A linear regression model for non-oriented dyadic data with interactive individual effects

Yassine Sbai Sassi *

September 2023

Click here for to the most up to date version.

Abstract

We propose a two step rate optimal estimator for an undirected dyadic linear regression model with interactive unit-specific effects. The estimator remains consistent when the individual effects are additive rather than interactive. We observe that the unit-specific effects alter the eigenvalue distribution of the data's matrix representation in significant and distinctive ways. We offer a correction for the *ordinary least squares*' objective function to attenuate the statistical noise that arises due to the individual effects, and in some cases, completely eliminate it. The new objective function is similar to the *least squares* estimator's objective function from the large N large T literature (Bai (2009)). In general, the objective function is ill behaved and admits multiple local minima. Following a novel proof strategy, we show that in the presence of interactive effects, an iterative process in line with Bai (2009)'s converges to a global minimizer and is asymptotically normal when initiated properly. The new proof strategy suggests a computationally more advantageous and asymptotically equivalent estimator. While the iterative process does not converge when the individual effects are additive, we show that the alternative estimator remains consistent for all slope parameters.

*Department of Economics, University of California - Berkeley, e-mail: yassine@berkeley.edu. I am thankful to Bryan Graham, Michael Jansson, Demian Pouzo and Kévin Dano for helpful conversations. All errors are mine.

Introduction

Linear regression models with individual-specific effects are widely used to fit data with network structures. Such linear models were used to explain trade flows between countries (Anderson and van Wincoop (2003), Fally (2015)), to fit matched employer-employee data (Abowd et al. (1999), Bonhomme et al. (2019)), or to study teacher effects on student performance (Jackson et al. (2014)), to mention a few examples. In applications, these linear regression models are most often used with a particular specification of the individual-specific effect. A popular specification consists of including the individual effects additively. This is the approach taken for instance in Abowd et al. (1999) and Jackson et al. (2014). Broadly, three types of estimators are used under this specification. The two way fixed effects estimator (Abowd et al. (1999)) exploits the additivity of the model to eliminate the individual effects. After double differencing, the initial model is turned into a regular linear regression model (free of the individual effects), and estimators are obtained by least squares on the transformed model. When the data is non bipartite with a number N of agents (respectively, when it is bipartite, with N and M agents on each side), the two way fixed estimator of the slope parameters converges at the optimal rate of N (resp. $\sqrt{NM}$). The two way fixed estimator comes with a significant caveat: the slope parameters on any agent-specific observable covariates disappear in the double differencing process, in the same way as the individual effects. Those can be recovered in a second stage by ordinary least squares if we further assume the individual effects to be exogenous with respect to the additive observable attributes. The second stage OLS estimators for the slope parameters on the additive covariates converges at a $\sqrt{N}$ rate.

A second approach appeals to the standard OLS estimator (e.g. Rose (2004), Fafchamps and Gubert (2007)). In the dyadic linear regression setting, the OLS estimator is in general $\sqrt{N}$ consistent for all the parameters. Given that for some covariates the two way fixed effects estimator can provide N-consistent estimators, the OLS estimator is severely inefficient.

Other approaches consist of estimating a fixed effects model, by regressing the output variable on the covariates, individual indicators and interactions of individual interactions. Examples abound in the large N large T panel data literature. Bai (2009), Moon and Weidner (2015) and Moon and Weidner (2017) study the *least squares* (LS) estimator, obtained by treating the individual and time effects as nuisance parameters estimated by minimizing the squared errors. The least squares estimator is shown to converge at the optimal $\sqrt{NT}$ rate. However, the LS estimator is obtained by minimizing the objective function over $K+T+N$ parameters (K being the dimension of the slope parameter, N and T the dimensions of the cross-sectional and time effects), which poses computational challenges. Bai (2009) proposes an iterative minimization routine that is guaranteed to converge to a stationary point. However, the objective function can be ill-behaved, potentially admitting multiple stationary points. Moon and Weidner (2023) propose an iterative process that returns an estimator that is asymptotically equivalent to the LS estimator after just 2 iterations, when initiated with a consistent but potentially rate inferior estimator. Each of the iterations requires the resolution of a high dimensional minimization problem.

This paper studies symmetric non-oriented network regression models with interactive effects. We propose a modification over the *ordinary least squares* estimator's objective function to obtain a new low dimensional objective function. We exploit the matrix structure of network data and identify the individual effects' footprint on the spectrum of the output matrix. We then correct for the unobservables' effect on the spectrum. We show that the estimator obtained through the minimization of the new objective function has a similar asymptotic behavior to the LS estimator from the large N large T panel data literature.[1]

We propose an iterative process to solve the new minimization problem. Following Sargan (1964)'s standard argument, the iterative process is guaranteed to converge to a stationary point. However, the new objective function can have multiple local minima. We show that if the iterative process is initiated by a consistent but potentially rate inferior estimator, then the iterations converge to a global minimizer. We study the asymptotic behavior of that specific global minimizer.

Interestingly, we show that in theory, no finite number of iterations is enough to jump from the inferior initial rate of convergence to the optimal N rate. To escape the inferior rate, the number of iterations ought to be indexed by the sample size, which is computationally problematic. Building on our results on the distribution of a single iteration estimator, we propose an equivalent estimator that only requires 1 iteration when the initial estimator is $\sqrt{N^2}$ consistent, or 2 iterations when the initial estimator is $\sqrt{N}$ consistent, substantially reducing the computational burden.

Throughout the paper, an initial $\sqrt{N}$- consistent estimator is assumed to be available. In the context of dyadic (network) regression, the individual effects are generally assumed to be centered and independent from the observable regressors (e.g. Graham (2020), Graham et al. (2021)). When that is the case, the OLS estimator is $\sqrt{N}$-consistent and is a good candidate for the initiation phase. Generally, for dyadic data, one can always extract an i.i.d. subsample of a size of order N (for instance by only keeping the observations with indices

[1] The language in this statement is kept intentionally loose. The usual assumptions in the large N large T panel data literature (e.g. Bai (2009), Moon and Weidner (2015), Moon and Weidner (2017), etc) exclude the network models that this paper studies. However, the proofs in that literature can be marginally modified to obtain results on the LS estimator in our context. Later in this paper (especially following corollary 3), we precise in what sense the estimator proposed in this paper compares to the LS estimator.

$\{1,2\}$, $\{3,4\}$, ..., $\{N-1,N\}$, which are i.i.d. observations since no index appears more than once), then employ whatever cross-sectional estimation procedure is suitable for the context at hand (for instance, using an instrumental variable for the unobservable effects on the i.i.d. subsample). This would typically yield a $\sqrt{N}$-consistent estimator. The results in the paper are more general in the sense that they allow for an arbitrary correlation between the regressors and the individual effects, as long as an initial estimator is available.

The next section introduces the setup and lays out the main intuitions leading up to the definition of the new estimator. Section 3 discusses the estimator's theoretical properties and numerical implementation. Section 3 proposes estimators for the covariance matrix. Section 5 examines the asymptotic distribution of the alternative estimator in a specification without interactive effects. Finally, section 6 shows the results from Monte Carlo simulations and from an empirical illustration and the last section concludes. All proofs are deffered to the end of the paper (appendix A).

1 Minimizing the least eigenvalues: definitions and main results

Consider the model:

$$Y_{ij} = X_{ij}'\beta_0 + \gamma(A_i + A_j) + \delta(A_i \times A_j) + V_{ij} \tag{1}$$

for all $i \neq j$, where A_i's are i.i.d centered random variables with finite fourth moments. The V_{ij}'s are i.i.d centered square integrable random variables with $V_{ij} = V_{ji}$, $\beta_0 := (\beta_{0,1}, .., \beta_{0,L})$ is the parameter of interest and $\gamma \geq 0$ and $\delta \in \{-1, 0, +1\}$ are unknown nuisance parameters.[2] The covariates X are such that for all i, j, l such that $i \neq j$: $X_{ij,l} = X_{ji,l} = \phi(X_i, X_j, W_{ij})$, for some (possibly unknown) function ϕ, i.i.d random variables X_i and $i.i.d.$ variables W_{ij}. By convention, $X_{ii,l} = 0$ for all i and l and the first covariate is the intercept (i.e. $X_{ij,1} = 1$ for all $i \neq j$). [TBC: see how i should present the correlation] X_{ij} is arbitrarily correlated with A_i and A_j, but independent from V_{ij}.

When $\delta \neq 0$, the model in equation (1) can also be re-expressed:

$$Y_{ij} = \sum_{l=1}^{L} \beta_{0,l} X_{ij,l} - \delta\gamma^2 + \delta(A_i + \gamma)(A_j + \gamma) + V_{ij} \tag{2}$$

[2] γ is set to be positive because the model (1) can also be re-expressed as $Y_{ij} = X_{ij}\beta + (-\gamma)((-A_i) + (-A_j)) + \delta \times (-A_i) \times (-A_j) + V_{ij}$. The sign of γ is not identified.

which reduces the study of the model (1) to that of:

$$Y_{ij} = \sum_{l=1}^{L} \mu_{0,l} X_{ij,l} + \delta U_i U_j + V_{ij} \tag{3}$$

where the errors $U_i := \gamma + A_i$ are no longer assumed to be centered. All the slope parameters remain unchanged as you move from model (1) to (2) (or (3)), only the intercept is altered by the correction term "$-\delta\gamma^2$" in equation (2). Therefore, any "good" estimators for the parameters of the model (3) also provide good estimators for the parameters in models (1) and (2), except perhaps for their intercepts. Because the intercept is shifted by $-\delta\gamma^2$ when we move from the original model (1) to model (3), the OLS estimator of the intercept would need to be corrected to account for the shift. That is done in Proposition 7 and relegated to the appendix.

Let N be the sample size (number of nodes or agents i). Denote Y and V the $N \times N$ matrices with entries Y_{ij}, V_{ij} and X_l the matrix with entries $X_{ij,l}$ for every $l = 1 \ldots L$. Y and X_l's diagonal entries are equal to zero. V's ith diagonal term is equal to $\delta(E(U_1^2) - U_i^2)$. Finally, stack the individual random effects into a vector denoted U. This allows for the formulation of model 3 in a compact matrix form :

$$Y = \sum_{l=1}^{L} \mu_{0,l} X_l + \delta U U' + V - \delta E(U_1^2) I_N \tag{4}$$

I_N being the identity matrix of dimension N. Let $M(\mu)$ be the matrix of residuals corresponding to μ:

$$M(\mu) := Y - \sum_{l=1}^{L} \mu_l X_l = \sum_{l=1}^{L} (\mu_{0,l} - \mu_l) X_l + \delta U U' + V - \delta E(U_1^2) I_N \tag{5}$$

For any $N \times N$ matrix M, $Tr(M)$ denotes M's trace, $\lambda_1(M) \geq \lambda_2(M) \geq \ldots \lambda_N(M)$ are M's eigenvalues ranked from largest to smallest. We study the estimator that minimizes the objective function

$$g_N(\mu) := \sum_{i=2}^{N} \lambda_i \left(M(\mu)^2 \right) \tag{6}$$

that is, the sum of $M(\mu)^2$'s $N-1$ smallest eigenvalues; g_N is a modification over the ordinary least squares' objective function, observe:

$$g_N(\mu) = \sum_{i=2}^{N} \lambda_i \left(M(\mu)^2 \right) = Tr\left(M(\mu)^2 \right) - \lambda_1 \left(M(\mu)^2 \right) = \sum_{i,j} \left(Y_{ij} - \sum_{l=1}^{L} \mu_l X_{ij,l} \right)^2 - \lambda_1 \left(M(\mu)^2 \right)$$

and note that the OLS estimator minimizes the sum of squared residuals $\sum_{i,j}\left(Y_{ij}-\sum_{l=1}^{L}\mu_l X_{ij,l}\right)^2$. The OLS estimator is efficient when the interactive term U_iU_j is absent from equation (3), and as we discuss further later in this paper, the interactive term mostly impacts $M(\mu)$'s largest eigenvalue, once μ is *close enough* to the true μ_0. The new objective function g_N mechanically removes the largest eigenvalue, the one baring most of UU' impact on the sum of squared errors (c.f. section 2 for a detailed discussion).

