M(1+\gamma)-\gamma^2}{M^2}+\tilde{D}\frac{\gamma(\gamma+M(M+M\gamma-2\gamma-1))}{M^2}\\&=-(\tilde{A}\frac{\gamma}{M}+\tilde{B}\frac{\gamma(\gamma+M-1)+M}{M}+\tilde{E}\frac{\gamma M(1+\gamma)-\gamma^2}{M})\end{aligned}$$

We now proceed to solve for each of these coefficients strictly as a function of γ. First, we solve for the coefficient on y_{12n}.

By substituting for $\tilde{A}$ and $\tilde{E}$ from Property 1, we can write the coefficient on y_{12n} in the following way

$$\tilde{A}+2\gamma\tilde{B}+\gamma^2\tilde{E}=\frac{M^2-V\gamma^2(\gamma-M)^2}{2(\gamma-M)^2}+\tilde{B}(1+\gamma)^2$$

To solve for this as a function of γ we need to substitute in for $\tilde{B}$ and V. Substituting in for $\tilde{B}$ and V from Property 1 yields

$$=\frac{(\gamma(\gamma-(2+\gamma)M)+M^2)(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)-\gamma^2M^2(\gamma^2-2\gamma M-\gamma^2M)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}$$

We can re-arrange this expression in the following manner:

$$\begin{aligned}=&\quad\frac{(\gamma^2-2\gamma M+M^2)(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}\\&-\frac{\gamma^2M(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2+\gamma^2M-2\gamma M^2-\gamma^2M^2)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}\end{aligned}$$

where we split the expression simply for ease of presentation. The numerator in the second line simplifies greatly, such that the entire expression simplifies to

$$=\frac{(\gamma^2-2\gamma M+M^2)(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)-\gamma^2M(\gamma^2-2\gamma M+M^2)}{2(\gamma-M)^2(\gamma^2-\gamma(2+\gamma)M+(1+\gamma)^2M^2)}$$

The numerator then factors to produce

$$= \frac{(\gamma - M)^2(M + \gamma(M-1))^2}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

Finally, we cancel out the common terms in the numerator and denominator to yield

$$\tilde{A} + 2\gamma\tilde{B} + \gamma^2\tilde{E} = \frac{(M + \gamma(M-1))^2}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

This gives us the coefficient on y_{11n} and y_{12n} in the expression for e_{11n} as a function of γ. Now we proceed to solve for the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$ as a function of γ.

Using Property 1, we can write the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$ in the following fashion:

$$\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma + M - 1) + M}{M} + \tilde{E}\frac{\gamma M(1+\gamma) - \gamma^2}{M} = \tilde{B}(1+\gamma)^2 + \frac{\gamma M}{2(\gamma - M)^2} - \frac{V(\gamma M(1+\gamma) - \gamma^2)}{2M}$$

Substituting for $\tilde{B}$ and re-arranging yields

$$\frac{\gamma^2 - \gamma M - \gamma^2 M}{2(\gamma - M)^2} - \frac{V(\gamma M(1+\gamma) - \gamma^2)}{2M}$$

Substituting for V from Property 1 and finding a common denominator yields

$$\frac{(\gamma^2 - \gamma M - \gamma^2 M)(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2) - \gamma M(\gamma - 2M - \gamma M)(\gamma M(1+\gamma) - \gamma^2)}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

After some manipulation the numerator of the above expression simplifies to yield

$$\frac{-(\gamma - M)^2(\gamma M(1+\gamma) - \gamma^2)}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

Canceling out the common terms in the numerator and denominator yields

$$\frac{-(\gamma M(1+\gamma) - \gamma^2)}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

Finally we can substitute our simplified versions of the coefficients on y_{11n}, y_{12n}, $\sum_{j=2}^{M+1} y_{j2n}$, and $\sum_{j=2}^{M+1} y_{j2n}$ back into the equation for e_{11n}, described in Equation (4).

$$e_{11n} = \left(\frac{(M + \gamma(M-1))^2}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}(y_{11n} - y_{12n}) + \frac{\gamma(M + \gamma(M-1))}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}\sum_{j=2}^{M+1}(y_{j2n} - y_{j1n})\right)$$

This simplifies further to produce

$$e_{11n} = \frac{(M + \gamma(M-1))}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}\left((M + \gamma(M-1))(y_{11n} - y_{12n}) + \gamma\sum_{j=2}^{M+1}(y_{j2n} - y_{j1n})\right)$$

We now have the component of the least squares problem that corresponds to the residual for student 1 in block n as a function of γ with the α's concentrated out. Next, we need to find similar expressions for e_{12n} and e_{itn}.

Finding a version of e_{12n} as function of γ is simple since it takes a form that is essentially identical to e_{11n}. The expression for e_{12n} is given by

$$e_{12n} = y_{12n} - \alpha_{1n} - \frac{\gamma}{M} \sum_{j=2}^{M+1} \alpha_{j2n}$$

which after substituting for α using the results from Lemma 1 yields

$$\begin{aligned} e_{12n} &= y_{12n} \left(1 - \tilde{A} - 2\gamma \tilde{B} - \tilde{C}\frac{\gamma^2}{M} - \tilde{D}\frac{\gamma^2(M-1)}{M}\right) - y_{11n} \left(\tilde{A} + 2\gamma \tilde{B} + \gamma^2 \tilde{E}\right) \\ &- \left(\sum_{j=2}^{M+1} y_{j2n}\right) \left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma + M - 1) + M}{M} + \tilde{C}\frac{\gamma M(1+\gamma) - \gamma^2}{M^2} + \tilde{D}\frac{\gamma(\gamma + M(M + M\gamma - 2\gamma - 1))}{M^2}\right) \\ &- \left(\sum_{j=2}^{M+1} y_{j1n}\right) \left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{\gamma(\gamma + M - 1) + M}{M} + \tilde{E}\frac{\gamma M(1+\gamma) - \gamma^2}{M}\right) \end{aligned}$$

This equation is identical to the equation for e_{11n} except that all the time subscripts are changed. However, we know from Property 2 that the coefficients on y_{11n} and y_{12n} are equal in this expression and that coefficients on $\sum_{j=2}^{M+1} y_{j2n}$ and $\sum_{j=2}^{M+1} y_{j2n}$ are equal but of the opposite sign. Thus, $e_{12n} = -e_{11n}$.

To get the final piece of the least squares problem with the α's concentrated out we need to substitute for α in e_{itn}, where $i > 1$. To find a simplified formula for e_{itn} consider first substituting in for α in e_{21n}. The formula for e_{21n} from the basic least squares problem can be written as follows:

$$e_{21n} = y_{21n} - \alpha_{21n} - \frac{\gamma}{M}\left(\alpha_{1n} + \sum_{j=3}^{M+1} \alpha_{j1n}\right)$$

Substituting in for α from Lemma 1 and combining like terms yields the following expression:

$$
\begin{aligned}
e_{21n} \quad = \quad & y_{21n}\left(1 - \tilde{A}\frac{\gamma^2}{M^2} - 2\tilde{B}\frac{M\gamma + \gamma^2(M-1)}{M^2} - \tilde{C}\frac{M^2 + \gamma^2(M-1)}{M^2} - \tilde{D}\frac{2\gamma M(M-1) + \gamma^2(M-1)(M-2)}{M^2}\right) \\
& - y_{11n}\left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{M + (M-1)\gamma + \gamma^2}{M} + \tilde{C}\frac{\gamma M + (M-1)\gamma^2}{M^2} + \tilde{D}\frac{(M-1)\gamma(M + (M-1)\gamma)}{M^2}\right) \\
& - y_{12n}\left(\tilde{A}\frac{\gamma}{M} + \tilde{B}\frac{M + (M-1)\gamma + \gamma^2}{M} + \tilde{E}\frac{\gamma M + (M-1)\gamma^2}{M}\right) \\
& - \left(\sum_{j=3}^{M+1} y_{j1n}\right)\left(\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{C}\frac{2\gamma M + \gamma^2(M-2)}{M^2} + \tilde{D}\frac{(M + \gamma(M-2))^2 + (M-1)\gamma}{M^2}\right. \\
& - \left(\sum_{j=2}^{M+1} y_{j2n}\right)\left(\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{E}\frac{(M + (M-1)\gamma)^2}{M^2}\right)
\end{aligned}
$$

To simplify the above expression, we follow the same strategy employed in simplifying e_{11n}.

Property 3

The coefficients on y_{11n} and y_{12n} in the equation for e_{21n} are equal in magnitude but of the opposite sign. The same relationship exists between the coefficients on $\sum_{j=3}^{M+1} y_{j1n}$ and $\sum_{j=2}^{M+1} y_{j2n}$. In addition, the coefficient on y_{21n} is identical to the coefficient on $\sum_{j=3}^{M+1} y_{j1n}$.

