A Online Appendix to Large Scale Ideal Point Estimation

Michael Peress, SUNY-Stony Brook

A.1 Computing Starting Values for Bridged Chambers

Suppose that we are bridging together J chambers connected by bridge voters. My approach first estimates ideal points for each of the J chambers individually. Let $\hat{\alpha}_{jn}$ denote the ideal point of individual n in chamber j. Let I^{α}_{jn} be equal to 1 if individual n is in chamber j and 0 otherwise. The estimates for each chamber will be on a different scale. To pool this information together, I transform these chamber estimates onto a common scale. I denote the transformation parameters by $\tilde{C}_j$ and $\tilde{d}_j$ and I denote the (commonly-scaled) starting values by $\tilde{\alpha}_n$.

My approach solves the optimization problem,

$$(\tilde{\alpha}, \tilde{C}, \tilde{d}) = \underset{\alpha, C, d}{\arg\min} \sum_{n=1}^{N} \sum_{j=1}^{J} I^{\alpha}_{nj} (\hat{\alpha}_{nj} - C_j \alpha_n - d_j)^2 \tag{33}$$

The parameters $\tilde{C}_j$ and $\tilde{d}_j$ as chosen such that the chamber-by-chamber estimates are as close as possible to being linear transformations of the commonly-scaled starting values. This approach resembles the dynamic ideal point adjustment approach of Groseclose, Levitt and Snyder (1999), the linear mapping approach of Shor and McCarty (2011), and the "black box" approach of Poole (1998), but differs in some key ways. Applying Shor and McCarty (2011)'s linear mapping approach will generally produce good starting values in the case where a single chamber is connected through bridge voters to every other chamber. We would like to have an algorithm that works more generally however. For example, in connected sessions of the U.S. Senate, no single session contains bridge voters to every other session. In addition, in Shor and McCarty (2011), the information is not pooled in the same way—my estimates make use of all available information to estimate the starting values.

Groseclose, Levitt and Snyder (1999) come closer to developing an approach that can yield starting values for dynamic Senate estimates, but there are two limitations of applying their approach directly. First, the goal of their approach is to obtain transformed chamber-by-chamber estimates while we would like to obtain a single estimate for each bridge voter so that those bridge voters can be used to connect that scales across different chambers. Second, Groseclose, Levitt and Snyder's approach involves an iterative solver which itself can get stuck in a local minimum. Since my goal is to obtain starting values to prevent the penalized maximum likelihood estimator from getting stuck in a local maximum, employing a starting value algorithm that itself can get stuck without good starting values would be problematic.[13] Instead, the objective function I suggest is quadratic—meaning the minimum can be computed by solving a sparse linear system of equations—avoiding the need for an iterative nonlinear solver. Poole's (1998) objective function is quadratic, but he nonetheless employs an iterative solver similar to the one Groseclose, Levitt and Snyder employ.

The objective function specified above does not have a unique minimum and the commonly-scaled ideal points $\tilde{\alpha}$ can be linearly transformed without affecting the value of the objective function. I have found two effective approaches approaches for guaranteeing a unique minimum. The first is to choose a chamber j^* and to constrain the transformation to be the identity transformation for this chamber, i.e. $\tilde{C}_{j^*} = I$ and $\tilde{d}_{j^*} = 0$. This approach tends to be effective when there is a single chamber that is connected to most other chambers, as is the case in Shor and McCarty (2011). This approach tends to perform poorly when some chambers are only connected to some other chambers through a very long sequence of con-

[13]To be clear, this should not be read as a criticism of Groseclose, Levitt and Snyder—their approach was effective on their intended application.

nections, as is the case when connecting the 1st congress to the 113th congress using bridge voters. The second approach is to penalize the objective function above for deviations from $C_j = I$ and $d_j = 0$ for all j, which is more effective when some chambers are only connected through a long sequence of connections.[14]