Moon and Weidner (2015) show that in the setting of the large N large T panel regression model (or similarly, in the context of oriented dyadic linear regression models), the objective function in (7) is obtained from the objective function of the *least squares* estimator. In our context, however, the two objective functions are different because of the zeros on the diagonal of the matrix $M(\mu)$ in (5).[3]

The problem of minimizing g_N can't be solved in closed form and the function g_N is in general ill behaved (globally not smooth and potentially admitting multiple local minima). Following Bai (2009), we propose an iterative process and show that it converges to a global minimizer when initiated properly. In studying the minimizer of the objective function (7), we take a different route than the common route in the panel data literature. Bai (2009) offers the iterative process as a practical method of minimizing g_N, but studies the asymptotic properties of the minimizer independently of how it is obtained in practice. The function g_N can admit multiple stationary points and the iterations are not guaranteed to converge to the intended argmin. To illustrate this point, figure (1) shows a plot of the function g_N for the model $Y_{ij}=\mu_0+U_iU_j+V_{ij}$; $\sigma_U=\sigma_V=1$, $E(U)=1$, $N=100$ and $\mu_0=1$.

In this paper, we study the global minimizer specifically by analysing the effect of individual iterations, then combining the effects of successive iterations. In addition to guaranteeing convergence to the desired minimizer (conditional on proper initialization), this proof strategy also offers the simple shortcuts at the origin of our equivalent and computationally more efficient alternative estimator. Let f_N be the function

$$f_N:\mu\rightarrow\left(\sum_{i\neq j}X'_{ij}X_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i(\mu)\nu_j(\mu)X'_{jk}X_{ik}\right)^{-1}\left(\sum_{i\neq j}X'_{ij}Y_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i(\mu)\nu_j(\mu)X'_{jk}Y_{ik}\right)$$

[3]The LS objective function: $S_N(\mu,U) := \frac{1}{N(N-1)}\sum_{i\neq j}(Y_{ij}-\sum_k\mu_kX_{ij,k}-U_iU_j)^2 = \frac{1}{N(N-1)}\sum_{i,j}(Y_{ij}-\sum_k\mu_kX_{ij,k}-U_iU_j)^2-\frac{\sum_i U_i^4}{N(N-1)}$, and Moon and Weidner (2015) show that $\frac{g_N(\mu)}{N(N-1)} = arg\min_U \frac{1}{N(N-1)}\sum_{i,j}(Y_{ij}-\sum_k\mu_kX_{ij,k}-U_iU_j)^2$. Even assuming U is uniformly bounded: $\frac{\sum_i U_i^4}{N(N-1)}=O_p\left(\frac{1}{N}\right)$ (uniformly), extending standard results (e.g. Arcones (1998)) , we would need a $O_p\left(\frac{1}{N^2}\right)$ error to obtain the asymptotic equivalence of the minimizers of both objective functions.

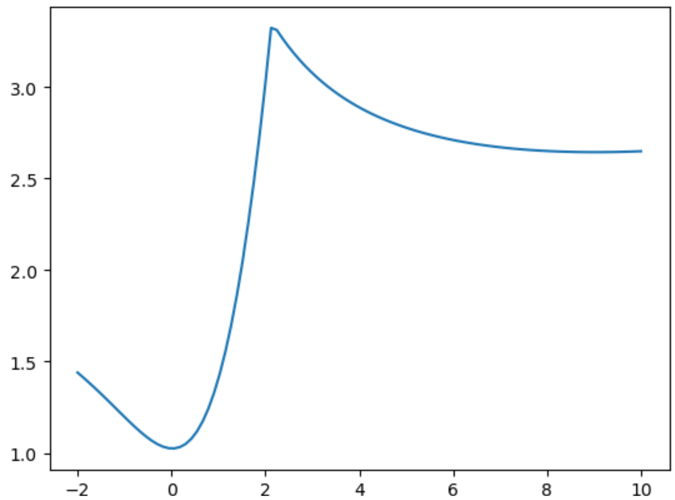

Figure 1: *The graph of the function g_N for the model $Y_{ij} = \mu_0 + U_iU_j + V_{ij}$; $\sigma_U = \sigma_V = 1$, $E(U) = 1$, $N = 100$ and $\mu_0 = 1$. The values of μ are on the X-axis, and the corresponding $f_N(\mu)$ is on the Y-axis.*

Lemma 1. *Assume $E(X_{12}'X_{12})$ is invertible. With probability approaching 1, the problem*

$$\min_{\mu\in\mathbb{R}^L} \sum_{i=2}^{N} \lambda_i \left(M(\mu)^2\right) \tag{7}$$

admits a solution for N large enough. Moreover, μ^* is a minimizer of (7) if and only if it is a solution to the fixed point problem:

$$\mu = f_N(\mu) \tag{8}$$

where $\nu(\mu)$ is the normalized $\left(||\nu(\mu)||_2 = 1\right)$ eigenvector of $M(\mu)$ corresponding to $M(\mu)'s$ largest eigenvalue.

Proof. See section A.4. □

The condition on $E(X_{12}'X_{12})$ is a standard non-collinearity condition. In addition to guaranteeing the existence of a solution, Lemma 1 provides a practical tool to study the behavior of estimators obtained through the optimization problem (7). Intuitively, equation (8) is a first order condition of a minimization problem that is equivalent to (7). Let

$\mu^* \in \arg\min_{\mu\in\mathbb{R}^L} \sum_{i=2}^N \lambda_i \left(M(\mu)\right)^2$ and note

$$
\begin{aligned}
&\mu^* \in \arg\min_{\mu\in\mathbb{R}^L} \sum_{i=2}^N \lambda_i \left(M(\mu)\right)^2 \\
&\qquad\qquad \iff \mu^* \in arg\min_{\mu} \sum_{i\neq j} \left(Y_{ij} - \sum_{l=1}^L \mu_l X_{ij,l}\right)^2 - \max_{\nu:||\nu||=1} \nu' M(\mu)^2 \nu \\
&\qquad\qquad \Rightarrow \mu^* \in arg\min_{\mu} \sum_{i\neq j} \left(Y_{ij} - \sum_{l=1}^L \mu_l X_{ij,l}\right)^2 - \nu(\mu^*)' M(\mu)^2 \nu(\mu^*);\ \text{where } \nu(\mu) \in \arg\max_{\nu:||\nu||=1} \nu' M(\mu)^2 \nu \\
&\qquad\qquad \Rightarrow \mu^* \in arg\min_{\mu} \left(Y_{ij} - \sum_{l=1}^L \mu_l X_{ij,l}\right)^2 - \sum_{i,j,k\neq i,j} \nu_i(\mu^*)\nu_j(\mu^*) \left(Y_{ik} - \sum_{l=1}^L \mu_l X_{ik,l}\right) \left(Y_{kj} - \sum_{l=1}^L \mu_l X_{kj,l}\right)
\end{aligned}
$$

The last equality allows for the expression of μ^* as the minimizer of a smooth and convex function over $\mathbb{R}^L$ (in fact, strictly convex with probability 1, when N is large enough):

$$
\mu \to \sum_{i\neq j} \left(Y_{ij} - \sum_{l=1}^L \mu_l X_{ij,l}\right)^2 - \sum_{i,j,k\neq i,j} \nu_i(\mu^*)\nu_j(\mu^*) \left(Y_{ik} - \sum_{l=1}^L \mu_l X_{ik,l}\right) \left(Y_{kj} - \sum_{l=1}^L \mu_l X_{kj,l}\right)
$$

the first order condition results in the fixed point problem (8). The proof in section A.4 closely follows this sketch.

Lemma 1 does not guarantee the uniqueness of the solution to the minimization problem (7). The iteration process just described, when it converges, could converge to one of many potential fixed point of (8) (solutions to (7)). Additionally, the iteration process could be explosive, leading the iterations to diverge rather than approach one of the fixed points. The function f_N is generally ill-behaved. In general, it is neither convex, nor quasi-convex, nor differentiable. Figure 2 illustrates f_N's behavior for the simplest model nested in model (3): $Y_{ij} = \mu_0 + U_i U_j + V_{ij}$ for $\sigma_U = \sigma_V = 1$, $\mu_0 = 1$ and for $N = 100$. In this example, f_N is convex between ≈ -0.5 and ≈ 2, it has a point of inflexion, smoothly switching convexity at ≈ -0.5. f_N is not differentiable at ≈ 2. However, f_N has a unique minimum (on the interval displayed in figure 2), that is close to the true parameter $\mu = \mu_0 = 1$. The figure also points to the direction that the results in the sequel will follow: I show that with high probability, f_N is well behaved in a shrinking neighborhood of μ_0; μ_0 being unknown, knowledge of a *good enough* first stage estimator will be essential throughout the paper. In particular, we study the estimator defined defined in (7) by studying single successive iterations on the fixed point problem (8). It turns out that when the iteration process is initiated with a *good* first stage estimator, e.g. OLS when the individual effects are assumed to be independent

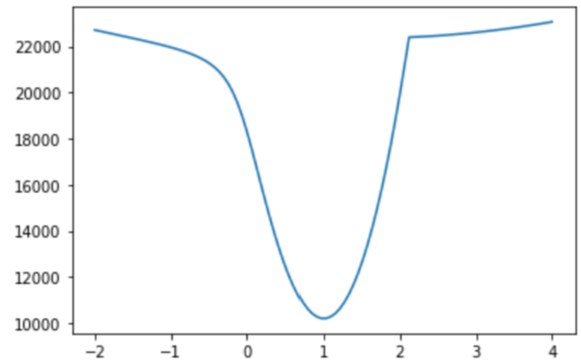

Figure 2: *The graph of the function f_N for the model $Y_{ij} = \mu + U_i U_j + V_{ij}$; $\sigma_U = \sigma_V = 1$, $N = 100$. The values of μ are on the X-axis, and the corresponding $f_N(\mu)$ is on the Y-axis.*

of the observable covariate, the process converges to a fixed point or a minimizer (formal statements are presented in Corollary 2 in the following section).

The estimator(s) studied in this paper are obtained by iterating equation (8), that is, by plugging some "reasonable" initial estimator in the right hand side of (8) to obtain what we show is a more precise estimator on the left hand side, then iterating this process as needed until the true fixed point distribution is achieved. We are ready to state our main result:

Theorem 1. *Let $\tilde{\mu}$ be an estimator such that $\tilde{\mu} - \mu_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$. Assume that the vector of covariates is not perfectly collinear with the individual errors, that is: for any vector $\lambda \in \mathbb{R}^L$, $\mathbb{P}(\lambda' X_{12} = U_1 U_2) < 1$. Assume that the U's have at least 4 finite moments, that $Var(U) = \sigma_U^2 \neq 0$ and that the V's have at least 2 finite moments.*

Define the sequence $\hat{\mu}_m$ by: $\hat{\mu}_0 := \tilde{\mu}$ and $\hat{\mu}_{m+1} := f_N(\hat{\mu}_m)$, and let $\hat{\mu}^* := \limsup_m \hat{\mu}_m$. Then with probability approaching 1 $\hat{\mu}^* = \lim_{m \to +\infty} \hat{\mu}_m$ and $\hat{\mu}^*$ is a solution to (7). Moreover:

$$N(\hat{\mu}^* - \mu_0) \to_d \mathcal{N}\left(0, 2\sigma_V^2 \Sigma^{-1}\right) \tag{9}$$

for

$$\Sigma := \left(E(X_{12} X_{12}') + \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12}) E(U_1 U_2 X_{12})' - \frac{2}{E(U_1^2)} E(U_1 U_3 X_{12} X_{23}')) \right)$$

Proof. Immediately follows from corollary 2 and proposition 4. □

Notice that $\hat{\mu}^*$ is asymptotically biased. This is a manifestation of the incidental parameter problem and in line with the behavior of other estimators that incidentally estimate the individual effects (e.g. Bai (2009) , Moon and Weidner (2017)). In section 4, we offer a bias correction by proposing a consistent estimator for the bias term.