Proof of Property 3

The first step is to examine the coefficients on y_{11n} and y_{12n}. Our work is simple here since the coefficients on y_{11n} and y_{12n} in the expression for e_{21n} and the coefficients on $\sum_{j=2}^{M+1} y_{j1n}$ and $\sum_{j=2}^{M+1} y_{j2n}$ in the expression for e_{11n} are exactly the same. Thus, we know they are opposite in sign, and of magnitude

$$\frac{(\gamma M(1+\gamma) - \gamma^2)}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2 M^2)}$$

by Property 2.

Now we turn to the coefficients on y_{21n}, $\sum_{j=3}^{M+1} y_{j1n}$, and $\sum_{j=2}^{M+1} y_{j2n}$. Using the results from Property 1 relating $\tilde{A}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ to $\tilde{B}$, we can re-write the coefficient on $\sum_{j=3}^{M+1} y_{j1n}$ in the following fashion:

$$
\begin{aligned}
&\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{C}\frac{2\gamma M + \gamma^2(M-2)}{M^2} + \tilde{D}\frac{(M + \gamma(M-2))^2 + (M-1)\gamma^2}{M^2} \\
&\qquad = \tilde{B}(1+\gamma)^2 + \frac{V(\gamma^2 - 2\gamma M(1+\gamma) + M^2(1+\gamma)^2)}{2M^2} + \frac{2\gamma M(2+\gamma) - 3\gamma^2}{2(\gamma - M)^2}
\end{aligned}
$$

Next, substituting in for $\tilde{A}$, $\tilde{D}$, and $\tilde{E}$ in the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$ and simplifying yields

$$\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{E}\frac{(M+(M-1)\gamma)^2}{M^2}$$
$$= \tilde{B}(1+\gamma)^2 - \frac{V(\gamma^2 - 2\gamma M(1+\gamma) + M^2(1+\gamma)^2)}{2M^2} + \frac{\gamma^2}{2(\gamma - M)^2}$$

Adding together the simplified expressions for the coefficients on $\sum_{j=3}^{M+1} y_{j1n}$ and $\sum_{j=2}^{M+1} y_{j2n}$ yields

$$2\tilde{B}(1+\gamma)^2 + \frac{\gamma^2}{2(\gamma-M)^2} + \frac{2\gamma M(2+\gamma) - 3\gamma^2}{2(\gamma-M)^2}$$

after the terms including V cancel each other. Substituting in our expression for $\tilde{B}$ yields

$$\frac{\gamma^2 - \gamma(2+\gamma)M}{(\gamma-M)^2} + \frac{2\gamma M + \gamma^2 M - \gamma^2}{(\gamma-M)^2}$$

All the terms in the above expression cancel out, indicating that

$$\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{C}\frac{2\gamma M + \gamma^2(M-2)}{M^2} + \tilde{D}\frac{(M+\gamma(M-2))^2 + (M-1)\gamma^2}{M^2}$$
$$= -\left(\tilde{A}\frac{\gamma^2}{M^2} + 2\tilde{B}\frac{\gamma M + \gamma^2(M-1)}{M^2} + \tilde{E}\frac{(M+(M-1)\gamma)^2}{M^2}\right)$$

or that the coefficients on $\sum_{j=3}^{M+1} y_{j1n}$ and $\sum_{j=2}^{M+1} y_{j2n}$ are equal in magnitude but of the opposite sign.

Finally, we can substitute in for $\tilde{A}$, $\tilde{C}$, $\tilde{D}$, and $\tilde{E}$ as function of $\tilde{B}$ from Property 1 in the coefficient for y_{21n}. After some simplification we can show that this coefficient can be written as

$$-\tilde{B}(1+\gamma)^2 - \frac{V(\gamma^2 - 2\gamma M(1+\gamma) + M^2(1+\gamma)^2)}{2M^2} - \frac{2\gamma M(2+\gamma) - 3\gamma^2}{2(\gamma-M)^2}$$

Comparing this to the coefficient on $\sum_{j=3}^{M+1} y_{j1n}$ as shown above indicates that these two expressions are exactly the same, except that the signs are flipped on all the terms. Thus, the coefficients for y_{21n} and $\sum_{j=3}^{M+1} y_{j1n}$ are equal in magnitude but of the opposite sign.
QED

All that remains is to find the expression for these three coefficients as a function of γ. We can work with the easiest formula since they are all identical. Recall that the coefficient on $\sum_{j=2}^{M+1} y_{j2n}$ can be written

$$\tilde{B}(1+\gamma)^2 - \frac{V(\gamma^2 - 2\gamma M(1+\gamma) + M^2(1+\gamma)^2)}{2M^2} + \frac{\gamma^2}{2(\gamma-M)^2}$$

Substituting in for V yields

$$\tilde{B}(1+\gamma)^2 - \frac{\gamma(\gamma M + 2M - \gamma)(M^2(1+\gamma)^2 + \gamma^2 - 2\gamma M(1+\gamma))}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)} + \frac{\gamma^2}{2(\gamma - M)^2}$$

Finding a common denominator and re-arranging yields

$$\tilde{B}(1+\gamma)^2 + \frac{\gamma^2((\gamma - M)^2 + \gamma M(2M - \gamma + \gamma M)) - \gamma(\gamma M + 2M - \gamma)((\gamma - M)^2 + \gamma M(2M + \gamma M - 2\gamma))}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

Finally, substituting in for $\tilde{B}$, finding a common denominator, and eliminating terms yields.

$$\frac{\gamma^2((\gamma - M)^2 + \gamma M(2M - \gamma + \gamma M)) - \gamma^2(\gamma M + 2M - \gamma)}{2(\gamma - M)^2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

The above expression simplifies further to

$$\frac{\gamma^2}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

Now we have expressions for all the terms in the equation for e_{21n}. We can substitute back in and write the residual as a simple function of γ.

$$e_{21n} = \frac{\gamma(M + \gamma(M-1))}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}(y_{11n} - y_{12n}) + \frac{\gamma^2}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)} \sum_{j=2}^{M+1}(y_{j2n} - y_{j1n})$$

Notice that in the residual for e_{21n}, we can combine y_{21n} and $\sum_{j=3}^{M+1} y_{j1n}$ since they share the exact same coefficient. This means that the form of e_{j1n} for all $j > 1$ will take the exact form as the equation for e_{21n}. In addition, if we were to write down the equation for e_{22n}, it would take the exact same form as the equation for e_{21n}, except the coefficients would be swapped across the two time periods. As a result, $e_{22n} = -e_{21n}$. These relationships will allow us to greatly simplify the least squares problem.

We can simplify the solution for e_{21n} by factoring out the common terms in the numerator and denominator of each term. Doing so yields

$$e_{21n} = \frac{\gamma}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}\left((M + \gamma(M-1))(y_{11n} - y_{12n}) + \gamma \sum_{j=2}^{M+1}(y_{j2n} - y_{j1n})\right)$$

Finally we have all the components of the least squares problem strictly as functions of y, γ, and M. Re-writing the least squares problem in terms of the residuals yields

$$\min_{\gamma} \sum_{n=1}^{\mathcal{N}} \left(e_{11n}^2 + e_{12n}^2 + \sum_{j=2}^{M+1}(e_{i1n}^2 + e_{i2n}^2)\right)$$

Using the fact that $e_{12n} = -e_{11n}$, $e_{j1n} = e_{21n}$ for i greater than 3, and $e_{22n} = -e_{21n}$ we can simplify the above expression to

$$\min_{\gamma} \; 2\sum_{n=1}^{\mathcal{N}} \left(e_{11n}^2 + Me_{21n}^2\right)$$

Now substituting in for the residuals using the results previously derived yields

$$\min_{\gamma} \; 2\sum_{n=1}^{\mathcal{N}} \left[\frac{(M+\gamma(M-1))^2}{4(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)^2} \left((M+\gamma(M-1))(y_{11n} - y_{12n}) + \gamma \sum_{j=2}^{M+1} (y_{j2n} - y_{j1n}) \right)^2 \right.$$
$$\left. + \frac{\gamma^2 M}{4(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)^2} \left((M+\gamma(M-1))(y_{11n} - y_{12n}) + \gamma \sum_{j=2}^{M+1} (y_{j2n} - y_{j1n}) \right)^2 \right]$$

Notice that the terms inside the squares are exactly the same. We can re-arrange the above expression by combining like terms.

$$\min_{\gamma} \; 2\sum_{n=1}^{\mathcal{N}} \left[\frac{(M+\gamma(M-1))^2 + \gamma^2 M}{4(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)^2} \left((M+\gamma(M-1))(y_{11n} - y_{12n}) + \gamma \sum_{j=2}^{M+1} (y_{j2n} - y_{j1n}) \right)^2 \right]$$

We can pull the first term inside the summation out since it does not depend on n. In addition, this term can be simplified since the numerator and denominator take the same basic form. Finally, we are left with the following least squares problem,

$$\min_{\gamma} \; \frac{\sum_{n=1}^{\mathcal{N}} \left((M+\gamma(M-1))(y_{11n} - y_{12n}) + \gamma \sum_{j=2}^{M+1} (y_{j2n} - y_{j1n}) \right)^2}{2(\gamma^2 - \gamma(2+\gamma)M + (1+\gamma)^2M^2)}$$