While the approach above will generate starting values for α in the case where all the chambers are connected by bridge voters, we need to have an approach that also works when some chambers are only connected by bridge votes. My approach works as follows—I first identify all clusters of chambers that are connected by bridge voters and apply the above approach for generating starting values of α. Then, for each cluster of chambers, I estimate starting values for a and b using vote-specific probits. Each of these chamber clusters will be connected by bridge votes, which will be characterized by a and b estimated on different scales. For each chamber-cluster j, let $I^{\delta}_{jt} = 1$ if vote t occurs in chamber-cluster j. To transform these onto the same scale, I solve,

$$(\tilde{a}, \tilde{b}, \tilde{C}, \tilde{d}) = \underset{a,b,C,d}{\arg\min} \sum_{t=1}^{T} \sum_{j=1}^{J} I^{\delta}_{tj} [(\hat{a}_{tj} - a_t + b_t'(C_j^{-1})' d_j)^2 + (\hat{b}_{tj} - C_j^{-1} b_t)^2] \tag{34}$$

where the form of the objective function is motivated from equations 27 and 28. To obtain a quadratic objective function, this problem is re-parameterized as,

$$(\tilde{a}, \tilde{b}, \tilde{E}, \tilde{f}) = \underset{a,b,E,f}{\arg\min} \sum_{t=1}^{T} \sum_{j=1}^{J} I^{\delta}_{tj} [(\hat{a}_{tj} - a_t + b_t' f_j)^2 + (\hat{b}_{tj} - E_j b_t)^2] \tag{35}$$

I then use the linear transformations given by $\tilde{C}_j = \tilde{E}_j^{-1}$ and $\tilde{d} = \tilde{E}_j^{-1} \tilde{f}_j$ to transform the chamber-cluster values of α onto the common scale. Then (as before), I adjust the scale of

[14]This approach favors orientations which don't flip the signs of the ideal points. Before this approach is applied, I normalize the chamber-by-chamber ideal points to have mean zero and variance equal to the identity matrix. I then flip the signs of the ideal points along each dimension by applying a singular value decomposition to the ideal points and then flip the sign of the ideal points whenever the SVD weights have a negative sign. It is possible to avoid this pre-processing step if the penalty terms are formed differently. In the one-dimensional case, my approach penalized deviations from $C_j = 1$. One could instead penalize deviations from $C_j^2 = 1$ or $|C_j| = 1$, but each of these would lead to a non-quadratic objective function.

the starting values (α, a, b) to match that scale of the prior using equation 29.

One question to ask is whether this two stage approach is necessary—can α and (a, b) be transformed in one step rather than two? While it is possible to form such an objection function, the objective function will not be quadratic. Absent the quadratic form, the starting value optimization problem may be vulnerable to the same problem of local solutions, making the approach less than helpful.

A.2 Proof of Proposition 1

For an $m \times n$ matrix A, define the matrix norm $\|A\|_{\max} = \max_{1 \leq i \leq m, 1 \leq j \leq n} |A_{ij}|$. Note that,

$$\tfrac{1}{T} H = \begin{pmatrix} \frac{1}{T} H_{11} & \frac{1}{T} H_{12} \\ \frac{1}{T} H_{21} & \frac{1}{T} H_{22} \end{pmatrix} \tag{36}$$

and define,

$$\tfrac{1}{T} H_0 = \begin{pmatrix} \frac{1}{T} H_{11} & 0 \\ 0 & \frac{1}{T} H_{22} \end{pmatrix} \tag{37}$$

To prove the result, I make the following assumptions:

Assumption A.1. $\alpha_n \in \mathrm{A}$ *for all* n *and* $(a_t, b_t) \in \Delta$ *for all* t*, where* A *and* Δ *are bounded.*

Assumption A.2. $N, T \to \infty$ *with* $\frac{N}{T} \to \kappa$ *where* $\kappa > 0$*.*

Assumption A.1 assumes that the parameters live in a bounded set and is standard when deriving theoretical properties of nonlinear estimators. Assumption A.2 assumes that both the number of individuals and the number of items are growing and also assumes that they grow at the same rate.