The covariance matrix Σ is equal to the asymptotic covariance D_0 in Bai (2009) (if we were to use the LS estimator on an oriented network model of the form $Y_{ij} = X_{ij}\beta + A_i B_j + V_{ij}$ to fit Bai (2009))'s assumptions). The coefficient 2 in $\hat{\mu}^*$'s asymptotic variance is a sample size adjustment due the fact that out model is symmetric and that the actual number of observations is $\frac{N(N-1)}{2}$ rather than N^2 if the model were oriented.

Theorem 1 shows that if we iterate for *long enough*, we approach a minimizer of the objective function g_N with high probability. The theorem does however not provide any guidance regarding the number of iterations that are required to "sufficiently" approach the optimum. In fact, we show (in proposition 1) that if we initiate with a $\sqrt{N}$-consistent estimator, no finite number of iterations is sufficient to escape the $\sqrt{N}$ rate of convergence. For any hope of achieving a superior rate, the number of iterations needs to be indexed by the sample size N, which is computationally challenging. One exception to this curse are noteworthy. If the individual effect are centered and independent of the observable regressors X, then proposition 2 shows that even initiating with a $\sqrt{N}$-consistent estimator, one iteration is enough to obtain an estimator with the same asymptotic properties of $\hat{\mu}^*$. In that case also, interestingly, the OLS estimator is N-consistent but has a non standard asymptotic distribution (c.f. for instance Menzel (2021)). Moreover, under these assumptions, $\hat{\mu}^*$ is easily shown to be asymptotically efficient (refer to the discussion following proposition 2).

We circumvent the debate around the appropriate number of iterations by proposing an alternative estimator that is asymptotically equivalent to the "oracle" $\hat{\mu}^*$. The alternative estimator only requires 2 iterations on the function f_N, significantly limiting the computational burden. First, starting with a $\sqrt{N}$-consistent estimator $\tilde{\mu}$, define the matrix

$$\hat{K}_N := \left(\sum_{i \neq j} X_{i,j} X'_{i,j,k} - \sum_{i \neq j,k} \nu_i(\tilde{\mu}) X_{i,j} X'_{j,k} \nu_k(\tilde{\mu}) \right)^{-1} \\ \times \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} X'_{j,k} \nu_k(\tilde{\mu}) - \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} \nu_j(\tilde{\mu}) \right) \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} \nu_j(\tilde{\mu}) \right)' \right)$$

then follow the three steps:

1. Run one iteration to get $\hat{\mu}_1 := f_N(\tilde{\mu})$
2. Compute $\check{\mu}_1 := (I_L - \hat{K})^{-1} \hat{\mu}_1 + (I_L - (I_L - \hat{K})^{-1}) \tilde{\mu}$
3. Iterate on $\check{\mu}_1$ to get $\hat{\mu}_2 := f_N(\check{\mu}_1)$

4. Compute $\check{\mu}_2 := (I_L - \hat{K})^{-1}\hat{\mu}_2 + (I_L - (I_L - \hat{K})^{-1})\check{\mu}_1$

we show (proposition 5) that $\hat{K}_N$ is a consistent estimator for a matrix K that is central to our analysis of the iteration process. More on the definition of K, its role, and why the 4 steps above deliver an estimator with the desired properties comes in section 3. We conclude this section by stating the paper's second main theorem:

Theorem 2. *Under the assumptions of theorem 1*

$$N(\check{\mu}_2 - \hat{\mu}^*) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

Proof. Immediately follows from corollaries 3 and 4. □

When $\delta = 0$ in model (1), however, the iterations become explosive (see proposition 6) and do not converge. μ^* is therefore not well defined. The alternative estimator remains consistent for all slope parameters, excluding the interecept, when the individual effects A in equation (1) are independent of the regressors X:

Theorem 3. *Assume A is independent of X, $\delta = 0$, and $\tilde{\mu}$ is $\sqrt{N}$-consistent for μ_0. Then:*

$$\begin{aligned}\sqrt{N} diag(0, 1, \ldots, 1)(\check{\mu}_2 - \mu_0) &= diag(0, 1, \ldots, 1)\left(\Gamma_1(c)\sqrt{N}(\tilde{\mu} - \mu_0) + \Gamma_2(c)\frac{\sqrt{N}}{N^2}\sum_{ij} X_{ij}A_j\right) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= O_p(1)\end{aligned}$$

for some $c \sim 1 - 2Bernouilli(\frac{1}{2})$ and some deterministic matrix valued functions $\Gamma_1(.)$ and $\Gamma_2(.)$.

Proof. Follows from proposition 6 (section 5) and lemma 6 in appendix A.13. □

The two following sections discuss the intuition behind the *least eigenvalue estimator* and break down the intermediary results leading to theorems 1 and 2. Section 2 details how the interactive term in equation (3) affects the spectrum of the matrix $M(\mu)$ and why minimizing the function g_N is a sensible choice. Section 3 outlines the theoretical results starting from the behavior of a single iteration estimator and leading up to the construction of the alternative estimator $\check{\mu}_2$.

2 Some intuition

Consider the ordinary least squares estimator on model 3, defined by

$$\begin{aligned}
\hat{\mu}_{OLS} &:= \arg\min_{\mu\in\mathbb{R}^L} \sum_{i,j} \left(Y_{ij} - \sum_{l=1}^{L} \mu_l X_{ij,l} \right)^2 \\
&= \arg\min_{\mu\in\mathbb{R}^L} Trace \left(\left(Y - \sum_{l=1}^{L} \mu_l X_l \right)^2 \right) \\
&= \arg\min_{\mu\in\mathbb{R}^L} \sum_{i=1}^{N} \left(\lambda_i \left(Y - \sum_{l=1}^{L} \mu_l X_l \right)^2 \right) \\
&=: \arg\min_{\mu\in\mathbb{R}^L} \sum_{i=1}^{N} \lambda_i \left(M(\mu)^2 \right)
\end{aligned} \tag{10}$$

Equation (10) indicates that the OLS estimator can also be defined as a minimizer of the average squared eigenvalues of the matrix $M(\mu)$. Let's examine the distribution of $M(\mu)$'s eigenvalues for values of μ that are "close" to the true value μ_0, assuming $\delta = 1$ (the treatment for $\delta = -1$ is similar). Begin with the value $\mu = \mu_0$, that is, let's look at the distribution of the eigenvalues of the matrix $UU' + V$. [4] Figure 3 shows the histogram of the eigenvalues of the simulated matrix $\frac{1}{\sqrt{N}}(UU' + V)$, where the U's and V's are i.i.d standard normal and the sample size is set to $N = 2000$.

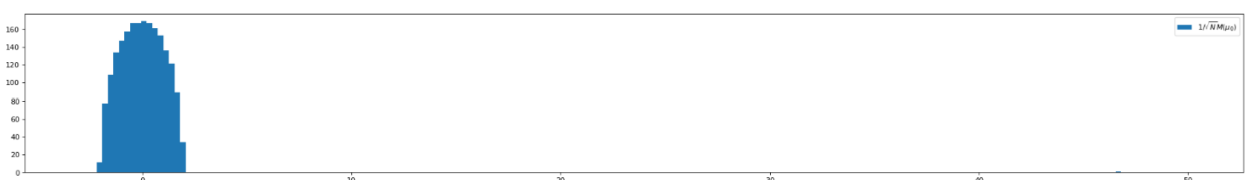

Figure 3: *A histogram for $\frac{1}{\sqrt{N}}M(\mu_0)$'s eigenvalues; $\sigma_u = \sigma_v = 1$, $N = 2000$*

The histogram in figure 3 shows two distinct parts: to the left, a block of eigenvalues concentrated between values ~ -2 and $\sim +2$, and a single eigenvalue, further to the right, at around value ~ 46. After proper rescaling (and ignoring the single eigenvalue to the left for the rescaled histogram to fit on a page) the block of eigenvalues to the left has the shape of a semi-circle as shown in figure 4 .

[4] We ignore the effect of the matrix $E(U_1^2)I_N$ in the discussion that follows. $E(U_1^2)I_N$ simply shifts all eigenvalues by the same quantity $E(U_1^2)$. The shift size will turn out to be of a low order of magnitude compared to the bulk of $UU' + V$'s eigenvalues and its effect will be negligible anyways.

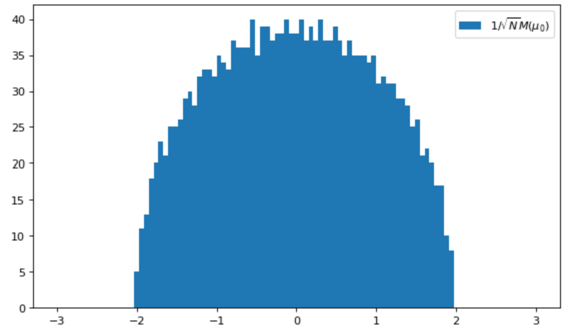

Figure 4: *A zoom into the semi circle (the left block in Figure 3)*

To rationalize the shape of the histogram 3, let's examine the eigenvalues of each of the terms composing $M(\mu_0)$. The matrix UU' is of rank 1, its unique non null eigenvalue is equal to $U'U = \sum_i U_i^2$ which is of the same order as $NE(U_1^2)$ when N is large enough.

Figure 5 shows the histogram of V's eigenvalues. The two histograms in 4 and 5 are seemingly identical. Only $M(\mu_0)$'s outlier eigenvalue (the one approximately equal to 46) is absent from V's histogram. This should come as no surprise: the matrix $M(\mu_0)$ is a rank 1 deformation of V. The impact of rank 1 deformations on the eigenvalues of the original matrix (V here) is well studied (e.g. Bunch et al. (1978)). Because UU's unique eigenvalue is positive, modifying V through UU' shifts all of V's eigenvalues upwards such that V's eigenvalues are interlaced with $V + UU'$'s, that is, for $i = 2, ..., N$:

$$\lambda_i(V) \leq \lambda_i\left(V + UU'\right) \leq \lambda_{i-1}(V)$$

and

$$\lambda_1(V) \leq \lambda_1\left(V + UU'\right)$$

Provided that V's eigenvalues (rescaled by $\frac{1}{\sqrt{N}}$) are concentrated roughly between -2 and 2, then the inequalities above predict that $V + UU'$'s $N - 1$ smallest eigenvalues will be only shifted by a small amount, which explains why the figures 4 and 5 are not visually distinguishible.

The semi-circle in figure 4 is reminiscent of Weigner's semi-circle law in the random matrix literature (see for instance Benaych-Georges and Knowles (2016)). Weigner's law states that the empirical distribution of the eigenvalues of a random symmetric matrix with centered square integrable entries "converges" (in a sense that is made precise below) to a distribution with a semi-circular probability density function. In particular, Füredi and Komlós (1981) show that V's largest eigenvalue is of order $\sqrt{N}$ with probability approaching 1 as N grows.

These observations combined suggest the following rough interpretation of the histogram 3:

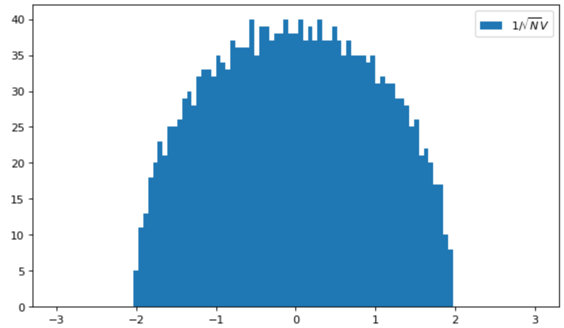

Figure 5: *A histogram for* $\frac{1}{\sqrt{N}}V$*'s eigenvalues;* $\sigma_v = 1$, $N = 2000$

$M(\mu_0)$'s $N-1$ smallest eigenvalues are of order $\sqrt{N}$ and are "very close" to V's eigenvalues, whereas the largest eigenvalue is due to the UU' deformation and is of order N.