If we define

$$\begin{aligned} W_{11n} &= M + \gamma(M-1) = -W_{12n} \\ W_{j1n} &= -\gamma = -W_{j2n} \; \forall \; j > 1 \end{aligned}$$

we can re-write the least squares problem in the following simplified manner,

$$\min_{\gamma} \; \sum_{n=1}^{\mathcal{N}} \frac{\left(\sum_{t=1}^{2} \sum_{j=1}^{M+1} W_{jtn} y_{jtn} \right)^2}{\sum_{t=1}^{2} \sum_{j=1}^{M+1} W_{jtn}^2}$$

QED

Proof of Lemma 3.G

Recall that $q(w, \gamma)$ is given by

$$q(w,\gamma) = \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt} y_{jt}\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}$$

where the W's are defined in the outline of Lemma 2.G. Substituting in for y_{jt} with the data generating process yields:

$$q(w,\gamma) = \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}\left[\alpha_{jto} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1}\alpha_{kto} + \epsilon_{jt}\right]\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}$$

Collecting the α_{jto} terms yields:

$$q(w,\gamma) = \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}\left(W_{jt} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1} W_{kt}\right)\alpha_{jto} + W_{jt}\epsilon_{jt}\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2} \tag{6}$$

Note that the coefficient on α_{1o} is given by the weight at $t = 1$ plus the weight at $t = 2$. Since $W_{1t} = -W_{2t}$, the coefficient on α_{1o} is zero.

Now consider the coefficient on α_{j1o} for $j > 1$. Substituting in for the weights yields:

$$W_{j1} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1} W_{k1} = -\gamma + \left[1 + \gamma - \frac{\gamma}{M}\right]\gamma_o - \left[\frac{\gamma(M-1)}{M}\right]\gamma_o$$

which reduces to:

$$W_{j1} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1} W_{k1} = (\gamma_o - \gamma)$$

We then know that the coefficient on α_{j2o} will be the negative of the coefficient on α_{j1o}. Substituting for these expressions in (6) yields:

$$\begin{aligned} q(w,\gamma) &= \left[\left((\gamma_o - \gamma)\left(\sum_{j=2}^{M+1}\alpha_{j1o}\right) + (\gamma - \gamma_o)\left(\sum_{j=2}^{M+1}\alpha_{j2o}\right) + \sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}\epsilon_{jt}\right)^2\right] \\ &\quad \Bigg/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right] \end{aligned}$$

Taking expectations:

$$\begin{aligned} E[q(w,\gamma)] &= \Bigg\{\Bigg[\Bigg((\gamma_o - \gamma)\left(\sum_{j=2}^{M+1}\alpha_{j1o}\right) + (\gamma - \gamma_o)\left(\sum_{j=2}^{M+1}\alpha_{j2o}\right) \\ &\quad + \sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}\epsilon_{jt}\Bigg)^2\Bigg] \Bigg/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right]\Bigg\} \end{aligned}$$

Expanding the square and noting that 1) $E(\alpha_{jto}\epsilon_{kt'}) = 0$ for all j, k, t, t' by assumption 2 and 2) $E(\epsilon_{jt}\epsilon_{kt'}) = 0$ for all $j \neq k$ or $t \neq t'$ by assumption 1 yields:

$$
\begin{aligned}
E[q(w,\gamma)] &= E\Bigg\{\Bigg[\left((\gamma_o - \gamma)\left(\sum_{j=2}^{M+1}\alpha_{j1o}\right) + (\gamma - \gamma_o)\left(\sum_{j=2}^{M+1}\alpha_{j2o}\right)\right)^2 \\
&\quad + \left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2 \epsilon_{jt}^2\right)\Bigg] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right]\Bigg\}
\end{aligned}
$$

which can be rewritten as:

$$
\begin{aligned}
E[q(w,\gamma)] &= (\gamma - \gamma_o)^2 E\left[\left(\sum_{j=2}^{M+1}(\alpha_{j2o} - \alpha_{j1o})\right)^2\right] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right] \\
&\quad + \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2 E\left(\epsilon_{jt}^2\right)\right] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right]
\end{aligned}
$$

Further, $E(\epsilon_{jt}^2) = E(\epsilon_{kt}^2)$ as the labeling of time periods is arbitrary and the expectation is unconditional. We can then express the expectation over the squared errors solely as a function of the first observation's squared error:

$$
\begin{aligned}
E[q(w,\gamma)] &= (\gamma - \gamma_o)^2 E\left[\left(\sum_{j=2}^{M+1}(\alpha_{j2o} - \alpha_{j1o})\right)^2\right] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right] \\
&\quad + \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2 E\left(\epsilon_{1t}^2\right)\right] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right]
\end{aligned}
$$

Note that the weights in the numerator of the second expectation are the same weights as in the denominator. Further, unconditionally $E(\epsilon_{11}^2) = E(\epsilon_{12}^2)$. Taking the unconditional expectation then yields:

$$
\begin{aligned}
E[q(w,\gamma)] &= (\gamma - \gamma_o)^2 E\left[\left(\sum_{j=2}^{M+1}(\alpha_{j2o} - \alpha_{j1o})\right)^2\right] \Big/ \left[\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right] \\
&\quad + E\left(\epsilon^2\right)
\end{aligned}
$$

The first term in the above expression is strictly greater than 0 for all $\gamma \neq \gamma_o$ and the second term does not depend upon γ. As a result, $E[q(w,\gamma_o)] < E[q(w,\gamma)]$ for all $\gamma \in \Gamma$ when $\gamma \neq \gamma_o$. QED.

Proof of Lemma 4.G

Uniform convergence requires that

$$\max_{\gamma\in\Gamma}\left|\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}} q(w_n,\gamma) - E[q(w,\gamma)]\right| \xrightarrow{p} 0$$

Theorem 12.1 in Wooldridge states four conditions that the data and q must satisfy in order for the above condition to hold.

1. Γ is compact

   This condition is satisfied by assumption 6.

2. For each $\gamma \in \Gamma$, $q(\cdot,\gamma)$ is Borel measurable on $\mathcal{W}$

   $q(\cdot,\gamma)$ is measurable with respect to product σ-algebra of $\mathcal{B}(\mathbb{R}^{2+2\overline{M}}) \times 2^{\mathcal{J}}$, where $2^{\mathcal{J}}$ is the power set over the possible class sizes.

3. For each $w \in \mathcal{W}$, $q(w,\cdot)$ is continuous on Γ

   Our concentrated objective function is continuous in γ.

4. $|q(w,\gamma)| \le b(w)$ for all $\gamma \in \Gamma$, where b is a nonnegative function on $\mathcal{W}$ such that $E[b(w)] < \infty$

   Recall that $q(w,\gamma)$ is given by:

$$q(w,\gamma) = \frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}y_{jt}\right)^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}$$

   Expanding the square and noting that $W_{jt}^2 y_{jt}^2 + W_{kt'}^2 y_{kt'}^2 \ge 2W_{jt}W_{kt'}y_{jt}y_{kt'}$ for all j,k,t,t' (the triangle inequality), we have:

$$q(w,\gamma) \le \frac{(2+M)\left(\sum_{j=1}^{M+1} W_{jt}^2 y_{jt}^2\right)}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}$$

   where the the leading term arises from replacing all the cross products using the triangle inequality.

   Note that each of the terms in the numerator and denominator are positive and that all weights in the numerator are also in the denominator, implying that:

$$q(w,\gamma) < (2+2M)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} y_{jt}^2\right) = b(w)$$

where we have shown that $b(w) > q(w, \gamma)$ for all w.

We now show that $E[b(w)] < \infty$. Note that $E[b(w)]$ is given by:

$$E[b(w)] = E\left[(2+M)\sum_{t=1}^{2}\sum_{j=1}^{M+1} y_{jt}^2\right]$$

Substituting in for y with the data generating process into the inner expectation yields:

$$E[b(w)] = (2+2M)E\left[\sum_{t=1}^{2}\sum_{j=1}^{M+1}\left(\alpha_{jto} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1}\alpha_{kto} + \epsilon_{jt}\right)^2\right]$$

Repeatedly using the triangle inequality after expanding the square implies:

$$E[b(w)] \leq (2+2M)E\left[\sum_{t=1}^{2}\sum_{j=1}^{M+1}(M+2)\left(\alpha_{jto}^2 + \frac{\gamma_o^2}{M^2}\sum_{k\neq j}^{M+1}\alpha_{kto}^2 + \epsilon_{jt}^2\right)\right]$$

Collecting α_{jto} terms and recognizing that $\gamma_o^2/M \leq \gamma_o^2$ implies that:

$$E[b(w)] \leq (2+2M)E\left[(1+\gamma_o^2)\sum_{t=1}^{2}\sum_{j=1}^{M+1}(M+2)\left(\alpha_{jto}^2 + \epsilon_{jt}^2\right)\right]$$

We can take the expectation operator through yielding:

$$E[b(w)] \leq (2+2M)(1+\gamma_o^2)\sum_{t=1}^{2}\sum_{j=1}^{M+1}(M+2)\left[E\left(\alpha_{jto}^2\right) + E\left(\epsilon_{jt}^2\right)\right]$$

Assumptions 3, 4, and 6 ensure that $E(\alpha_{jto}^2)$, $E(\epsilon_{jt}^2)$, and γ_o are all finite. Thus, $E[b(w)] < \infty$.