The result is given below:

Proposition A.1. *Suppose that Assumptions A.1 and A.2 hold. Then* $\left\|\frac{1}{T}(H - H_0)\right\|_{\max} \xrightarrow{prob.} 0$.

Proof. Define,

$$z_{nt} = 1\{y_{nt} = 2\}\frac{\Phi(a_t + b_t'\alpha_n)\phi'(a_t + b_t'\alpha_n) - \phi(a_t + b_t'\alpha_n)^2}{\Phi(a_t + b_t'\alpha_n)^2} \tag{38}$$

$$+ 1\{y_{nt} = 1\}\frac{-(1 - \Phi(a_t + b_t'\alpha_n))\phi'(a_t + b_t'\alpha_n) - \phi(a_t + b_t'\alpha_n)^2}{(1 - \Phi(a_t + b_t'\alpha_n))^2} \tag{39}$$

$$w_{nt} = 1\{y_{nt} = 2\}\frac{\phi(a_t + b_t'\alpha_n)}{\Phi(a_t + b_t'\alpha_n)} - 1\{y_{nt} = 1\}\frac{-\phi(a_t + b_t'\alpha_n)}{1 - \Phi(a_t + b_t'\alpha_n)} \tag{40}$$

and note that $-1 < z_{nt} < 0$. We have that,

$$H_{11,nn} = \sum_{t=1}^{T} z_{nt} b_t b_t' \tag{41}$$

$$H_{22,tt} = \frac{1}{N}\sum_{n=1}^{N} z_{nt}(1, \alpha_n)(1, \alpha_n)' \tag{42}$$

$$H_{12,nt} = z_{nt} b_t (1, \alpha_n)' + (0, w_{nt}, ..., w_{nt}) \tag{43}$$

$$H_{21} = H_{12}' \tag{44}$$

Note that,

$$\frac{1}{T}H - \frac{1}{T}H_0 = \begin{pmatrix} 0 & \frac{1}{T}H_{12} \\ \frac{1}{T}H_{21} & 0 \end{pmatrix} \tag{45}$$

The elements of H involve bounded random variables (y_{nt}) multiplied by continuous functions, which by Assumption A.1, take values on a bounded set. It follows that there exists a $C > 0$ such that $|H_{ij}| \leq C$ for all i and j. We have that,

$$\left\|\tfrac{1}{T}H - \tfrac{1}{T}H_0\right\|_{\max} = \left\|\begin{pmatrix} 0 & \tfrac{1}{T}H_{12} \\ \tfrac{1}{T}H_{21} & 0 \end{pmatrix}\right\|_{\max} = \tfrac{1}{T}\left\|\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix}\right\|_{\max} \leq \tfrac{1}{T}C \tag{46}$$

Setting $\delta = \frac{1}{T}C$, we have that,

$$\Pr\left(\left\|\tfrac{1}{T}H - \tfrac{1}{T}H_0\right\|_{\max} > \delta\right) = 0 \tag{47}$$

Therefore, it follows that for all $\varepsilon > 0$ and $\delta > 0$, there exists an N and T large enough so that $\Pr\left(\left\|\frac{1}{T}H - \frac{1}{T}H_0\right\|_{\max} > \delta\right) < \varepsilon$, proving the result.

□

A.3 Additional Theoretical Result

In Appendix A.2, I provided conditions under which the Hessian converges to a block diagonal form in large samples. Standard errors would usually be calculated by inverting the projected Hessian (where the projection handles the fact that parameter constraints must be imposed in order to achieve identification). Here, I provide conditions under which the inverse of the projected Hessian converges to the inverse of the block diagonal approximation of the projected Hessian.