Let's extend these intuitions to values of μ that are different from the true parameter μ_0. If μ is too far from μ_0, then the term $\sum_{l=1}^{L}(\mu_{0,l}-\mu_l)X_l$ in equation (5) can become dominant and dwarf the contributions of V and UU' in $M(\mu)$'s eigenvalue distribution. In the other extreme, when the candidate μ is"very close" to μ_0, then the contribution of the covariates' term becomes negligible and we obtain a histogram that is similar to the one in figure 3.

The values of μ that are abberantly far from μ_0 lead to the eigenvalues of $\sum_{l=1}^{L}(\mu_{0,l}-\mu_l)X_l$ being of a higher than order $\sqrt{N}$. Subsequently, they are easy to eliminate as they produce a histogram that is grossly different from the one in figure 3. However, this rough discrimination strategy will be ineffective for values of μ that return a term $\sum_{l=1}^{L}(\mu_{0,l}-\mu_l)X_l$ of order $\sqrt{N}$ or lower. In any case, for model (3), when U is independent of X, the OLS estimator is known to be at least $\sqrt{N}$ consistent in general (See for instance Menzel (2021) or section 4 in Graham (2020)).[5] The following lemma shows that any $\sqrt{N}$ estimator is in fact "close enough" for our purposes.

Lemma 2. *For any* $N \times N$ *matrix* X_N *with entries* X_{ij} *that can be represented as:* $X_{ij} =_d \psi(X_i, X_j, W_{ij})$ *for some i.i.d. and finite dimensional* X_i*'s and some i.i.d* $W'_{ij}s$ *that are i.i.d. and independent from the* X_i*'s. Assume* X_{ij} *has at least 4 finite moments. We have that*

$$\max_i |\lambda_i(X_N)| = O_p(N)$$

Proof. Refer to subsection A.3. □

The lemma implies that any initial estimator $\tilde{\mu}$ that is $\sqrt{N}$ consistent - like the OLS estimator - would yield a covariates' term such that $\sum_{l=1}^{L}(\mu_{0,l}-\tilde{\mu}_l)X_l = O_p(\sqrt{N})$. It would

[5] In fact, if $E(U_1) \neq 0$, the OLS estimator of the intercept is biased. In proposition 7 (appendix A.1), we show how that bias can be corrected to obtain a $\sqrt{N}$-consistent estimator.

produce an eigenvalue histogram for $M(\tilde{\mu})$ that is similar to figure 3 with one outlier eigenvalue of order N, due to the rank 1 modification UU', and a cloud of eigenvalues that are of a smaller order $\sqrt{N}$ but that need not form a semi-circle this time.

Provided that the candidate μ is close enough to μ_0, the largest eigenvalue of $M(\mu)$ is, at least up to a first order approximation, closely tied to the error term UU'. Notice that, absent the UU' from the model 3 (or the random effects A_i and A_j from the model (1)), we would be back to the standard linear regression model with i.i.d. and exogenous noise V_{ij}. In that case, we know that OLS is efficient, and since the sample size is $\frac{N(N-1)}{2}$, the rate of convergence of the OLS estimator would be N, rather than $\sqrt{N}$ under models (1) or (3).

An appealing idea is then to modify the objective function in the matrix form definition of the OLS in (10) to remove the contribution of the random effects. Following the intuition laid down so far, this can for instance be done by removing $M(\mu)$'s largest eigenvalue from the sum of squared errors before minimizing. The new estimator would be a solution to the minimization problem

$$\min_{\mu \in \mathbb{R}^L} \sum_{i=2}^{N} \lambda_i \left(M(\mu)^2 \right) = g_N(\mu) \tag{11}$$

3 Single iteration analysis

The first result examines a single iteration of the fixed point problem (8).

Proposition 1. *Consider the model 3:*

$$Y_{ij} = \sum_{k=1}^{K} \mu_{0,k} X_{ij,k} + \delta U_i U_j + V_{ij} = X_{ij}\mu_0 + \delta U_i U_j + V_{ij}$$

where $\delta \in \{-1, +1\}$. Assume that the vector of covariates is not perfectly colinear with the individual errors, that is: for any vector $\lambda \in \mathbb{R}^L$, $\mathbb{P}(\lambda' X_{12} = U_1 U_2) < 1$. Assume that the U's have at least 4 finite moments, that $Var(U) = \sigma_U^2 \neq 0$ and that the V's have at least 2 finite moments. Given a first stage estimator $\tilde{\mu}$ such that $||\tilde{\mu} - \mu_0|| = O_p\left(\frac{1}{\sqrt{N}}\right)$, the single iteration estimator

$$\hat{\mu} := \left(\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\mu}) \nu_j(\tilde{\mu}) X_{jk}' X_{ik} \right)^{-1} \left(\sum_{i \neq j} X_{ij}' Y_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\mu}) \nu_j(\tilde{\mu}) X_{jk}' Y_{ik} \right) \tag{12}$$

satisfies

$$\sqrt{N}(\hat{\mu} - \mu_0) = K\sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{13}$$

for

$$K := \frac{1}{E(U_1^2)} \left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}'X_{32}) \right)^{-1} \left(E(U_1U_3X_{12}X_{23}) - \frac{1}{E(U_1^2)} E(U_1X_{12}U_2)E(U_1U_2X_{1}$$

A detailed proof is presented in section A.5. Proposition 3 shows that $\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}'X_{32}\right.$ is invertible, insuring that K is well defined.

Equation (13) describes how the distribution of the single iteration estimator relates to the first stage estimator's. An immediate corollary of proposition (1) is that the single iteration estimator $\hat{\mu}$ is consistent and converges to μ_0 at least as at a rate of $\sqrt{N}$. Also, up to a first order approximation, the single stage estimator depends linearly on the initial $\tilde{\mu}$.

Whether the iteration process improves the quality of estimation depends on the matrix K. When the individual effects U are independent of the regressors X and when $E(U_1) = 0$ in proposition 1, the matrix K is null and equation (13) becomes

$$\hat{\mu} - \mu_0 = O_p\left(\frac{1}{N}\right).$$

After a single iteration, we are able to achieve the optimal rate of convergence N. Unfortunately, proposition 1 does not provide the asymptotic distribution of $\hat{\mu}$ or the effect iterations have beyond the first iteration. To answer both these questions, we need to *zoom* into the $O_p\left(\frac{1}{\sqrt{N}}\right)$ term in equation (13) and determine how it depends on the first stage estimator $\tilde{\mu}$ and/or how it behaves asymptotically. The next proposition and its proof in appendix A.5 address this case.

Proposition 2. *In addition to the assumptions in proposition 1, assume that the individual effects are independent of the regressors X_{ij} and that $E(U_i) = 0$, then*

$$N(\hat{\mu} - \mu_0) \to_d \mathcal{N}\left(0, 2\sigma_v^2 E(X_{12}X_{12}')^{-1}\right) \tag{14}$$

A "one step theorem" applies, one iteration is enough to achieve full efficiency. The argument proving the efficiency of $\hat{\mu}$ in proposition 2 is simple: consider the alternative model $Y_{ij} = \sum_{l=1}^{L} \mu_{0,l} X_{ij,l} + V_{ij} = X_{ij}\mu_0 + V_{ij}$ is *i.i.d.* with *i.i.d.* errors V_{ij}. In this model, the ordinary least squares estimator is known to be efficient and asymptotically normal, with asymptotic covariance matrix $2\sigma_v^2 E(X_{12}X_{12}')^{-1}$ - the same asymptotic distribution as in (14) (see for instance Chamberlain (1987) or Newey (1990)). Given that our model of interest (3) is noisier than the alternative model, the following corollary holds.

Corollary 1. *Under the assumptions of proposition 1, when $E(U_i) = 0$, the single iteration estimator defined in (13) is semi-parametrically efficient.*

When $K \neq 0$, the one step theorem no longer applies. After any finite number of iterations, the new estimator is still $\sqrt{N}$-consistent. To understand the role of K when $K \neq 0$, consider the simple case where we have a single regressor ($L = 1$). K becomes a scalar and when $|K| < 1$, $\hat{\mu}$ is to a first order closer to μ_0 than $\tilde{\mu}$. If the first stage estimator is asymptotically normal (the standard *ordinary least squares* estimator for example) with an asymptotic variance of $\tilde{\sigma}^2$, then $\hat{\mu}$ is normally distributed with variance $K^2\tilde{\sigma}^2 < \tilde{\sigma}^2$. Moreover, as we iterate, the variance decays exponentially in the number of iterations. Conversely, if $|K| > 1$, iterations produce noisier estimators, and the variance explodes exponentially with the number of iterations. Finally, if $|K| = 1$, then the new estimator is asymptotically equivalent to the first stage estimator, iteration is neither useful nor harmful.

Simplify further, and assume that the single regressor is in fact just a constant $X_{ij} = 1$, that is, we are interested in estimating the mean of Y_{ij}. The constant K becomes $K = \frac{E(U_1)^2}{E(U_1^2)}$ which is positive and strictly smaller than 1 (since by assumption $\sigma_U^2 > 0$), the iterations improve estimation quality.

When $L > 1$, K is a matrix. Rather than comparing K to 1, the relevant comparison is now between K and I_L - the identity matrix of dimension L - in the partial order on symmetric matrices. When $K^2 > I_L$, that is, when K's eigenvalues are all larger than 1 in absolute value, the successive iterations follow an explosive path of covariance matrices. The conclusions are similar to the univariate setting in the two cases: $K^2 < I_L$ or $K^2 = I_L$. In the multivariate case however, these three cases are not exhaustive, since $>$ here is only a partial order. Fortunately, the next proposition shows that the only possibile case, given our assumptions, is in fact $0 < K < I_L$.

Proposition 3. *Under the assumptions of proposition 1, the matrix* $\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)$ *is definite positive and all the eigenvalues of the matrix*

$$K := \frac{1}{E(U_1^2)}\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1}\left(E(U_1U_3X_{12}X_{23}) - \frac{1}{E(U_1^2)}E(U_1X_{12}U_2)E(U_1U_2X_{1\ldots}\right.$$

are positive and strictly smaller than 1.

Proof. cf. section A.6 □

Together, the propositions 1 and 3 imply that given a $\sqrt{N}$-consistent initial estimator and a fixed $\epsilon > 0$, we iterate the process described in the equation (12) to obtain a new $\sqrt{N}$-consistent estimator with a variance that is smaller than ϵ (or ϵI_L in the multivariate case). This strongly suggests that an estimator with a faster than $\sqrt{N}$ rate of convergence exists. In fact, using a simple trick, the propositions 8 and 3 provide a rate N (a rate optimal) estimator.[6]

Another corollary of proposition 1 is that, if f_N has a fixed point $\hat{\mu}^*$ that is $\sqrt{N}$-consistent, then equation (13) yields:

$$(I-K)\sqrt{N}(\hat{\mu}^* - \mu_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

so $\hat{\mu}^*$ is in fact N-consistent. Proposition 1 is silent about the exact asymptotic distribution of $\hat{\mu}^*$ and about its existence.

To establish the existence of a fixed point, notice that equation (13) has the flavor of Taylor expansion, where the matrix K would represent a gradient. Because the matrix K has a spectral radius that is smaller than 1 (proposition 3), then f_N must be contracting in a local sense. Then (a variation on) the Banach fixed point theorem should prove existence. This intuition is the main idea for the proof for the next corollary.

Corollary 2. *Let $\tilde{\mu}$ be an estimator such that $\tilde{\mu} - \mu_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$. Fix $\kappa \in (\lambda_1(K), 1)$ and some $C > 0$. With probability approaching 1:*

1. *The function f_N in equation (8) is differentiable in the closed ball $B(\mu_0, \frac{C}{\sqrt{N}})$ centered at μ_0 and with radius $\frac{C}{\sqrt{N}}$.*
2. $\sup_{\mu \in B(\mu_0, \frac{C}{\sqrt{N}})} ||f'_N(\mu)|| \leq \kappa < 1$

Moreover, define the sequence $\hat{\mu}_m$ by: $\hat{\mu}_0 := \tilde{\mu}$ and $\hat{\mu}_{m+1} := f_N(\hat{\mu}_m)$, and $\hat{\mu}^* := \limsup_m \hat{\mu}_m$. Then $\hat{\mu}^* - \mu_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$ and with probability approaching 1 $\hat{\mu}^* = \lim_{m\to+\infty} \hat{\mu}_m$ and $\hat{\mu}^*$ is a solution to (7).