QED

Proof of Lemma 5.G

To establish asymptotic normality, we now show that the six conditions of Theorem 12.3 in Wooldridge (2002) are satisfied.

1. γ_o must be in the interior of Γ

   This condition is satisfied by assumption 6.

2. Each element of $H(w, \gamma)$ is bounded in absolute value by a function $b(w)$ where $E[b(w)] < \infty$

Recall that $q(w,\gamma)$ be written as:

$$
\begin{aligned}
q(w,\gamma) &= \frac{\left(\sum_{t=1}^{t}\sum_{j=1}^{M+1} W_{jt}y_{jt}\right)^2}{\sum_{t=1}^{t}\sum_{j=1}^{M+1} W_{jt}^2} \\
&= \frac{\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W_{jt}W_{kt'}y_{jt}y_{kt'}}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2}
\end{aligned}
$$

Denoting W'_{jt} as the first partial derivative with respect to γ,

$$
\begin{aligned}
W'_{11} &= [(1+2\gamma-M)(M-1)+1] = -W'_{12} \\
W'_{j1} &= -2\gamma+M = -W'_{j2} \qquad \text{for all } j>1
\end{aligned}
$$

Note W''_{jt}, the second partial derivative of W_{jt} with respect to γ, are all zero

We can then write the score as:

$$
\begin{aligned}
s(w,\gamma) &= \frac{2\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W'_{jt}W_{kt'}y_{jt}y_{kt'}}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2} \\
&- \frac{\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^2}
\end{aligned}
$$

and the hessian as:

$$
\begin{aligned}
H(w,\gamma) &= \frac{2\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W'_{jt}W'_{kt'}y_{jt}y_{kt'}}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2} \\
&- \frac{4\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W'_{jt}W_{kt'}y_{jt}y_{kt'}\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^2} \\
&- \frac{\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W'_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^2} \\
&+ \frac{\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W_{jt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^3}
\end{aligned}
$$

We need to derive a bounding function such that $b(w) \geq |H(w,\gamma)|$ for all $\gamma \in \Gamma$. Note that:

$$
\begin{aligned}
|H(w,\gamma)| \leq & \frac{\left|2\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1}W'_{jt}W'_{kt'}y_{jt}y_{kt'}\right|}{\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2} \\
+ & \frac{\left|4\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1}W'_{jt}W_{kt'}y_{jt}y_{kt'}\right)\right|\left|\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)\right|}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2} \\
+ & \frac{\left|\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1}W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\right|\left|\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W'_{jt}\right)\right|}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2} \\
+ & \frac{\left|\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1}W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\right|\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^3}
\end{aligned}
$$

Repeatedly applying the triangle inequality and collecting terms yields:

$$
\begin{aligned}
|H(w,\gamma)| \leq & \frac{2\,(2+2M)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)y_{jt}^2\right)}{\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2} \\
+ & \frac{4\,(2+2M)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)y_{jt}^2\right)\left|\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)\right|}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2} \\
+ & \frac{(2+2M)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2y_{jt}^2\right)\left|\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}(W'_{jt}W'_{jt})\right)\right|}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2} \\
+ & \frac{(2+2M)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2y_{jt}^2\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^3}
\end{aligned}
$$

Denote W_{jt}^* as the weight given to y_{jt}^2 in the above expression:

$$
\begin{aligned}
W_{jt}^* = & \frac{2\,(2+2M)\left[((W'_{jt})^2+W_{jt}^2\right]}{\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2} \\
+ & \frac{4\,(2+2M)\,(W'^2_{jt}+W_{jt}^2)\left|\left(2\sum_{t=1}^{2}\sum_{k=1}^{M+1}W'_{kt}W_{kt}\right)\right|}{\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2\right)^2} \\
+ & \frac{(2+2M)\,W_{jt}^2\left|\left(2\sum_{t=1}^{2}\sum_{k=1}^{M+1}(W'_{kt}W'_{kt})\right)\right|}{\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2\right)^2} \\
+ & \frac{(2+2M)\left(W_{jt}W_{jt'}y_{jt}\right)\left(2\sum_{t=1}^{2}\sum_{k=1}^{M+1}W'_{kt}W_{kt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2\right)^3}
\end{aligned}
$$

implying that:

$$|H(w,\gamma)| \leq \sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^{*}y_{jt}^{2}$$

Note that W_{jt}^{*} is function only of the class sizes and γ and for any class sizes and γ it is finite. Since the expression on the left hand side of the above equation is increasing in W_{jt}^{*}, define B_{jt}^{*} as:

$$B_{jt}^{*} = \max_{\gamma} W_{jt}^{*}$$

which exists and is finite due to all elements of Γ being finite. Our bounding function is then:

$$b(w) = \sum_{t=1}^{2}\sum_{j=1}^{M+1} B_{jt}^{*}y_{jt}^{2}$$

We then need to establish that $E[b(w)] < \infty$.

$$\begin{aligned} E[b(w)] &= \sum_{t=1}^{2}\sum_{j=1}^{M+1} B_{jt}^{*}E(y_{jt}^{2}) \\ &= \sum_{t=1}^{2}\sum_{j=1}^{M+1} B_{jt}^{*}E\left[\left(\alpha_{jto} + \frac{\gamma_o}{M}\sum_{k\neq j}^{M+1}\alpha_{kto} + \epsilon_{jt}\right)^{2}\right] \end{aligned}$$

Repeatedly using the triangle inequality after expanding the square implies:

$$E[b(w)] \leq \sum_{t=1}^{2}\sum_{j=1}^{M+1} B_{jt}^{*}(M+2)E\left[\alpha_{jto}^{2} + \frac{\gamma_o^{2}}{M^{2}}\sum_{k\neq j}^{M+1}\alpha_{kto}^{2} + \epsilon_{jt}^{2}\right]$$

Collecting α_{jto} terms and recognizing that $\gamma_o^2/M \leq \gamma_o^2$ implies that:

$$E[b(w)] \leq (1+\gamma_o^{2})\sum_{t=1}^{2}\sum_{j=1}^{M+1} B_{jt}^{*}(M+2)E\left[\left(\alpha_{jto}^{2} + \epsilon_{jt}^{2}\right)\right]$$

Assumptions 3, 4, and 6 ensure that B^{*}, $E(\alpha_{jto}^{2})$, $E(\epsilon_{jt}^{2})$, and γ_o are all finite, implying that $E[b(w)] < \infty$.

3. $s(w,\cdot)$ is continuously differentiable on the interior of Γ for all $w \in \mathcal{W}$

   Since $H(w,\gamma)$ is continuous in γ, $s(w,\cdot)$ is continuously differentiable.

4. $A_o \equiv E[H(w,\gamma_o)]$ is positive definite

   With only one parameter, this implies that the Hessian is strictly greater than zero when evaluated at the true γ. To test this condition, we evaluate the expected Hessian at

γ_o. We first note that we can interchange the expectations and the partial derivatives: $E[H(w,\gamma)] = \partial^2 E[q(w,\gamma)]/\partial\gamma^2$. From Lemma 3.G, we know that $E[q(w,\gamma)]$ can be written as:

$$E[q(w,\gamma)]{=}(\gamma-\gamma_o)^2 E\Bigg[\left(\sum_{j=2}^{M+1}(\alpha_{j2o}-\alpha_{j1o})\right)^2 / \sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\Bigg] + E\left(\epsilon^2\right)$$

Note that γ affects two terms: $(\gamma-\gamma_o)^2$ and the denominator. However, because we are going to evaluate the expected Hessian at γ_o, we only need the second derivative of the first term, $(\gamma-\gamma_o)^2$. All of the other partial derivatives will either be multiplied by $(\gamma-\gamma_o)^2$ or $(\gamma-\gamma_o)$, both of which are zero when $\gamma=\gamma_o$. The second derivative of $(\gamma-\gamma_o)^2$ with respect to γ is positive. This second derivative is then multiplied by the expectation of a squared object in the numerator and divided by the sum of squared objects in the denominator. Thus, the expectation of the Hessian evaluated at γ_o is strictly positive.

5. $E[s(w,\gamma_o)] = 0$

   Note that $E[s(w,\gamma)] = \partial E[q(w,\gamma)]/\partial\gamma$. Differentiating with respect to γ leaves terms that are multiplied by $(\gamma-\gamma_o)$ or by $(\gamma-\gamma_o)^2$, implying that if we evaluate the derivative at $\gamma=\gamma_0$ then the expected score is zero.