Before presenting the next result, I introduce some useful notation and definitions. For an $m \times n$ matrix A, define the matrix norms $\|A\|_{\max} = \max\limits_{1 \leq i \leq m, 1 \leq j \leq n} |A_{ij}|$, $\|A\|_1 = \max\limits_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}|$, $\|A\|_\infty = \max\limits_{1 \leq i \leq n} \sum_{j=1}^n |A_{ij}|$, and $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}(A)$ denotes the largest singular value of A. Note that all four matrix norms satisfy the triangular inequality, but only $\|.\|_1$,

$\|.\|_\infty$, and $\|.\|_2$ are sub-multiplicative, i.e. $\|AB\|_1 \leq \|A\|_1 \|B\|_1$, $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$, and $\|AB\|_2 \leq \|A\|_2 \|B\|_2$ for conformant matrices A and B. If A is symmetric negative definite, then $\|A\|_2 = -\lambda_{\min}(A)$ where $\lambda_{\min}(A)$ is the minimum eigenvalue of A. Note that $\|A\|_1 \leq \sqrt{m}\,\|A\|_2$, $\|A\|_1 \leq n\,\|A\|_{\max}$, $\|A\|_{\max} \leq \|A\|_2$, and if A is symmetric, $\|A\|_1 = \|A\|_\infty$ (Golub and Van Loan, 1996).

In the proof below, I make use of the following result.

Lemma A.1. $\|AB\|_{\max} \leq \|A\|_{\max} \|B\|_1$ *and* $\|AB\|_{\max} \leq \|A\|_\infty \|B\|_{\max}$

Proof. $\|AB\|_{\max} = \max\limits_{i,j} |[AB]_{ij}| = \max\limits_{i,j} |\sum_k A_{ik}B_{kj}| \leq \max\limits_{i,j} \sum_k |A_{ik}| * |B_{kj}| \leq \max\limits_{i,j,l} \sum_k |A_{il}| * |B_{kj}| = \|A\|_{\max} \max\limits_{j} \sum_k |B_{kj}| = \|A\|_{\max} \|B\|_1$. Similar logic implies the second result. $\square$

For symmetric matrices, it is also the case that $\|AB\|_{\max} \leq \|A\|_1 \|B\|_{\max}$ and $\|AB\|_{\max} \leq \|A\|_{\max} \|B\|_\infty$. Throughout, let $S = ND + T(D+1)$ denote the dimension of H.

Consider now the projected Hessian which would be inverted in order to yield the asymptotic variance matrix. Consider a set of constraints on α of the form $C\alpha = d$ necessary to achieve identification, where d has $D(D+1)$ rows. For simplicity, I consider the case where $D = 1$. There exists a basis for the space of α satisfying the linear equation. Let w be the vectors that make up a basis. There exists a projection matrix Z such that $\{\alpha : C\alpha = d\} = \{\alpha : \alpha = Zw + \theta, w \in \mathbb{R}^{N-2}\}$. The projected Hessian is given by $\tilde{H} = \begin{bmatrix} Z' & 0 \\ 0 & I \end{bmatrix} H \begin{bmatrix} Z & 0 \\ 0 & I \end{bmatrix}$ and the projected version of H_0 is given by $\tilde{H}_0 = \begin{bmatrix} Z' & 0 \\ 0 & I \end{bmatrix} H_0 \begin{bmatrix} Z & 0 \\ 0 & I \end{bmatrix}$. We choose a basis such that $Z = \begin{bmatrix} I \\ \tau \end{bmatrix}$. We partition $H_{11} = \begin{bmatrix} H_{11pp} & 0 \\ 0 & H_{11qq} \end{bmatrix}$ and $H_{12} = \begin{bmatrix} H_{12p} & H_{12q} \end{bmatrix}$ in which case,