Proof. Cf. Section A.7 . □

So $\hat{\mu}^*$ exists with probability approaching 1 and is rate optimal. It is left to determine its asymptotic distribution. We need to compute a higher order term in the expansion (13) of proposition 1. That is the purpose of proposition 4.

Proposition 4. *Under the assumptions of proposition 1, if the first stage estimator is such that $\tilde{\mu} - \mu_0 = O_p\left(\frac{1}{N}\right)$, then*

$$N(\hat{\mu} - \mu_0) = KN(\tilde{\mu} - \mu_0) + R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{15}$$

[6] A similar idea is used for instance to construct a "Generalized Jackknife estimator" (e.g. Powell et al. (1989), Cattaneo et al. (2013)). In the context of proposition 1, the term $\sqrt{N}(\tilde{\mu} - \mu_0)$ is eliminated by taking the convex combination, in the same fashion that the bias is removed in the generalized jackknife by taking a convex combination of estimator with the same bias.

with

$$R_N \to_d \left(E(X_{12}X_{12}') - \frac{E(U_1)^2}{E(U_1^2)} E(X_{12}X_{32}') \right)^{-1} \mathcal{N}\left(0, 2\sigma_V^2 \Sigma\right)$$

for

$$\Sigma := \left(E(X_{12}X_{12}') + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12})E(U_1U_2X_{12})' - \frac{2}{E(U_1^2)} E(U_1U_3X_{12}X_{23}')) \right)$$

Proof. C.f. Section A.8 □

Because of the presence of the residual R_N in equation (15), the new expansion is fundamentally different from the previous one ((13) in proposition 1). The effect of an iteration on the estimation quality is now ambiguous and depends on how the first stage estimator $\tilde{\mu}$ relates to the residual R_N. Even if they were independent, it is not clear whether iteration improves estimation. Unfortunately, even though proposition 2 provides the asymptotic distribution of R_N, that is not enough to fully characterize the distribution of the single iteration estimator. For that, we would need the joint distribution of the first stage $\tilde{\mu}$ and R_N, which is challenging even for a single iteration. However, we can see that because K has a smaller than one spectrum (by proposition 3), as we iterate, the contribution of the initial (first stage or input) estimator fades away. Intuitively, starting with some $N-$consistent first stage $\tilde{\mu}(=: \hat{\mu}_0)$, from (15):

$$\begin{aligned} N(\hat{\mu}_m - \mu_0) &\approx K^m N(\tilde{\mu} - \mu_0) + \sum_{i=0}^{m-1} K^i R_N \\ &\approx K^m N(\tilde{\mu} - \mu_0) + (I_L - K^m)(I_L - K)^{-1} R_N \\ &\approx (I_L - K)^{-1} R_N; \text{ when } m \text{ is lage.} \end{aligned}$$

So the limit distribution (when m approaches infinity) should not depend on the initial estimator $\tilde{\mu}$. Corollary 3 formalizes these thoughts.

Corollary 3. *Let $\tilde{\mu}$ be a $\sqrt{N}$ consistent estimator. Define the sequence $\hat{\mu}_m$: $\hat{\mu}_0 := \tilde{\mu}$ and $\hat{\mu}_{m+1} := f_N(\hat{\mu}_m)$ for all $m \geq 0$, and let $\hat{\mu}^* := \limsup_m \hat{\mu}_m$. Then*

$$N(\hat{\mu}^* - \mu_0) = (I - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and with probability approaching 1 $\hat{\mu}^*$ is a solution to (7). Therefore

$$N(\hat{\mu}^* - \mu_0) \to_d \frac{2\delta}{E(U_1^2)} \Sigma^{-1} E(U_1^3 U_2 X_{12}) + \mathcal{N}\left(0, 2\sigma_V^2 \Sigma^{-1}\right) \tag{16}$$

for

$$\Sigma := \left(E(X_{12}X_{12}') + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12})E(U_1U_2X_{12})' - \frac{2}{E(U_1^2)} E(U_1U_3X_{12}X_{23}')) \right)$$

Proof. Immediately follows from proposition 2 and corollary 2. □

The covariance matrix Σ is equal to the asymptotic covariance D_0 in Bai (2009) (if we were to use the LS estimator on an oriented network model of the form $Y_{ij} = X_{ij}\beta + A_iB_j + V_{ij}$ to fit Bai (2009))'s assumptions). The coefficient 2 in $\hat{\mu}^*$'s asymptotic variance and that we don't see in Bai (2009) is simply a sample size adjustment due the fact that out model is symmetric and that the actual number of observations is $\frac{N(N-1)}{2}$ rather than N^2 had the model been oriented.

The corollary shows that if we initiate a sequence $\hat{\mu}_0 := \tilde{\mu}$, for some initial $\sqrt{N}$-consistent estimator $\tilde{\mu}$, and then we iterate "infinitely many" $\hat{\mu}_{m+1} := f_N(\hat{\mu}_m)$ as in corollary 2, then with high probability $\hat{\mu}_m$ approaches a fixed point $\hat{\mu}^*$. As is standard in numerical optimization methods, "infinitely many" repetitions can in practice be read as "sufficiently many repetitions". None of the results so far in this paper provides any guidance regarding how many repetitions are enough. In fact, one of proposition 1's corollaries can be concerning: equation (13) establishes that if we initiate with a $\sqrt{N}$ consistent estimator, then we can only hope the iteration process to return $\sqrt{N}$-consistent estimators if we stop after a finite number of iterations. Therefore, from equation (37), we get a sense of what a lower bound on the number of iterations should be, and it is rather massive. The number of iterations should be a diverging function of the sample size N for us to have any hope to escape the $\sqrt{N}$ rate of convergence. How fast the number of iterations grows with N will have an effect on the rate of convergence of the final estimator, but it is hard to tell what the proper order of magnitude is. It is even less clear what the rate of convergence would be if the number of iterations is indexed on some stoppage criterion on the value of the objective function, as is usually the case in standard numerical optimization algorithms. In simulations, the question of the number of iterations does not seem to be problematic. The standard optimization methods deliver distributions that are in line with the predictions of the asymptotic results presented so far, in particular the asymptotic distribution of $\hat{\mu}^*$ in corollary 3.

Fortunately, the propositions 1 and 2 can be put to use differently to extract an estimator that is asymptotically equivalent to the minimizer $\hat{\mu}^*$. The alternative estimator requires exactly 2 iterations over the function f_N and is therefore numerically more efficient. Using the alternative estimator, we can circumvent the concerns we highlighted around the number of iterations that are sufficient to achieve the desired asymptotic distribution.

The equivalent estimator

First, assume that the matrix K is observed. Let $\tilde{\mu}$ be an initial $\sqrt{N}$ consistent estimator and let $\hat{\mu}_1$ be the estimator returned in the equation (12) after a single iteration. Write $\breve{\mu}_1^* := G\hat{\mu}_1 + (I_L - G)\tilde{\mu}$, for some fixed $L \times L$ matrix G and where I_L is the identity matrix of dimension L. We will choose the matrix G so that $\breve{\mu}_1^*$ converges at rate N. Write

$$
\begin{aligned}
\sqrt{N}(\breve{\mu}_1^* - \mu_0) &= \sqrt{N}G(\hat{\mu}_1 - \mu_0) + \sqrt{N}(I_L - G)(\tilde{\mu} - \mu_0) \\
&= (GK + I_L - G)\sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= \big(I_L - G(I_L - K)\big)\sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

choosing G such that $I_L - G(I_L - K) = 0$ - i.e. $G = (I_L - K)^{-1}$ - yields a rate N estimator. Note that by the proposition 3, $I_L - K$ is invertible and G is well defined.

In practice, the matrix K is not observed. Instead, it needs to be estimated and plugged in to generate an estimator for G. Assume we have a consistent estimator $\hat{K}$ for K. Define $\hat{G} := (I_L - \hat{K})^{-1}$ and $\breve{\mu}_1 := \hat{G}\hat{\mu} + (I_L - \hat{G})\tilde{\mu}$. As for $\breve{\mu}^*$

$$
\begin{aligned}
\sqrt{N}(\breve{\mu}_1 - \mu_0) &= \sqrt{N}\hat{G}(\hat{\mu} - \mu_0) + \sqrt{N}(I_L - \hat{G})(\tilde{\mu} - \mu_0) \\
&= \left(I_L - (I_L - \hat{K})^{-1}(I_L - K)\right)\sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= (I_L - \hat{K})^{-1}\left(K - \hat{K}\right)\sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned} \tag{17}
$$

If $\hat{K}$ is a $\sqrt{N}$ - consistent estimator for K, that is, if $\hat{K} - K = O_p\left(\frac{1}{\sqrt{N}}\right)$, then the new estimator $\breve{\mu}$ is rate optimal. The following proposition offers an example of a $\sqrt{N}$ consistent estimator for K.

Proposition 5. *Let $\tilde{\mu}$ be a $\sqrt{N}$-consistent estimator for μ_0. Define:*

$$
\hat{K}_N := \left(\sum_{i \neq j} X_{i,j}X'_{i,j,k} - \sum_{i \neq j,k} \nu_i(\tilde{\mu})X_{i,j}X'_{j,k}\nu_k(\tilde{\mu})\right)^{-1}
$$

$$\times \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} X'_{j,k} \nu_k(\tilde{\mu}) - \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} \nu_j(\tilde{\mu}) \right) \left(\sum_{i \neq j} \nu_i(\tilde{\mu}) X_{i,j} \nu_j(\tilde{\mu}) \right)' \right)$$

Then

$$\hat{K} - K = O_p\left(\frac{1}{\sqrt{N}}\right)$$

Proof. c.f. section A.9 □

Proposition 5 allows for the construction of an estimator that is rate optimal. However, studying the asymptotic distribution of $\check{\mu}_1$ defined in (17) is challenging. It requires that we determine the joint asymptotic distribution of $\hat{K}$, $\sqrt{N}(\tilde{\mu} - \mu_0)$ and the residual of order $O_p\left(\frac{1}{\sqrt{N}}\right)$ in equation (17). However, as for the study of the fixed point $\hat{\mu}^*$, as we iterate, the effect of first stage estimator fades away. Rather than iterating here again, we use the same linear combination trick that allows us again to achieve the "infinite iterations" distribution using one iteration only.

Let $\hat{\mu}_2 := f_N(\check{\mu}_1)$ and define $\check{\mu}_2 := \hat{G}\hat{\mu}_2 + (I_L - \hat{G})\check{\mu}_1$. Following the steps in equation (17),

$$\begin{aligned} N(\check{\mu}_2 - \mu_0) &= (I_L - \hat{K})^{-1}\left(K - \hat{K}\right) N(\check{\mu}_1 - \mu_0) + (I_L - \hat{K})^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= (I_L - K)^{-1} R_N + (I_L - \hat{K})^{-1}\left(K - \hat{K}\right) N(\check{\mu}_1 - \mu_0) + (K - \hat{K})^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= (I_L - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \end{aligned} \tag{18}$$

the last equality is a consequence of proposition 5. This proves that $\check{\mu}_2$ is asymptotically equivalent to $\hat{\mu}^*$, the fixed point studied through corollary 3.