6. Each element of $s(w,\gamma_o)$ has finite second moment.

$$E[s(w,\gamma)^2] = E\Bigg(\Bigg[\frac{2\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W'_{jt}W_{kt'}y_{jt}y_{kt'}}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2} - \frac{\left(\sum_{t=1}^{2}\sum_{t'=1}^{2}\sum_{j=1}^{M+1}\sum_{k=1}^{M+1} W_{jt}W_{kt'}y_{jt}y_{kt'}\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^2}\Bigg]^2\Bigg)$$

Applying the triangle inequality and collecting terms yields:

$$\begin{aligned} E[s(w,\gamma)^2] \;\; &\leq \;\; E\Bigg(\Bigg[\frac{2\,(2+2M)\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)y_{jt}^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2} \\ &\qquad - \frac{\left((2+2M)\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2 y_{jt}^2\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1} W'_{jt}W_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1} W_{jt}^2\right)^2}\Bigg]^2\Bigg)\end{aligned}$$

Repeatedly applying the triangle inequality we can write:

$$
\begin{aligned}
E[s(w,\gamma)^2] \;\leq\; & E\left(2\left[\frac{2(2+2M)\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)y_{jt}^2}{\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2}\right]^2\right) \\
& + E\left(2\left[\frac{\left((2+2M)\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2y_{jt}^2\right)\left(2\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2}\right]^2\right) \\
\leq\; & E\left(8(2+2M)^2\left[\frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)y_{jt}^2\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2}\right]\right) \\
& + E\left(8(2+2M)^2\left[\frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2y_{jt}^2\right)^2\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^4}\right]\right) \\
\leq\; & E\left(8(2+2M)^3\left[\frac{\sum_{t=1}^{2}\sum_{j=1}^{M+1}((W'_{jt})^2+W_{jt}^2)^2y_{jt}^4}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^2}\right]\right) \\
& + E\left(8(2+2M)^3\left[\frac{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^4y_{jt}^4\right)\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W'_{jt}W_{jt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{j=1}^{M+1}W_{jt}^2\right)^4}\right]\right)
\end{aligned}
$$

Note that the expectation is taken with respect to the y's conditional on γ. Denote W^*_{jt} as the aggregate weight given to y^4_{jt} in the above expression:

$$
W^*_{jt} \;=\; \frac{8(2+2M)^3((W'_{jt})^2+W_{jt}^2)^2}{\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2\right)^2} + \frac{8(2+2M)^3(W_{jt}^4\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W'_{kt}W_{kt}\right)^2}{\left(\sum_{t=1}^{2}\sum_{k=1}^{M+1}W_{kt}^2\right)^4}
$$

where we know that W^*_{jt} is finite as the denominator is greater than zero, M is finite, and γ is finite. Substituting in with W^*_{jt} in the inequality yields:

$$
E[s(w,\gamma)^2] \;\leq\; \sum_{t=1}^{2}\sum_{j=1}^{M+1} W^*_{jt}E(y^4_{jt})
$$

Substituting in for y_{jt} and repeatedly applying the triangle inequality yields:

$$
\begin{aligned}
E[s(w,\gamma)^2] \;\leq\; & \sum_{t=1}^{2}\sum_{j=1}^{M+1} W^*_{jt}E\left[\left(\alpha_{jto}+\frac{\gamma_o}{M}\sum_{k\neq j}^{M+1}\alpha_{kto}+\epsilon_{jt}\right)^4\right] \\
\leq\; & \sum_{t=1}^{2}\sum_{j=1}^{M+1} W^*_{jt}E\left[(M+2)^2\left(\alpha^2_{jto}+\left(\frac{\gamma_o}{M}\right)^2\sum_{k\neq j}^{M+1}\alpha^2_{kto}+\epsilon^2_{jt}\right)^2\right] \\
\leq\; & \sum_{t=1}^{2}\sum_{j=1}^{M+1} W^*_{jt}E\left[(M+2)^3\left(\alpha^4_{jto}+\left(\frac{\gamma_o}{M}\right)^4\sum_{k\neq j}^{M+1}\alpha^4_{kto}+\epsilon^4_{jt}\right)\right]
\end{aligned}
$$

Collecting terms we have:

$$
E[s(w,\gamma)^2] \;\leq\; \sum_{t=1}^{2}\sum_{j=1}^{M+1} W^*_{jt}\left[(M+2)^3\left(1+\frac{\gamma_o^4}{M^3}\right)E\left(\alpha^4_{jto}\right)+E(\epsilon^4_{jt})\right]
$$

W^*_{jt}, γ_o, and M are all finite and since the fourth moments of α and ϵ's are finite by assumptions 3 and 4, the expression is finite for all $\gamma \in \Gamma$ and for all M.

QED $\square$

1.1 Proof of Theorem 1: Accumulation (A2)

Proof. In the case of accumulation, we face the following least squares problem

$$\min_{\alpha^*,\gamma} \sum_{i=1}^{N} \left[(Y_{i1n} - \alpha^*_{i2})^2 + \left(Y_{i2n} - \alpha^*_{i2} - \frac{\gamma}{M_{2n}} \sum_{j \in \mathbb{M}_{2n\sim i}} \alpha^*_{j2} \right)^2 \right] \tag{7}$$

where α^*_{it} is defined in the text as the initial ability plus the accumulated peer effects through period t. Similar to the proof under A1, we illustrate the proof assuming that students are grouped with at most one other student at any point in time and that students are observed for at most two time periods. In addition, within each class there is only one student that is observed for two periods.[2]

Consider the set of students that are observed for two time periods. Each of these students has one peer in period one and one peer in period two. Denote a student block as one student observed for two periods plus his two peers. There are then $\mathcal{N}$ blocks of students, one block for each student observed twice, with three students in each block. Denote the first student in each block as the student who is observed twice where α_{1n} is the *initial* individual effect. The initial individual effect for the first classmate in block n is α_{2n}, while the initial individual effect for the second classmate in block n is α_{3n}.

Define $\alpha^*_{1n} = \alpha_{1n} + \gamma\alpha_{2n}$ and $\alpha^*_{2n} = \alpha_{2n} + \gamma\alpha_{1n}$. The optimization problem is then

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left((y_{11n} - \alpha^*_{1n})^2 + (y_{12n} - \alpha^*_{1n} - \gamma\alpha_{3n})^2 + (y_{2n} - \alpha^*_{2n})^2 + (y_{3n} - \alpha_{3n} - \gamma\alpha^*_{1n})^2 \right)$$

Within each block there are four terms, two residuals for the student observed twice, and a residual for the peer in each period. Our proof then consists of the following five lemmas, each of which is proven later in this appendix.

We first show that the α's can be written as closed form expressions of γ and the data.

[2] See the proof of Theorem 1 for further discussion of these simplifications.

Lemma 1

The vector of unobserved student abilities, α, can be concentrated out of the least squares problem and written strictly as a function of γ and y. Ability for the student in block n observed in both periods is given by

$$\alpha_{1n}^* = \frac{(1+\gamma^2)y_{11n} + (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{31n}}{2-\gamma^2+\gamma^4}$$

while the abilities for the peers in block n are given by

$$\alpha_{2n}^* = y_{2n}$$

and

$$\alpha_{3n} = \frac{-2\gamma y_{11n} + \gamma^3 y_{12n} + (2-\gamma^2)y_{31n}}{2-\gamma^2+\gamma^4}$$

We then show the form of the minimization problem when the α's are concentrated out.

Lemma 2

Concentrating the α's out of the original least squares problem results in an optimization problem over γ that takes the following form

$$\min_{\gamma} \; \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \frac{\left((\gamma^2-1)y_{11n} + y_{12n} - \gamma y_{3n}\right)^2}{2-\gamma^2+\gamma^4}$$

Our nonlinear least squares problem now has only one parameter, γ. We are now in a position to investigate the properties of our estimator of γ_o. For ease of notation, define $q(w,\gamma)$ as

$$q(w,\gamma) = \frac{\left((\gamma^2-1)y_{11} + y_{12} - \gamma y_3\right)^2}{2-\gamma^2+\gamma^4}$$

where $w \equiv \mathbf{y}$. We let $\mathcal{W}$ denote the subset of $\mathbb{R}^4$ representing the possible values of w. Our key result is then Lemma 3, which establishes identification.

Lemma 3

$$E[q(w,\gamma_o)] < E[q(w,\gamma)], \quad \forall\, \gamma \in \Gamma, \;\; \gamma \neq \gamma_o$$

Theorem 12.2 of Wooldridge (2002) establishes that sufficient conditions for consistency are identification and uniform convergence. Having already established identification, Lemma 4 shows uniform convergence.

Lemma 4

$$\max_{\gamma \in \Gamma} \left| \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} q(w_n, \gamma) - E[q(w, \gamma)] \right| \xrightarrow{p} 0$$

Consistency then follows from Theorem 12.2 of Wooldridge: $\hat{\gamma} \xrightarrow{p} \gamma_o$.