$$\tilde{H} = \begin{bmatrix} I & \tau' & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} H_{11pp} & 0 & H_{12p} \\ 0 & H_{11qq} & H_{12q} \\ H'_{12p} & H'_{12q} & H_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ \tau & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} H_{11pp} + \tau' H_{11qq}\tau & H_{12p} + \tau' H_{12q} \\ H'_{12p} + H'_{12q}\tau & H_{22} \end{bmatrix}$$

$$\tilde{H}_0 = \begin{bmatrix} I & \tau' & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} H_{11pp} & 0 & 0 \\ 0 & H_{11qq} & 0 \\ 0 & 0 & H_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ \tau & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} H_{11pp} + \tau' H_{11qq}\tau & 0 \\ 0 & H_{22} \end{bmatrix}$$

Further suppose that the columns of τ, τ_n, have $\tau_n \in \left\{ \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$, where $\frac{1}{N}\sum_{n=1}^{N} 1\{\tau_n = \begin{bmatrix} -1 \\ 0 \end{bmatrix}\} \to p_1 > 0$ and $\frac{1}{N}\sum_{n=1}^{N} 1\{\tau_n = \begin{bmatrix} 0 \\ -1 \end{bmatrix}\} \to p_2 > 0$ as $N \to \infty$. This would be the case if the constraints $C\alpha = d$ took the form of setting the means of two groups equal to fixed values, with the proportion in each of those two groups being bounded away from zero as the sample size increases.

Let $\lambda_i(A)$ denote the eigenvalues of matrix A, let $\lambda_{\min}(A)$ denote the minimum eigenvalue of A, and let $\lambda_{\max}(A)$ denote the maximum eigenvalue of A. Recall that for a symmetric matrix with non-zero eigenvalues, the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A. This means that for a negative definite matrix, $\lambda_{\min}(A^{-1}) = \frac{1}{\lambda_{\max}(A)}$ and $\lambda_{\max}(A^{-1}) = \frac{1}{\lambda_{\min}(A)}$.

Assumption A.3. *There exist $C_2 > C_1 > 0$ such that $-C_2 \leq \lambda_i(\frac{1}{T}H_0) \leq -C_1$ and $\lambda_i(\frac{1}{T}\tilde{H}) \leq -C_1$ for all i with probability approaching 1.*

Proposition A.2. *Suppose that Assumptions A.1 through A.3 hold. Then $\left\| T(\tilde{H}^{-1} - \tilde{H}_0^{-1}) \right\|_{\max} \xrightarrow{prob.}$*

0.

Proof. We start by noting that,

$$T\tilde{H}^{-1} - T\tilde{H}_0^{-1} = T\tilde{H}^{-1}(\tfrac{1}{T}\tilde{H}_0 - \tfrac{1}{T}\tilde{H})T\tilde{H}_0^{-1} \tag{48}$$

Applying various norm properties,

$$\left\|T\tilde{H}^{-1} - T\tilde{H}_0^{-1}\right\|_{\max} = \left\|T\tilde{H}^{-1}(\tfrac{1}{T}\tilde{H}_0 - \tfrac{1}{T}\tilde{H})T\tilde{H}_0^{-1}\right\|_{\max} \tag{49}$$

$$\leq \left\|\tfrac{1}{T}\tilde{H}_0 - \tfrac{1}{T}\tilde{H}\right\|_{\max}\left\|T\tilde{H}^{-1}\right\|_1\left\|T\tilde{H}_0^{-1}\right\|_1 ‘$$

$$\leq \left\|\tfrac{1}{T}\tilde{H}_0 - \tfrac{1}{T}\tilde{H}\right\|_{\max}\sqrt{S}\left\|T\tilde{H}^{-1}\right\|_2\left\|T\tilde{H}_0^{-1}\right\|_1 = \left\|\tfrac{\sqrt{S}}{T}\tilde{H}_0 - \tfrac{\sqrt{S}}{T}\tilde{H}\right\|_{\max}\left\|T\tilde{H}^{-1}\right\|_2\left\|T\tilde{H}_0^{-1}\right\|_1$$