To summerize, the alternative estimation procedure follows these steps:

1. Run one iteration to get $\hat{\mu}_1 := f_N(\tilde{\mu})$
2. Compute $\check{\mu}_1 := (I_L - \hat{K})^{-1}\hat{\mu}_1 + (I_L - (I_L - \hat{K})^{-1})\tilde{\mu}$
3. Iterate on $\check{\mu}_1$ to get $\hat{\mu}_2 := f_N(\check{\mu}_1)$
4. Compute $\check{\mu}_2 := (I_L - \hat{K})^{-1}\hat{\mu}_2 + (I_L - (I_L - \hat{K})^{-1})\check{\mu}_1$

Corollary 4.

$$N(\check{\mu}_2 - \mu_0) = (I_L - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{19}$$

Proof. See the steps leading to equation (17). □

4 Variance estimation

To be able to do inference on the (asymptotically equivalent) estimators presented in the previous section. We need to provide a consistent estimator for the covariance matrix $2\sigma_V^2\Sigma^{-1}$ (proposition 3).

To provide a consistent estimator for the bias term, we can use $\hat{K}$ defined in proposition 5 as a consistent estimator for K. The matrice $E(X_{12}X_{12}')$ the vector $E(X_{12})$ can be estimated through their sample analogues. δ, $E(U_1^2)$, $E(U_1U_3X_{12}X_{32}')$, $E(U_1^3U_2X_{12})$ and $E(U_1U_2X_{12})$ are left to be estimated. Assume that $\delta = 1$, section 2 explained how the eigenvector corresponding to the largest eigenvalue of $M(\tilde{\mu})$ is a good approximation to the normalized vector $\frac{U}{||U||_2}$. Moreover, the largest eigenvalue informs about $U'U$, the norm of the vector U. Combining both, we can recover an estimator for U. When δ is -1, then we reason in terms of the largest eigenvalue in absolute value, and its corresponding eigenvalue. The difference when $\delta = -1$ is that the corresponding eigenvalue in fact estimates $-\frac{U}{||U||_2}$ rather than $\frac{U}{||U||_2}$ and a sign correction is necessary. The sign of δ is, with probability approaching 1, the sign of the largest eigenvalue in absolute value. These ideas are formalized through lemma 3.

Lemma 3. *Under the conditions and notation of proposition 1, denote $\hat{U}_i = \sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu_i(\tilde{\mu})$. We have*

1. $\frac{\sum_i U_i^2 - \hat{U}_i^2}{N} = O_p\left(\frac{1}{N}\right)$,
2. $\frac{1}{N^3}\sum_{i\neq j, k\neq i,j} \hat{U}_i\hat{U}_k X_{ij}X_{jk}' \to_p E(U_1U_3X_{12}X_{32}')$
3. $\frac{1}{N^2}\sum_{i\neq j} \hat{U}_i\hat{U}_j X_{ij} \to_p E(U_1U_2X_{12})$
4. $\frac{1}{N^2}\sum_{i\neq j} X_{ij}X_{ij}' \to_p E(X_{12}X_{12}')$

Proof. Cf section A.10. □

Finally, to obtain an estimator for the covariance matrix, it is left to provide a consistent estimator for the variance σ_V^2. Let's go back to the model (2)

$$Y_{ij} = \sum_{l=1}^{L} \mu_{0,l} X_{ij,l} + \delta U_i U_j + V_{ij}$$

First, observe that

$$\frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij} - X_{ij}\check{\mu}_2\right)^2 = \frac{1}{N^2}\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right)^2 + (\check{\mu}_2 - \mu_0)'\left(\sum_{i\neq j} X_{ij}'X_{ij}\right)(\check{\mu}_2 - \mu_0)$$

$$
\begin{aligned}
&- 2\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right) X_{ij}(\check{\mu}_2 - \mu_0) \\
&= \frac{1}{N^2}\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right)^2 + O_p\left(\frac{1}{N}\right) \\
&= E(U_1^2)^2 + \sigma_V^2 + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

Therefore, given the estimator for $E(U_1^2)$ provided in lemma 3, we obtain a consistent estimator $\hat{\sigma}_V^2 := \frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij} - X_{ij}\check{\mu}_2\right)^2 - \left(\frac{\sum_i \hat{U}_i^2}{N}\right)^2$, where $\hat{U}_i$ are defined in lemma 3. To summarize:

Corollary 5. *Define* $\hat{\sigma}_V^2 := \frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij} - X_{ij}\check{\mu}_2\right)^2 - \left(\frac{\sum_i \hat{U}_i^2}{N}\right)^2$ *where* $\hat{U}_i$ *is defined in lemma 3. We have*

$$\hat{\sigma}_V^2 \to_p \sigma_V^2$$

Proof. Follows from the earlier observation that $\frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij} - X_{ij}\check{\mu}_2\right)^2 \to_p E(U_1^2)^2 + \sigma_V^2$ and the convergence of $\frac{\hat{U}_i^2}{N}$ to $E(U_1^2)$ (lemma 3). $\square$

5 The no interaction ($\delta = 0$) specification

For this section only, assume that the individual effects are independent of the regressors. In model (1), when $\delta = 0$, the model becomes:

$$Y_{ij} = \sum_l X_{ij,l}\mu_l + A_i + A_j + V_{ij} \tag{20}$$

or, in matrix form:

$$Y = \sum_l \mu_l X_l + A\iota' + \iota A' + V$$

Assume that at iteration m, we obtain an estimator $\hat{\mu}_m$ with

$$\hat{\mu}_m - \mu_0 = -\sigma_A c_{m,N} \times (1, 0, 0, ..., 0)' + Z_{m,N} + O_p\left(\frac{1}{N}\right) \tag{21}$$

Where $c_{m,N}$ is a binary variable taking values $c_{+,m}$ with probability $p_{m,N}$ or $c_{-,m}$ with probability $(1 - p_{m,N})$. $Z_{m,N}$ is a random variable such that for all m, $Z_{m,N} = O_p\left(\frac{1}{\sqrt{N}}\right)$. For instance, if the initial estimator is $\sqrt{N}$ consistent, then $c_{0,N} = 0$, and $Z_{0,N} = \tilde{\mu} - \mu_0$. The reason we allow for a bias in the intercept in equation (21) is that proposition 6 shows that even when we initiate with a consistent estimator under model (20), iterations introduce bias in the intercept.

Then

$$
\begin{aligned}
M(\hat{\mu}_m) &= \sum_l (\mu_{0,l} - \hat{\mu}_{m,l}) X_l + A\iota' + \iota A' + V \\
&= \sigma_A c_{m,N} \iota\iota' - \sum_l Z_{m,N,l} X_l + A\iota' + \iota A' + V - \sigma_A c_{m,N} I_N
\end{aligned}
$$

Lemma 4.

$$\sigma_A c_{m,N} \iota\iota' + A\iota' + \iota A' = e_{1,m} e'_{1,m} - e_{2,m} e'_{2,m}$$

where:

$$e_{1,m} := \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right) \iota + b_{m,N} A \tag{22}$$

$$e_{2,m} := \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right) \iota + b_{m,N} A \tag{23}$$

$$b_{m,N} := \left(\frac{N/4}{||A||^2 + N \frac{c_{m,N}^2}{4} \sigma_A^2 + c_{m,N} \sigma_A \iota' A} \right)^{\frac{1}{4}} \tag{24}$$

moreover:

$$e'_{1,m} e_{2,m} = 0$$

(the dependence on N is omitted in the notation for $e_{1,m}$ and $e_{2,m}$.)

Proof. See section A.11. □

Therefore, the error component of $M(\hat{\mu}_m)$ can be written as the difference of two rank 1 matrices, with eigenvalues that are of a similar order of magnitude ($||e_{1,m}|| \approx ||e_{1,m}||$) and opposite signs. We can state a result equivalent to proposition (13) in the $\delta \neq 0$ specification:

Proposition 6. *Under the specification(20), assume iterations are initiated with an estimator $\tilde{\mu} = \mu_0 - c\sigma_A (1, 0, 0, ..., 0)' + Z_N + O_p\left(\frac{1}{N}\right)$, for $Z_N = O_p\left(\frac{1}{\sqrt{N}}\right)$, if $\hat{\mu} = f_N(\tilde{\mu})$, then:*

$$
\begin{aligned}
\sqrt{N}(\hat{\mu} - \mu_0) = &-\sigma_A \left(h(c) + O_p\left(\frac{1}{\sqrt{N}} \right) \right) (1, 0, ..., 0)' \\
&+ M(c)^{-1} \left(E(X_{12} X'_{23}) + A(c, c_{1,N}) E(X_{12}) E(X'_{12}) \right) \sqrt{N} Z_{m,N}
\end{aligned}
$$

$$
+ B(c, c_{1,N}) \frac{1}{N\sqrt{N}} M(c)^{-1} \sum_{ij} X_{ij} A_j + O_p \left(\frac{1}{\sqrt{N}} \right)
$$

$$
= -\sigma_A \left(h(c) + O_p \left(\frac{1}{\sqrt{N}} \right) \right) (1, 0, ..., 0)' + O_p(1)
$$

where $|h|$ *is deterministic with* $|h(c)| = |h(-c)| > |c|$, $\forall c$, *and* $|h(c)| \to_{|c| \to \infty} \infty$, A *and* B *are deterministic scalar functions and* $c_{1N} \to_p 1 - 2 \times Bern(0.5)$,

Proof. Refer to appendix A.12. □

Two observations are in order. First, even when the initial estimator is consistent for all parameter including the intercept, in which case $c_0 = 0$, the first iteration estimator has a biased intercept, with a bias of order of magnitude $|\sigma_A h(c)| > 0$. Subsequently, the following iterations deliver estimators that are on an explosive path, since $|h(h(c))| = |h(|h(c)|)| > |h(c)|$ for all c. In particular, this implies that the iterative process described in theorem 1 cannot converge.

On the other side, proposition 6 guarantees that all coefficients other than the intercept remain $\sqrt{N}$ consistent following a single iteration, regardless of the bias in the initiating estimator.

6 Empirical illustration and simulation study

6.1 Empirical illustration

To illustrate the use of our new estimator in real world settings, we run our estimation procedure on trade data in line with Rose (2004). Rose (2004) uses a standard gravity model to examine whether joining the World Trade Organization increases trade. Using Rose (2004)'s data set, we estimate a gravity model for year 1999 by regressing $log(Trade)$ between the countries i and j, on indicators of whether both countries are in the World Trade Organization (WTO), only 1 is in the WTO, and a dummy variable GSP describing whether the countries extend each other preferential trade treatment under the Operation and Effects of the Generalized System of Preferences published by the UN. In addition to these three main variables of interest, and following Rose (2004), we regress on a number of other country pair observables (a total of 15 regressors, plus the intercept).

The data set concerns $N = 157$ countries. Out of the $\frac{N(N-1)}{2} = 12246$ possible country pairs, 7268 pairs show a non-null trade volume for the year 1999.

We estimate the regression coefficients using the standard OLS estimator (table 2) following Rose (2004)'s cross-sectional study (table 2 in Rose (2004)) then we use the *least eigevalues* *estimator* described in the earlier section of this paper (table 1). Comparing the two tables, as expected, the standard errors are lower for the least eigenvalues estimator than for the OLS.

Table 1: The least eigenvalues estimator for the slope parameters on trade data for year 1999. The explained variable is *log real trade*. The intercept is not reported.

Variable	Coefficient	Std. Error	t-statistic	P-value
Both in WTO	-0.479	0.072	-6.668	< 0.001
One in WTO	-0.322	0.070	-4.617	< 0.001
GSP	0.305	0.034	8.860	< 0.001
Log distance	-1.181	0.019	-61.441	< 0.001
Log product real GDP	0.829	0.010	83.992	< 0.001
Log product real GDP p/c	-0.033	0.012	-2.838	0.004
Regional FTA	0.679	0.081	8.432	< 0.001
Currency union	0.592	0.141	4.203	< 0.001
Common language	0.369	0.041	9.071	< 0.001
Land border	0.793	0.083	9.589	< 0.001
Number landlocked	-0.375	0.029	-12.901	< 0.001
Number islands	0.017	0.036	0.465	0.642
Log product land area	-0.071	0.008	-9.268	< 0.001
Common colonizer	0.863	0.055	15.825	< 0.001
Ever colony	1.246	0.105	11.888	< 0.001

6.2 Simulations

I run $S = 10000$ simulations on each of the 4 following designs, with a network of $N = 100$ nodes in each simulation.