Finally, we establish asymptotic normality of $\hat{\gamma}$. Denote $s(w, \gamma_o)$ and $H(w, \gamma_o)$ as the first and second derivative of $q(w, \gamma)$ evaluated at γ_o. Then, Lemma 5 completes the proof.

Lemma 5

$$\sqrt{\mathcal{N}}(\hat{\gamma} - \gamma_o) \xrightarrow{d} N(0, A_o^{-1} B_o A_o^{-1})$$

where

$$A_o \equiv E[H(w, \gamma_o)]$$

and

$$B_o \equiv E[s(w, \gamma_o)^2] = Var[s(w, \gamma_o)]$$

QED.

Proof of Lemma 1

Our objective is to show that the system of equations obtained by differentiating Equation (**??**) with respect to α can be expressed as a series of equations in terms of γ and y, and that these expressions are as given in Lemma 1. Again, conditional on γ, the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Thus, we can work with the system of first-order conditions within one block and then generalize the results to the full system of equations. The first-order condition for α_{1n}^* (student in each block who is observed in both time periods) is given by

$$0 = \frac{-2}{\mathcal{N}} \left[(y_{11n} - \alpha_{1n}^*) + (y_{12n} - \alpha_{1n}^* - \gamma \alpha_{3n}) + \gamma \left(y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}^* \right) \right]$$

while the first-order condition for α_{2n}^* and α_{3n} are respectively given by

$$0 = \frac{-2}{\mathcal{N}} \left(y_{2n} - \alpha_{2n}^* \right)$$

and

$$0 = \frac{-2}{\mathcal{N}} \left[(y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}^*) + \gamma \left(y_{12n} - \alpha_{1n}^* - \gamma \alpha_{3n} \right) \right]$$

Within each block, this yields a system of 3 equations and 3 unknown abilities. The solution for α^*_{2n} is simply given by y_{2n}. The first order condition α_{3n} can be re-arranged such that

$$\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n} - 2\gamma\alpha_{1n}}{1+\gamma^2}$$

and the first order condition for α^*_{1n} can be re-written as

$$\alpha^*_{1n} = \frac{y_{11n} + y_{12n} + \gamma y_{3n} - 2\gamma\alpha_{3n}}{2+\gamma^2}$$

Substituting the equation for α_{3n} into the equation for α^*_{1n} and combining terms yields

$$\alpha^*_{1n} = \frac{(1+\gamma^2)y_{11n} + (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{3n} + 4\gamma^2\alpha^*_{1n}}{(2+\gamma^2)(1+\gamma^2)}$$

Moving all terms containing an α_{1n} to the left hand side of the equation and simplifying results in

$$\alpha^*_{1n}\Big(\frac{2-\gamma^2+\gamma^4}{(2+\gamma^2)(1+\gamma^2)}\Big) = \frac{(1+\gamma^2)y_{11n} + (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{3n}}{(2+\gamma^2)(1+\gamma^2)}$$

Multiplying both sides of the equation by $\frac{(2+\gamma^2)(1+\gamma^2)}{2-\gamma^2+\gamma^4}$ yields the desired result,

$$\alpha^*_{1n} = \frac{(1+\gamma^2)y_{11n} + (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

Substituting the solution for α^*_{1n} into the first order condition for α_{3n} gives

$$\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n}}{1+\gamma^2} - \frac{2\gamma}{1+\gamma^2}\Big[\frac{(1+\gamma^2)y_{11n} + (1-\gamma^2)y_{12n} - \gamma(1-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}\Big]$$

Finding a common denominator and combining like terms yields

$$\alpha_{3n} = \frac{-2\gamma(1+\gamma^2)y_{11n} + \gamma^3(1+\gamma^2)y_{12n} + (2-\gamma^2)(1+\gamma^2)y_{3n}}{(1+\gamma^2)(2-\gamma^2+\gamma^4)}$$

All the $(1+\gamma^2)$ terms cancel, leaving the desired solution,

$$\alpha_{3n} = \frac{-2\gamma y_{11n} + \gamma^3 y_{12n} + (2-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

QED

Proof of Lemma 2

Lemma 1 provides a solution for α strictly as a function of y and γ. We can substitute this solution back into the original optimization problem to derive the result in Lemma 2.

Consider minimizing the sum of squared residuals within a particular block n. There are four residuals within each block, two for the student observed twice, and one each for the corresponding peer. We begin by simplifying the residual for the first observation of the student observed twice, which is given by the expression below

$$e_{11n} = y_{11n} - \alpha^*_{1n}$$

Substituting for α^*_{1n} with the results from Lemma 1 and finding a common denominator results in

$$e_{11n} = \frac{(2-\gamma^2+\gamma^4)y_{11n} - (1+\gamma^2)y_{11n} - (1-\gamma^2)y_{12n} + \gamma(1-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

Combining like terms in the numerator yields

$$e_{11n} = \frac{(1-\gamma^2)^2 y_{11n} - (1-\gamma^2)y_{12n} + \gamma(1-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

Next we want to find an expression for e_{12n}, the residual for individual one in period two, as a function of the data and γ. Recall that

$$e_{12n} = y_{12n} - \alpha^*_{1n} - \gamma\alpha_{3n}$$

Substituting for α^*_{1n} and α_{3n} and finding a common denominator yields

$$e_{12n} = \frac{(2-\gamma^2+\gamma^4)y_{12n} - (1+\gamma^2)y_{11n} - (1-\gamma^2)y_{12n} + \gamma(1-\gamma^2)y_{3n} + 2\gamma^2 y_{11n} - \gamma^4 y_{12n} - \gamma(2-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

Combining like terms yields

$$e_{12n} = \frac{(\gamma^2-1)y_{11n} + y_{12n} - \gamma y_{3n}}{2-\gamma^2+\gamma^4}$$

The residual for individual two is zero since $\alpha^*_{2n} = y_{2n}$ and

$$e_{2n} = y_{2n} - \alpha^*_{2n}$$

The residual for individual three is given by

$$e_{3n} = y_{3n} - \alpha_{3n} - \gamma\alpha^*_{1n}$$

Substituting for α_{3n} and α^*_{1n} from Lemma 1 and finding a common denominator leaves

$$e_{3n} = \frac{(2-\gamma^2+\gamma^4)y_{3n} + 2\gamma y_{11n} - \gamma^3 y_{12n} - (2-\gamma^2)y_{3n} - \gamma(1+\gamma^2)y_{11n} - \gamma(1-\gamma^2)y_{12n} + \gamma^2(1-\gamma^2)y_{3n}}{2-\gamma^2+\gamma^4}$$

Combining like terms yields

$$e_{3n} = \frac{\gamma(1-\gamma^2)y_{11n} - \gamma y_{12n} + \gamma^2 y_{3n}}{2-\gamma^2+\gamma^4}$$

Now we return to the optimization problem. The original optimization problem written as a function of the residuals in each block n takes the following form

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left(e_{11n}^2 + e_{12n}^2 + e_{2n}^2 + e_{3n}^2 \right)$$

We can substitute into the above formulation for each residual using the formulas previously derived. However, a cursory glance at the formulas for e_{11n}, e_{12n}, and e_{3n} reveals that

$$\begin{aligned} e_{11n} &= -(1-\gamma^2)e_{12n} \\ e_{3n} &= -\gamma e_{12n} \end{aligned}$$

Using these relationships along with the fact that $e_{2n} = 0$, we can re-write the least squares problem as

$$\min_{\alpha,\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left((2-\gamma^2+\gamma^4)e_{12n}^2 \right)$$

Substituting in with our solution for e_{12n} yields

$$\min_{\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \left((2-\gamma^2+\gamma^4) \frac{\left((\gamma^2-1)y_{11n} + y_{12n} - \gamma y_{3n}\right)^2}{(2-\gamma^2+\gamma^4)^2} \right)$$

Canceling terms results in the following optimization problem

$$\min_{\gamma} \frac{1}{\mathcal{N}} \sum_{n=1}^{\mathcal{N}} \frac{\left((\gamma^2-1)y_{11n} + y_{12n} - \gamma y_{3n}\right)^2}{2-\gamma^2+\gamma^4}$$

QED

Proof of Lemma 3

The population objective function as a function of γ is given by

$$E[q(w,\gamma)] = E\left[\frac{\left((\gamma^2-1)y_{11} + y_{12} - \gamma y_3\right)^2}{2-\gamma^2+\gamma^4} \right]$$

Substituting for y with the data generating process, canceling the appropriate terms and combining like terms in the numerator leaves

$$E[q(w,\gamma)] = E\left[\frac{\Big((\gamma^2-\gamma\gamma_o)\alpha_{1o}+(\gamma^2\gamma_o-\gamma\gamma_o^2)\alpha_{2o}+(\gamma_o-\gamma)\alpha_{3o}+(\gamma^2-1)\epsilon_{11}+\epsilon_{12}-\gamma\epsilon_{13}\Big)^2}{2-\gamma^2+\gamma^4}\right]$$