Note that,

$$\left\|\tfrac{\sqrt{S}}{T}\tilde{H}_0 - \tfrac{\sqrt{S}}{T}\tilde{H}\right\|_{\max} = \frac{\sqrt{S}}{T}\left\|\begin{bmatrix} 0 & H_{12p} + \tau' H_{12q} \\ H'_{12p} + H'_{12q}\tau & 0 \end{bmatrix}\right\|_{\max} \tag{50}$$

Using a similar argument to the one used in the proof of Proposition A.1, we have that $|H_{ij}| \leq C_3$ for $C_3 \geq 0$. Note that, $[\tau' H_{12q}]_{ij} = \sum_k \tau_{ki}[H_{12q}]_{kj}$. Since the number of rows of τ is 2 and the elements of τ are either zero or negative one, we have that $|[\tau' H_{12q}]_{ij}| \leq 2C_3$ from which we obtain $|[H_{12p} + \tau' H_{12q}]_{ij}| \leq 3C_3$. We therefore have $\left\|\frac{\sqrt{S}}{T}\tilde{H}_0 - \frac{\sqrt{S}}{T}\tilde{H}\right\|_{\max} \leq \frac{\sqrt{S}}{T}3C_3$.

Next, applying Assumption A.3 and the properties of eigenvalues of negative definite matrices, we have,

$$\left\|T\tilde{H}^{-1}\right\|_2 = -\lambda_{\max}(T\tilde{H}^{-1}) = \frac{1}{-\lambda_{\min}(\frac{1}{T}\tilde{H})} \leq \tfrac{1}{C_1} \tag{51}$$

with probability approaching 1.

Finally, we consider,

$$\left\|T\tilde{H}_0^{-1}\right\|_1 = \left\| \begin{bmatrix} T(H_{11pp} + \tau' H_{11qq}\tau)^{-1} & 0 \\ 0 & TH_{22}^{-1} \end{bmatrix} \right\|_1 \tag{52}$$

$$\leq \max\left\{\left\|T(H_{11pp} + \tau' H_{11qq}\tau)^{-1}\right\|_1, \left\|TH_{22}^{-1}\right\|_1\right\}$$

We have,

$$\left\|TH_{22}^{-1}\right\|_1 = \max_{1\leq t\leq T} \left\|TH_{22tt}^{-1}\right\|_1 \leq \sqrt{2} \max_{1\leq t\leq T} \left\|H_{22tt}^{-1}\right\|_2 = \sqrt{2} \max_{1\leq t\leq T}(-\lambda_{\min}(TH_{22tt}^{-1})) \tag{53}$$

$$= \sqrt{2} \max_{1\leq t\leq T} \frac{1}{(-\lambda_{\max}(\frac{1}{T}H_{22tt}))} \leq \frac{\sqrt{2}}{C_1}$$

with probability approaching 1. Using the Woodbury matrix identity (Golub and Van Loan, 1996), we can write,

$$T(H_{11pp} + \tau' H_{11qq}\tau)^{-1} = TH_{11pp}^{-1} - TH_{11pp}^{-1}\tau'(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau')^{-1}\tau TH_{11pp}^{-1} \tag{54}$$

Taking norms,

$$\left\|T(H_{11pp} + \tau' H_{11qq}\tau)^{-1}\right\|_1 = \left\|TH_{11pp}^{-1} - TH_{11pp}^{-1}\tau'(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau')^{-1}\tau TH_{11pp}^{-1}\right\|_1 \tag{55}$$

$$\leq \left\| TH_{11pp}^{-1} \right\|_1 + \left\| TH_{11pp}^{-1}\tau'(H_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau')^{-1}\tau TH_{11pp}^{-1} \right\|_1$$

$$\leq \left\| TH_{11pp}^{-1} \right\|_1 + \left\| TH_{11pp}^{-1} \right\|_1^2 \left\| \tau'(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau')^{-1}\tau \right\|_1$$