1. An intercept and an additive regressor , with $\gamma = 0$

$$Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_i A_j + V_{ij}$$

2. An intercept and a multiplicative regressor, with $\gamma = 0$

$$Y_{ij} := \beta_{0,1} + \beta_{0,2} X_i X_j + A_i A_j + V_{ij}$$

3. An intercept and an additive regressor, with $\gamma = 1$

$$Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_i + A_j + A_i A_j + V_{ij}$$

Table 2: The ordinary least squares estimator for the slope parameters on trade data for year 1999. The explained variable is *log real trade*. The intercept is not reported.

Variable	Coefficient	Std. Error	t-statistic	P-value
Both in WTO	-0.269	0.096	-2.798	0.005
One in WTO	-0.320	0.097	-3.312	0.001
GSP	0.199	0.045	4.418	< 0.001
Log distance	-1.073	0.025	-42.185	< 0.001
Log product real GDP	0.944	0.011	85.292	< 0.001
Log product real GDP p/c	-0.034	0.014	-2.501	0.012
Regional FTA	0.946	0.108	8.754	< 0.001
Currency union	0.757	0.194	3.906	< 0.001
Common language	0.415	0.053	7.860	< 0.001
Land border	0.965	0.114	8.451	< 0.001
Number landlocked	-0.540	0.034	-16.078	< 0.001
Number islands	0.022	0.041	0.540	0.589
Log product land area	-0.076	0.009	-8.526	< 0.001
Common colonizer	0.966	0.074	13.078	< 0.001
Ever colony	1.113	0.141	7.865	< 0.001

4. An intercept and a multiplicative regressor, with $\gamma = 1$

$$Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + A_i + A_j + A_iA_j + V_{ij}$$

for each of the two designs $X \sim Unif(0,1)$, $\beta_{0,1} = \beta_{0,2} = E(A_1^2) = E(V_{12}^2) = 1$.[7] The histograms for the estimated slope parameters $\beta_{0,2}$ are in the figures 6 to 9. In each graph, we show the histogram for the OLS estimator (in bleu) on the original model (1) as a benchmark, the estimator $\hat{\mu}_{EIG}$ defined in this paper in green. The OLS estimator is semi-parametrically efficient in the model without individual effects, $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_{ij} + V_{ij}$ as a "gold standard" in orange, it is also estimated for each of the simulations and displayed in orange in the figures 6 to 9 as an oracle estimator. The estimators for the intercepts are not shown since the slope parameter are our concern in this paper. As discussed in the introduction, our estimator is $N-$ consistent for $\beta_{0,1} - \delta\gamma = 1 - 1 = 0$ rather than for $\beta_{0,1} = 1$. The term $\delta\gamma$ can't be estimated at a higher rate than $\sqrt{N}$. Any estimator for $\beta_{0,1}$ based on our estimator and an estimated correction for $\delta\gamma$ would only yield a $\sqrt{N}$-consistent estimator, even though $\beta_{0,1} - \delta\gamma$ is estimated at rate N.

The first two histograms (figures 6 and 7) confirm the result in proposition 2. The histogram for the eigenvalue-corrected estimator (in green) is close to the oracle (orange). On

[7] I also generated simulations with $X \sim \mathcal{N}(0,1)$ or $X \sim 1 + \mathcal{N}(0,1)$ and the outcomes are similar.

both histogram, the OLS estimator (blue) seems to have a larger variance. In fact, the OLS estimator has a non standard asymptotic distribution (cf. Menzel (2021)) and its distribution is slightly skewed to the left. The skew is not visible in figures 6 and 7, because the variance of A is not large enough (see figure 10 for a version of figure 7 with a $Var(A) = 100$ and where the skew is now obvious on the OLS estimator, whereas the eigenvalue corrected estimator is unaffected).

Figures 8 and 9 show that the histogram of the OLS estimator (blue) is much less concentrated than the eigenvalue-corrected estimator (green). This reflects the prediction of corollary 4. The eigenvector corrected estimator is itself less efficient than the oracle (orange), but is rate optimal.

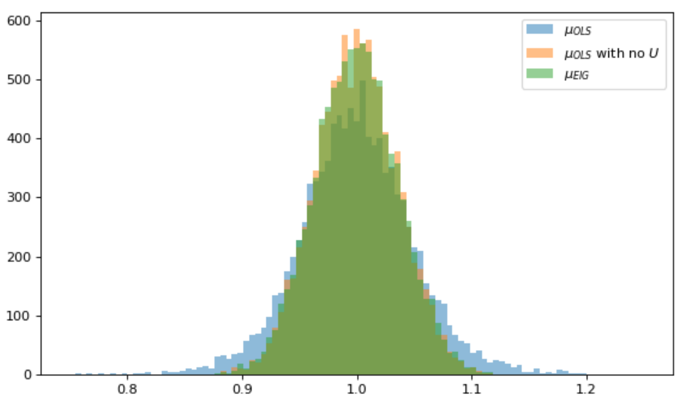

Figure 6: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_iA_j + V_{ij}$, and the "oracle" OLS estimator (orange) for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$.*

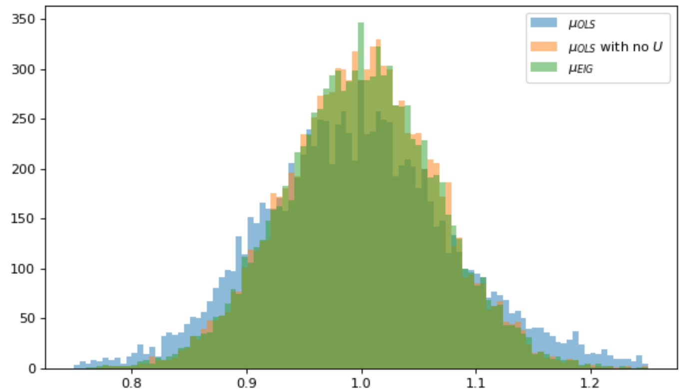

Figure 7: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*.*

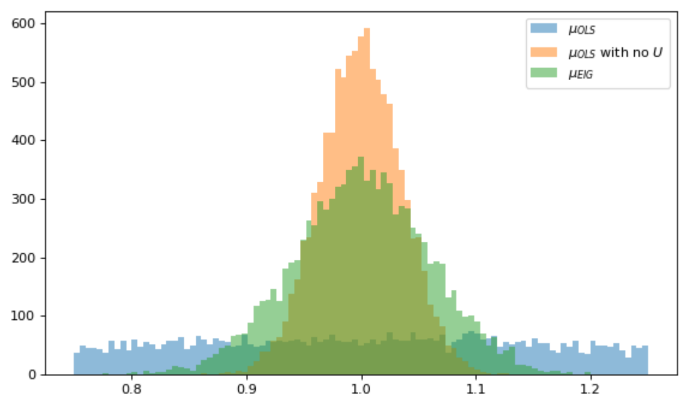

Figure 8: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_i + A_j + A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*.*

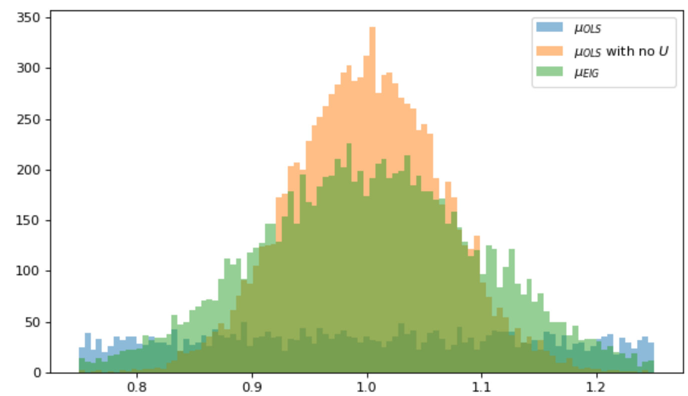

Figure 9: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + A_i + A_j + A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*.*

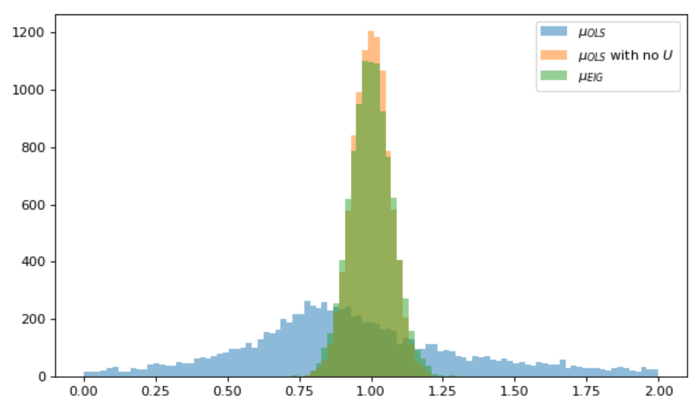

Figure 10: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + 10 \times A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*. Note the 10 factor multiplying the* A_iA_j *term to amplify the skew of the OLS estimator.*

7 Conclusion

In this paper, we proposed a new two iteration estimator for the dyadic non-oriented linear regression model with interactive effects. The new estimator is asymptotically equivalent to the *"infinite iterations"* estimator on an iterative process similar to Bai (2009)'s. The new estimator emerges from a new proof for the the iterations' limit distribution examining one iteration at a time. We also show that in the absence of interaction, the iterative process does not converge, with an estimated intercept that explodes through iterations . Because the alternative estimator requires only a finite number of iterations, and because iterations only bias the intecept, the alternative estimator is still well defined and is shown to be $\sqrt{N}$-consistent for all slope parameters, excluding the intercept.

Technically, studying the asymptotic distribution of $M(\mu)$'s largest eigenvalue up to a second order is the main challenge. The results in this papers hint at how similar 2 iteration estimators could be computed for models with higher order interactions. For higher order interactions, however, the proof would require the computation of the joint distribution of a number of largest eigenvalues, which would be technically challenging. I leave this extension for future work.

References

John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, 1999. doi: https://doi.org/10.1111/1468-0262.00020. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00020.

James E. Anderson and Eric van Wincoop. Gravity with gravitas: A solution to the border puzzle. *The American Economic Review*, 93(1):170–192, 2003. ISSN 00028282. URL http://www.jstor.org/stable/3132167.

Miguel A. Arcones. A remark on approximate m-estimators. *Statistics & Probability Letters*, 38(4):311–321, 1998. ISSN 0167-7152. doi: https://doi.org/10.1016/S0167-7152(98)00038-8. URL https://www.sciencedirect.com/science/article/pii/S0167715298000388.

Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279, 2009. doi: https://doi.org/10.3982/ECTA6135. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6135.

Florent Benaych-Georges and Antti Knowles. Lectures on the local semicircle law for wigner matrices, 2016. URL https://arxiv.org/abs/1601.04055.

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa. A distributional framework for matched employer employee data. *Econometrica*, 87(3):699–739, 2019. doi: https://doi.org/10.3982/ECTA15722. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15722.

James R. Bunch, Christopher P. Nielsen, and Danny C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 31(1):31–48, Mar 1978. ISSN 0945-3245. doi: 10.1007/BF01396012. URL https://doi.org/10.1007/BF01396012.

Matias D. Cattaneo, Richard K. Crump, and Michael Jansson. Generalized jackknife estimators of weighted average derivatives. *Journal of the American Statistical Association*, 108 (504):1243–1256, 2013. ISSN 01621459. URL http://www.jstor.org/stable/24247054.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(87)90015-7. URL https://www.sciencedirect.com/science/article/pii/0304407687900157.

Marcel Fafchamps and Flore Gubert. Risk sharing and network formation. *The American Economic Review*, 97(2):75–79, 2007. ISSN 00028282. URL http://www.jstor.org/stable/30034424.