Assuming that the ϵ's are uncorrelated with the α'sm we can re-write the above as

$$\begin{aligned} E[q(w,\gamma)] &= E\left[\frac{\Big((\gamma^2-\gamma\gamma_o)\alpha_{1o}+(\gamma^2\gamma_o-\gamma\gamma_o^2)\alpha_{2o}+(\gamma_o-\gamma)\alpha_{3o}\Big)^2}{2-\gamma^2+\gamma^4}\right] \\ &\quad +E\left[\frac{\Big((\gamma^2-1)\epsilon_{11}+\epsilon_{12}-\gamma\epsilon_{13}\Big)^2}{2-\gamma^2+\gamma^4}\right] \end{aligned}$$

Assuming that the ϵ's are uncorrelated with each other

$$\begin{aligned} E[q(w,\gamma)] &= E\left[\frac{\Big((\gamma^2-\gamma\gamma_o)\alpha_{1o}+(\gamma^2\gamma_o-\gamma\gamma_o^2)\alpha_{2o}+(\gamma_o-\gamma)\alpha_{3o}\Big)^2}{2-\gamma^2+\gamma^4}\right] \\ &\quad +E\left[\frac{(\gamma^2-1)^2\epsilon_{11}^2+\epsilon_{12}^2+\gamma^2\epsilon_{13}^2}{2-\gamma^2+\gamma^4}\right] \end{aligned}$$

Factoring out a $\gamma-\gamma_o$ from the first term leaves

$$\begin{aligned} E[q(w,\gamma)] &= E\left[\frac{(\gamma-\gamma_o)^2\Big(\gamma\alpha_{1o}+\gamma\gamma_o\alpha_{2o}+\alpha_{3o}\Big)^2}{2-\gamma^2+\gamma^4}\right] \\ &\quad +E\left[\frac{(\gamma^2-1)^2\epsilon_{11}^2+\epsilon_{12}^2+\gamma^2\epsilon_{13}^2}{2-\gamma^2+\gamma^4}\right] \end{aligned}$$

Finally, assumption 5 implies we can express the above equation as:

$$\begin{aligned} E[q(w,\gamma)] &= E\left[\frac{(\gamma-\gamma_o)^2\Big(\gamma\alpha_{1o}+\gamma\gamma_o\alpha_{2o}+\alpha_{3o}\Big)^2}{2-\gamma^2+\gamma^4}\right] \\ &\quad +E[\epsilon^2] \end{aligned}$$

For any $\gamma \neq \gamma_o$ the first term is always positive and the second term is not a function of γ.

QED.

Proof of Lemma 4

Uniform convergence, requires that

$$\max_{\gamma\in\Gamma}\left|\frac{1}{\mathcal{N}}\sum_{n=1}^{\mathcal{N}} q(w_n,\gamma)-E[q(w,\gamma)]\right| \xrightarrow{p} 0$$

Theorem 12.1 in Wooldridge states four conditions that the data and q must satisfy in order for the above condition to hold.

1. Γ is compact

   This condition is satisfied by assumption 6.

2. For each $\gamma\in\Gamma$, $q(\cdot,\gamma)$ is Borel measurable on $\mathcal{W}$

   Since $q(\cdot,\gamma)$ is a continuous function of w, it is also Borel measurable.

3. For each $w\in\mathcal{W}$, $q(w,\cdot)$ is continuous on Γ

   Our concentrated objective function is continuous in γ.

4. $|q(w,\gamma)| \le b(w)$ for all $\gamma\in\Gamma$, where b is a nonnegative function on $\mathcal{W}$ such that $E[b(w)]<\infty$

   Note that $q(w,\gamma)$ is always positive so we can ignore the absolute value. We derive a bounding function $b(w)$ in the following manner

$$\begin{aligned} q(w,\gamma) &= \frac{\left((\gamma^2-1)y_{11}+y_{12}-\gamma y_3\right)^2}{2-\gamma^2+\gamma^4} \\ &= \frac{(\gamma^2-1)^2y_{11}^2+y_{12n}^2+\gamma^2y_3^2+2(\gamma^2-1)y_{11}y_{12}-2\gamma(\gamma^2-1)y_{11}y_{3n}-2\gamma y_{12}y_3}{2-\gamma^2+\gamma^4} \\ &\le \frac{3(\gamma^2-1)^2y_{11}^2+3y_{12}^2+3\gamma^2y_3^2}{2-\gamma^2+\gamma^4} \end{aligned}$$

   where the last line follows from the triangle inequality. Since $\gamma^4-\gamma^2$ is always greater than -1, the following inequality must also be satisfied:

$$q(w,\gamma)\le 3(\gamma^4+1)y_{11}^2+3y_{12}^2+3\gamma^2y_3^2$$

   Let $\overline{\gamma}$ and $\underline{\gamma}$ denote the largest and smallest elements of the set Γ and denote $\gamma^*=\max\{\overline{\gamma},-\underline{\gamma}\}$. Our bounding function is then:

$$b(w)=3(\gamma^{*4}+1)y_{11}^2+3y_{12}^2+3\gamma^{*2}y_3^2$$

We now show that $E[b(w)] < \infty$, completing the proof. Since γ^* is finite, this amounts to establishing that $E[y_{11}^2]$, $E[y_{12}^2]$, and $E[y_3^2]$ are all finite. First consider $E[y_{12}^2]$:

$$E[y_{12}^2] = E[(\alpha_{1o} + \gamma_o\alpha_{2o} + \gamma_o\alpha_{3o} + \epsilon_{12})^2]$$

Repeatedly applying the triangle inequality yields:

$$E[y_{12}^2] \leq 4(E[\alpha_{1o}^2] + \gamma_o^2 E[\alpha_{2o}^2] + \gamma_o^2 E[\alpha_{3o}^2] + E[\epsilon_{12}^2])$$

Assumption 3 and 4 ensure that all the terms of the right hand side are finite. By a similar argument, it can be shown that all the terms in $E[b(w)]$ are finite.

QED

Proof of Lemma 5

Theorem 12.3 in Wooldridge(2002) states six conditions that must hold in order for $\hat{\gamma}$ to be distributed asymptotically normal.

Many of these conditions involve the first and second derivatives of $q(w,\gamma)$. The first and second derivatives are given by:

$$\begin{aligned} s(w,\gamma) &= 2\Big[\left(-3\gamma + 2\gamma^3 + \gamma^5\right) y_{11}^2 + \left(\gamma - 2\gamma^3\right) y_{12}^2 + \left(2\gamma - \gamma^5\right) y_3^2 + \left(2\gamma + 4\gamma^3 - 2\gamma^5\right) y_{11}y_{12} \\ &\quad + \left(2 - 5\gamma^2 - 2\gamma^4 + \gamma^6\right) y_{11}y_{12} + \left(-2 - \gamma^2 + 3\gamma^4\right) y_{12}y_3 \Big] \Big/ (2 - \gamma^2 + \gamma^4)^2 \end{aligned}$$

$$\begin{aligned} H(w,\gamma) &= -2\Big[(6 - 3\gamma^2 - 33\gamma^4 + 11\gamma^6 + 3\gamma^8)y_{11}^2 + (-2 + 9\gamma^2 + 9\gamma^4 - 10\gamma^6)y_{12}^2 \\ &\quad -(4 + 6\gamma^2 - 24\gamma^4 + \gamma^6 + 3\gamma^8)y_3^2 - 2(2 + 15\gamma^2 - 15\gamma^4 - 9\gamma^6 + 3\gamma^8)y_{11}y_{12} \\ &\quad +2\gamma(6 + 21\gamma^2 - 21\gamma^4 - 3\gamma^6 + \gamma^8)y_{11}y_3 + 2\gamma(6 - 19\gamma^2 - 3\gamma^4 + 6\gamma^6)y_{12}y_3\Big] \Big/ (2 - \gamma^2 + \gamma^4)^3 \end{aligned}$$

We now show that the six conditions of Theorem 12.3 in Wooldridge(2002) are satisfied. We will refer to the above formulations of the score and Hessian throughout.

1. γ_o must be in the interior of Γ

   This condition is satisfied by assumption 6.

2. $s(w,\cdot)$ is continuously differentiable on the interior of Γ for all $w \in \mathcal{W}$

   Since $H(w,\gamma)$ is continuous in γ, $s(w,\cdot)$ is continuously differentiable.