Using an argument similar to the one we provided for $\left\| TH_{22}^{-1} \right\|_1$, we have $\left\| TH_{11pp}^{-1} \right\|_1 \leq \frac{1}{C_1}$ with probability approaching 1. We then have,

$$\left\| \tau'(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau)^{-1}\tau \right\|_1 \leq N \left\| \tau'(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau)^{-1}\tau \right\|_{\max} \tag{56}$$

$$= N \max_{m,n} \left\| \tau_m(TH_{11qq}^{-1} + \tau TH_{11pp}^{-1}\tau)^{-1}\tau_n' \right\|_{\max} = N \max_{m,n} \left\| \tau_m(TH_{11qq}^{-1} + \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1}\tau_n' \right\|_{\max}$$

$$\leq N \max_{m,n} \left\| \tau_m \right\|_1 \left\| \tau_n \right\|_1 \left\| (TH_{11qq}^{-1} + \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1} \right\|_{\max} = N \left\| (TH_{11qq}^{-1} + \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1} \right\|_{\max}$$

$$= \left\| (\tfrac{1}{N} TH_{11qq}^{-1} + \tfrac{1}{N} \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1} \right\|_{\max} \leq \left\| (\tfrac{1}{N} TH_{11qq}^{-1} + \tfrac{1}{N} \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1} \right\|_2$$

$$= -\lambda_{\min}\left((\tfrac{1}{N} TH_{11qq}^{-1} + \tfrac{1}{N} \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r)^{-1} \right) = \frac{1}{-\lambda_{\max}\left(\frac{1}{N} TH_{11qq}^{-1} + \frac{1}{N} \sum_{r=1}^{N-2} \tau_r' TH_{11rr}^{-1}\tau_r \right)}$$

Define $R_{r1} = 1\left\{\tau_r = \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right\}$ and $R_{r2} = 1\left\{\tau_r = \begin{bmatrix} 0 \\ -1 \end{bmatrix}\right\}$. We have,

$$\tfrac{1}{N} TH^{-1}_{11qq} + \tfrac{1}{N}\sum_{r=1}^{N-2} \tau_r' TH^{-1}_{11rr}\tau_r \tag{57}$$

$$= \begin{bmatrix} \frac{1}{N} TH^{-1}_{11,N-1,N-1} + \frac{1}{N}\sum_{r=1}^{N-2} R_{r1} TH^{-1}_{11rr} & 0 \\ 0 & \frac{1}{N} TH^{-1}_{11,N,N} + \frac{1}{N}\sum_{r=1}^{N-2} R_{r2} TH^{-1}_{11rr} \end{bmatrix}$$

Using the fact that for diagonal matrices, the eigenvalues are equal to the diagonal elements,

$$-\lambda_{\max}\left(\tfrac{1}{N} TH^{-1}_{11qq} + \tfrac{1}{N}\sum_{r=1}^{N-2} \tau_r' TH^{-1}_{11rr}\tau_r\right) \tag{58}$$

$$= -\max\left\{\tfrac{1}{N} TH^{-1}_{11,N-1,N-1} + \tfrac{1}{N}\sum_{r=1}^{N-2} R_{r1} TH^{-1}_{11rr}, \tfrac{1}{N} TH^{-1}_{11,N-1,N-1} + \tfrac{1}{N}\sum_{r=1}^{N-2} R_{r2} TH^{-1}_{11rr}\right\}$$

$$= \min\left\{\tfrac{1}{N}(-TH^{-1}_{11,N-1,N-1}) + \tfrac{1}{N}\sum_{r=1}^{N-2} R_{r1}(-TH^{-1}_{11rr}), \tfrac{1}{N}(-TH^{-1}_{11,N-1,N-1}) + \tfrac{1}{N}\sum_{r=1}^{N-2} R_{r2}(-TH^{-1}_{11rr})\right\}$$

We can bound,

$$-TH^{-1}_{11rr} = -\lambda_i(TH^{-1}_{11rr}) = \frac{-1}{\lambda_i(\frac{1}{T}H_{11rr})} \geq \frac{1}{C_2} \tag{59}$$

with probability approaching 1, so that,

$$-\lambda_{\max}\left(\tfrac{1}{N} TH^{-1}_{11qq} + \tfrac{1}{N}\sum_{r=1}^{N-2} \tau_r' TH^{-1}_{11rr}\tau_r\right) \tag{60}$$

$$\geq \frac{1}{C_2} \frac{\min\left\{1 + \sum_{r=1}^{N-2} R_{r1}, 1 + \sum_{r=1}^{N-2} R_{r2}\right\}}{N}$$

and,

$$\frac{1}{-\lambda_{\max}\left(\frac{1}{N} T H_{11qq}^{-1} + \frac{1}{N} \sum_{r=1}^{N-2} \tau_r' T H_{11rr}^{-1} \tau_r\right)} \tag{61}$$

$$\leq \frac{C_2}{\frac{1}{N} \min\left\{1 + \sum_{r=1}^{N-2} R_{r1}, 1 + \sum_{r=1}^{N-2} R_{r2}\right\}}$$

Note that by assumption,

$$\frac{1}{N} \min\left\{1 + \sum_{r=1}^{N-2} R_{r1}, 1 + \sum_{r=1}^{N-2} R_{r2}\right\} \to \min\{p_1, p_2\} > 0 \tag{62}$$

as $N \to \infty$. Continuity of division when the denominator is not zero means that for each $\gamma > 0$, there exists an N, T large enough so that,

$$\left\|\tau'(T H_{11qq}^{-1} + \tau T H_{11pp}^{-1} \tau)^{-1} \tau\right\|_1 \leq \frac{C_2}{\min\{p_1, p_2\}} + \gamma \tag{63}$$

Under these conditions, we have $\|T(H_{11pp} + \tau' H_{11qq} \tau)^{-1}\|_1 \leq \frac{1}{C_1} + \frac{1}{C_1^2} \frac{C_2}{\min\{p_1, p_2\}} + \frac{1}{C_1^2}\gamma$ with probability approaching 1. Combined, these results yield,

$$\left\|T \tilde{H}_0^{-1}\right\|_1 \leq \max\left\{\frac{\sqrt{2}}{C_1}, \frac{1}{C_1} + \frac{1}{C_1^2} \frac{C_2}{\min\{p_1, p_2\}} + \frac{1}{C_1^2}\gamma\right\} \tag{64}$$

so that for each $\gamma > 0$, there exists an N, T large enough so that,

$$\left\|T \tilde{H}^{-1} - T \tilde{H}_0^{-1}\right\|_{\max} \leq \frac{\sqrt{S}}{T} \frac{3C_3}{C_1} \max\left\{\frac{\sqrt{2}}{C_1}, \frac{1}{C_1} + \frac{1}{C_1^2} \frac{C_2}{\min\{p_1, p_2\}} + \frac{1}{C_1^2}\gamma\right\} \tag{65}$$

Under Assumption A.2, we have that the right-hand side can be made arbitrarily small by increasing N and T. Selecting $\delta = \frac{\sqrt{S}}{T}\frac{3C_3}{C_1}\max\left\{\frac{\sqrt{2}}{C_1}, \frac{1}{C_1} + \frac{1}{C_1^2}\frac{C_2}{\min\{p_1,p_2\}} + \frac{1}{C_1^2}\gamma\right\}$, for all $\varepsilon > 0$, there exists an N, T large enough so that $\Pr\left(\left\|T\tilde{H}^{-1} - T\tilde{H}_0^{-1}\right\|_{\max} \leq \delta\right) \geq 1 - \varepsilon$, proving the result.

$\square$