3. Each element of $H(w,\gamma)$ is bounded in absolute value by a function $b(w)$ where $E[b(w)] < \infty$

   Taking absolute values through and noting that the denominator is always greater than one implies:

$$\begin{aligned} |H(w,\gamma)| &\leq 2\Big[(6+3\gamma^2+33\gamma^4+11\gamma^6+3\gamma^8)y_{11}^2+(2+9\gamma^2+9\gamma^4+10\gamma^6)y_{12}^2 \\ &+(4+6\gamma^2+24\gamma^4+\gamma^6+3\gamma^8)y_3^2+2(2+15\gamma^2+15\gamma^4+9\gamma^6+3\gamma^8)|y_{11}y_{12}| \\ &+2|\gamma|(6+21\gamma^2+21\gamma^4+3\gamma^6+\gamma^8)|y_{11}y_3|+2|\gamma|(6+19\gamma^2+3\gamma^4+6\gamma^6)|y_{12}y_3|\Big] \end{aligned}$$

   Applying the triangle inequality to the last three terms and collecting terms yields:

$$\begin{aligned} |H(w,\gamma)| &\leq 2\Big[(8+18\gamma^2+48\gamma^4+20\gamma^6+6\gamma^8+|\gamma|[6+21\gamma^2+21\gamma^4+3\gamma^6+\gamma^8])y_{11}^2 \\ &+(4+24\gamma^2+24\gamma^4+19\gamma^6+3\gamma^8+|\gamma|[6+19\gamma^2+3\gamma^4+6\gamma^6])y_{12}^2 \\ &+(4+6\gamma^2+24\gamma^4+\gamma^6+3\gamma^8+|\gamma|[12+30\gamma^2+24\gamma^4+9\gamma^6+\gamma^8])y_3^2\Big] \end{aligned}$$

   Our bounding function can then be found by setting γ to γ^* where $\gamma^* = \max\{-\underline{\gamma}, \overline{\gamma}\}$:

$$\begin{aligned} b(w) &= 2\Big[(8+18\gamma^{*2}+48\gamma^{*4}+20\gamma^{*6}+6\gamma^{*8}+\gamma^*[6+21\gamma^{*2}+21\gamma^{*4}+3\gamma^{*6}+\gamma^{*8}])y_{11}^2 \\ &+(4+24\gamma^{*2}+24\gamma^{*4}+19\gamma^{*6}+3\gamma^{*8}+|\gamma|[6+19\gamma^{*2}+3\gamma^{*4}+6\gamma^{*6}])y_{12}^2 \\ &+(4+6\gamma^{*2}+24\gamma^{*4}+\gamma^{*6}+3\gamma^{*8}+\gamma^*[12+30\gamma^{*2}+24\gamma^{*4}+9\gamma^{*6}+\gamma^{*8}])y_3^2\Big] \end{aligned}$$

   Since γ^* is finite and we have already established that $E[y_{11}^2]$, $E[y_{12}^2]$, and $E[y_3^2]$ are all finite, then $E[b(w)] < \infty]$, completing the proof.

4. $A_o \equiv E[H(w,\gamma_o)]$ is positive definite

   We first note that we can interchange the expectations and the partial derivatives: $E[H(w,\gamma)] = \partial^2 E[q(w,\gamma)]/\partial\gamma^2$. From Lemma 3, we know that $E[q(w,\gamma)]$ can be written as

$$E[q(w,\gamma)] = \frac{(\gamma-\gamma_o)^2 E\left[(\gamma\alpha_{1o}+\gamma\gamma_{2o}\alpha_{2o}-\alpha_3)^2\right]}{2-\gamma^2+\gamma^4} + E[\epsilon^2]$$

   Note that γ operates through three terms: $(\gamma-\gamma_o)^2$ in the numerator, through the expectation in the numerator, and the denominator. However, because we are going to evaluate the expected Hessian at γ_o, we only need the second derivative of the first term,

$(\gamma - \gamma_o)^2$. All of the other partial derivatives will either be multiplied by $(\gamma - \gamma_o)^2$ or $(\gamma - \gamma_o)$, both of which are zero when $\gamma = \gamma_o$. The second derivative of $(\gamma - \gamma_o)^2$ with respect to γ is positive. This second derivative is then multiplied by the expectation of a squared object in the numerator and divided by the sum of squared objects in the denominator. Thus, the expectation of the Hessian evaluated at γ_o is strictly positive.

5. $E[s(w, \gamma_o)] = 0$

   Note that $E[s(w, \gamma)] = \partial E[q(w, \gamma)]/\partial \gamma$. Differentiating $E[q(w, \gamma)]$ with respect to γ leaves terms that are multiplied by $(\gamma - \gamma_o)$ or by $(\gamma - \gamma_o)^2$, implying that if we evaluate the derivative at $\gamma = \gamma_o$ then the expected score is zero.

6. Each element of $s(w, \gamma_o)$ has finite second moment.

   Given that the score has only one element, this condition boils down to $E[s(w, \gamma_o)^2] < \infty$. To show this we square the score function and evaluate at the true γ.

$$\begin{aligned} E\left[s(w,\gamma_o)^2\right] &= E\Big[\left(-6\gamma_o + 4\gamma_o^3 + 2\gamma_o^5\right) y_{11}^2 + \left(2\gamma_o - 4\gamma_o^3\right) y_{12}^2 + \left(4\gamma_o - 2\gamma_o^5\right) y_3^2 + \left(4\gamma_o + 8\gamma_o^3 - 4\gamma_o^5\right) y_{11}y_{12} \\ &\quad + \left(4 - 10\gamma_o^2 - 4\gamma_o^4 + 2\gamma_o^6\right) y_{11}y_{13} + \left(-4 - 2\gamma_o^2 + 6\gamma_o^4\right) y_{12}y_3\Big]^2 \Big/ (2 - \gamma_o^2 + \gamma_o^4)^4 \end{aligned}$$

Note that the denominator is greater than one. Further, making all terms in the numerator positive will result in an increase in the left hand side:

$$\begin{aligned} E\left[s(w,\gamma_o)^2\right] &\leq E\Big[|\gamma_o|\left(6 + 4\gamma_o^2 + 2\gamma_o^4\right) y_{11}^2 + |\gamma_o|\left(2 + 4\gamma_o^2\right) y_{12}^2 + |\gamma_o|\left(4 + 2\gamma_o^4\right) y_3^2 \\ &\quad + |\gamma_o|\left(4 + 8\gamma_o^2 + 4\gamma_o^4\right) |y_{11}y_{12}| + \left(4 + 10\gamma_o^2 + 4\gamma_o^4 + 2\gamma_o^6\right) |y_{11}y_{13}| \\ &\quad + \left(4 + 2\gamma_o^2 + 6\gamma_o^4\right) |y_{12}y_3|\Big]^2 \end{aligned}$$

Applying the triangle inequality to remove the cross-y terms and collecting terms yields:

$$\begin{aligned} E\left[s(w,\gamma_o)^2\right] &\leq E\Big[\left(2 + 5\gamma_o^2 + 2\gamma_o^4 + \gamma_o^6 + |\gamma_o|\left[8 + 8\gamma_o^2 + 4\gamma_o^4\right]\right) y_{11}^2 \\ &\quad + \left(2 + \gamma_o^2 + 3\gamma_o^4 + |\gamma_o|\left[4 + 8\gamma_o^2 + 2\gamma_o^4\right]\right) y_{12}^2 \\ &\quad + \left(4 + 6\gamma_o^2 + 5\gamma_o^4 + \gamma_o^6 + |\gamma_o|\left[4 + 2\gamma_o^4\right]\right) y_3^2\Big]^2 \end{aligned}$$

Label the coefficients on y_{11}^2, y_{12}^2, and y_3^2 as A, B, and C, where all are positive and finite:

$$E\left[s(w,\gamma_o)^2\right] \leq E\left[Ay_{11}^2 + By_{12}^2 + Cy_3^2\right]^2$$

Taking the square through and again applying the triangle inequality yields:

$$E\left[s(w,\gamma_o)^2\right] \quad \leq \quad 3A^2 E\left[y_{11}^4\right] + 3B^2 E\left[y_{12}^4\right] + 3C^2 E\left[y_3^4\right]$$

With A, B, and C positive and finite, we now need to establish $E\left[y_{11}^4\right]$, $E\left[y_{12}^4\right]$, and $E\left[y_3^4\right]$ are finite. Consider $E\left[y_{12}^4\right]$ and substitute in for the DGP:

$$E\left[y_{12}^4\right] = E\left[(\alpha_{1o} + \gamma_o\alpha_{2o} + \gamma_o\alpha_{3o} + \epsilon_{11})^2\right]$$

Repeatedly applying the triangle inequality yields

$$\begin{aligned} E\left[y_{12}^4\right] &\leq 16E\left[\left(\alpha_{1o}^2 + \gamma_o^2\alpha_{2o}^2 + \gamma_o^2\alpha_{3o}^2 + \epsilon_{11}^2\right)^2\right] \\ &\leq 64\left(E\left[\alpha_{1o}^4\right] + \gamma_o^4 E\left[\alpha_{2o}^4\right] + \gamma_o^4 E\left[\alpha_{3o}^4\right] + E\left[\epsilon_{11}^4\right]\right) \end{aligned}$$

Assumptions 3 and 4 ensure that all of the terms on the right hand side of the inequality in the above equation are finite. Thus, $E[y_{12}^4]$ is finite. By a similar argument, it can be shown that all the terms in the expectation of the squared score are finite.

QED $\square$