Thibault Fally. Structural gravity and fixed effects. *Journal of International Economics*, 97 (1):76–85, 2015. ISSN 0022-1996. doi: https://doi.org/10.1016/j.jinteco.2015.05.005. URL https://www.sciencedirect.com/science/article/pii/S0022199615001026.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1 (3):233–241, Sep 1981. ISSN 1439-6912. doi: 10.1007/BF02579329. URL https://doi.org/10.1007/BF02579329.

Bryan S. Graham. Chapter 2 - network data. In Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, editors, *Handbook of Econometrics, Volume 7A*, volume 7 of *Handbook of Econometrics*, pages 111–218. Elsevier, 2020. doi: https://doi.org/10.1016/bs.hoe.2020.05.001. URL https://www.sciencedirect.com/science/article/pii/S1573441220300015.

Bryan S Graham, Fengshi Niu, and James L Powell. Minimax risk and uniform convergence rates for nonparametric dyadic regression. Working Paper 28548, National Bureau of Economic Research, March 2021. URL http://www.nber.org/papers/w28548.

Gary Harris and Clyde Martin. Shorter notes: The roots of a polynomial vary continuously as a function of the coefficients. *Proceedings of the American Mathematical Society*, 100(2): 390–392, 1987. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/2045978.

C. Kirabo Jackson, Jonah E. Rockoff, and Douglas O. Staiger. Teacher effects and teacher-related policies. *Annual Review of Economics*, 6(1):801–825, 2014. doi: 10.1146/annurev-economics-080213-040845. URL https://doi.org/10.1146/annurev-economics-080213-040845.

V.S. Korolyuk and Y.V. Borovskich. *Theory of U-Statistics*. Mathematics and Its Applications. Springer Netherlands, 2013. ISBN 9789401735155. URL https://books.google.com/books?id=RF7tCAAAQBAJ.

Jan R. Magnus. On differentiating eigenvalues and eigenvectors. *Econometric Theory*, 1(2): 179–191, 1985. ISSN 02664666, 14694360. URL http://www.jstor.org/stable/3532409.

Konrad Menzel. Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*, 89(5):2143–2188, 2021. doi: https://doi.org/10.3982/ECTA15383. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15383.

Hyungsik Roger Moon and Martin Weidner. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4):1543–1579, 2015. doi: https://doi.org/10.3982/ECTA9382. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9382.

Hyungsik Roger Moon and Martin Weidner. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1):158–195, 2017. ISSN 02664666, 14694360. URL http://www.jstor.org/stable/26172566.

Hyungsik Roger Moon and Martin Weidner. Nuclear norm regularized estimation of panel regression models, 2023.

Whitney K. Newey. Semiparametric efficiency bounds. *Journal of Applied Econometrics*, 5(2):99–135, 1990. doi: https://doi.org/10.1002/jae.3950050202. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.3950050202.

James L. Powell, James H. Stock, and Thomas M. Stoker. Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430, 1989. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1913713.

Andrew K. Rose. Do we really know that the wto increases trade? *American Economic Review*, 94(1):98–114, March 2004. doi: 10.1257/000282804322970724. URL https://www.aeaweb.org/articles?id=10.1257/000282804322970724.

John D Sargan. Wages and prices in the united kingdom: a study in econometric methodology. *Econometric analysis for national economic planning*, 16:25–54, 1964.

A Proofs and intermediary results

This section details the proofs of all the results in the paper. It begins by showing how the OLS estimator of the intercept (in model (1)) can be adjusted to obtain a $\sqrt{N}$- consistent estimator of the modified estimator (in model 2). Then we provide the technical ingredients (propositions 8 and 9) that our main results heavily rely on.

A.1 Adjustment to the intercept

Proposition 7. *Under model (1) and under the assumptions of theorem 1,* $\gamma \geq 0$, $\delta \in \{-1, 1\}$, $\sigma_U^2 = E(A_1^2) \neq 0$. *Let* $\tilde{\beta}_1$ *be a* $\sqrt{N}$*-consistent estimator of the intercept* β_0 *in equation (1). Then* $\mu_{0,1}$*, the intercept in the modified model (2) is equal to:* $\mu_{0,1} = \beta_{0,1} - \delta\gamma^2$*. Define*

$$\begin{aligned}
\epsilon_{ij} &:= \gamma(A_i + A_j) + \delta A_i A_j + V_{ij} \\
a &:= E(\epsilon_{12}\epsilon_{23}) = \gamma^2 E(A_i^2) \\
b &:= E(\epsilon_{12}\epsilon_{23}\epsilon_{31}) = 3\delta\gamma^2 E(A_i^2)^2 + \delta E(A_i^2)^3
\end{aligned}$$

Then $|\beta| = 3\delta\gamma^2 E(A_i^2)^2 + \delta E(A_i^2)^3$ *and* $E(A_i^2)$ *is the unique real root of the polynomial* $P(x; a, |b|) := x^3 + 3ax - |b|$ *. Denote*

$$\begin{aligned}
\hat{\epsilon}_{ij} &:= Y_{ij} - \sum_{l=1}^{L} X_{ij,l}\tilde{\beta}_l = \sum_{l=1}^{L} X_{ij,l}(\beta_{0,l} - \tilde{\beta}_l) + \epsilon_{ij} \\
\hat{a} &:= \frac{1}{N^3} \sum_{i \neq j \neq k} \hat{\epsilon}_{ij}\hat{\epsilon}_{ik} \\
\hat{b} &:= \frac{1}{N^3} \sum_{i \neq j \neq k} \hat{\epsilon}_{ij}\hat{\epsilon}_{ik}\hat{\epsilon}_{jk} \\
\tilde{\delta} &:= sign(\hat{b})
\end{aligned}$$

Let $\hat{\sigma}_U^2$ *be a real root of the polynomial* $P(x; \hat{a}, \hat{b})$ *and define* $\hat{\gamma}^2 := \frac{\hat{a}}{\hat{\sigma}_U^2}$*. We have*

$$\begin{aligned}
|\hat{\sigma}_U^2 - E(A_1^2)| &= O_p\left(\frac{1}{\sqrt{N}}\right) \\
|\hat{\gamma}^2 - \gamma^2| &= O_p\left(\frac{1}{\sqrt{N}}\right) \\
\mu_{0,1} - \tilde{\mu}_1 &= O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

where $\tilde{\mu}_1 := \hat{\beta}_0 - \hat{\delta}\hat{\gamma}^2$

Proof. That P has a unique real solution whenever $a \geq 0$ results from the observation that $\lim_{x \to -\infty} P(x; a, |b|) = -\infty$, $\lim_{x \to +\infty} P(x; a, |b|) = +\infty$ and $P(., a, b)$ is strictly increasing when $a \geq 0$.

Observe that

$$\begin{aligned} |\hat{a} - a| &= O_p\left(\frac{1}{\sqrt{N}}\right) \\ |\hat{b} - b| &= O_p\left(\frac{1}{\sqrt{N}}\right) \end{aligned}$$

The roots of a polynomial being continuous in its coefficients (e.g. Harris and Martin (1987)), the continuous mapping theorem proves the consistency of $\hat{\sigma}_U^2$.

Moreover, note that $\delta = sign(b)$, that $b \neq 0$, and that

In addition, by the mean value theorem, for some $(\bar{x}, \bar{a}, \bar{b})$ between $(E(U_1^2), a, b)$ and $(\hat{\sigma}_U^2, \hat{\alpha}, \hat{\beta})$

$$\begin{aligned} 0 = P(\hat{\sigma}_U^2; \hat{a}, \hat{b}) &= P(E(A_1^2); a, b) + \frac{\partial P}{\partial x}(\bar{x}, \bar{a}, \bar{b})(\hat{\sigma}_U^2 - E(A_1^2)) + \frac{\partial P}{\partial a}(\hat{a} - a) + \frac{\partial P}{\partial |b|}(|\hat{b}| - |b|) \\ &= (3\bar{x}^2 + 3\bar{a})(\hat{\sigma}_U^2 - E(A_1^2)) + 3\bar{x}(\hat{a} - a) - (|\hat{b}| - |b|) \end{aligned}$$

because $\left| |\hat{b}| - |b| \right| \leq |\hat{b} - b|$, then $|\hat{b}| - |b| = O_p\left(\frac{1}{\sqrt{N}}\right)$, implying:

$$\hat{\sigma}_U^2 - E(A_1^2) = \frac{-1}{(3\bar{x}^2 + 3\bar{a})}\left(3\bar{x}(\hat{a} - a) - (|\hat{b}| - |b|)\right) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

Finally

$$\begin{aligned} \hat{\gamma}^2 - \gamma^2 &= \frac{\hat{a}}{\hat{\sigma}_U^2} - \frac{a}{E(A_1^2)} \\ &= \frac{\hat{a}E(A_1^2) - a\hat{\sigma}_U^2}{\hat{\sigma}_U^2 E(A_1^2)} \\ &= \frac{(\hat{a} - a)E(A_1^2) + a(E(A_1^2) - \hat{\sigma}_U^2)}{\hat{\sigma}_U^2 E(A_1^2)} \\ &= O_p\left(\frac{1}{\sqrt{N}}\right) \end{aligned}$$

$\square$

A.2 On the distribution of the largest eigenvalue

Proposition 8. *Let $A = (a_{ij})$ be a matrix such that:*

$$a_{ij} = U_i U_j + V_{ij} \textit{ for all } i \neq j$$

and $a_{ii} = 0$ for all i, where the V_{ij}'s for $i \neq j$ are i.i.d. mean 0 random variables with variance σ_v and $V_{ij} = V_{ji}$, and the diagonal entries of V given by $V_{ii} = E(U_1)^2 - U_i^2$.
The U's are also i.i.d but not necessarily centered. Let $\lambda_1(A) > \lambda_2(A)$.... be A's eigenvalues. Then:

$$\lambda_1(A) = U'U + \frac{U'VU}{U'U} + \frac{U'V^2U}{(U'U)^2} - E(U_1)^2 + o_p(1)$$

Proof. The proof draws from Füredi and Komlós (1981). In all what follows,"with high probability (w.h.p.)" means"with probability approaching 1 as N grows". Write

$$A = UU' + V - E(U_1)^2 I_N$$

define $\tilde{A} := A + E(U_1)^2 I_N$ and decompose U into $U = v + r$ such that $r'v = 0$ and $\tilde{A}v = \lambda_1 v$. We first show that, with high probability, r is bounded.

Define:

$$S := \tilde{A}U = (U'U)U + VU = \lambda_1 v + Ar$$

define:

$$L := E(S|U) = (U'U)U$$

therefore:

$$L_i = (U'U)U_i$$

Notice

$$||Ar||^2 = r'\tilde{A}'\tilde{A}r \leq \lambda_2(\tilde{A}'\tilde{A}) \times ||r||^2 = max_{i>1}|\lambda_i(\tilde{A})|^2 \times ||r||^2$$

where the inequality results from the Courant-Fisher theorem (equation (11) in Füredi and Komlós (1981)) and the second equality results from: $\tilde{A}'\tilde{A} = \tilde{A}^2$. Therefore:

$$||\tilde{A}r|| \leq max_{i>1}|\lambda_i(\tilde{A})| \times ||r||$$

By a standard result on rank 1 modifications (e.g. Bunch et al. (1978)), for all $i > 1$

$$\lambda_i(V) \leq \lambda_i(\tilde{A}) \leq \lambda_{i-1}(V)$$

So :

$$max_{i>1}|\lambda_i(\tilde{A})| \leq \max\{|\lambda_N(V)|, \lambda_1(V)\}$$

By theorem 2 in Füredi and Komlós (1981) , almost surely:

$$\max\{|\lambda_N(V)|, \lambda_1(V)\} = 2\sigma_v\sqrt{N} + o\left(N^{\frac{1}{2}}\right)$$

so with high probability, for N large enough:

$$||\tilde{A}r|| \leq \max\{|\lambda_N(V)|, \lambda_1(V)\} \leq 3\sigma_v\sqrt{N}||r|| \tag{25}$$

Thus:

$$||\tilde{A}r - (U'U)r|| \geq (U'U)||r|| - ||\tilde{A}r|| \geq (U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})||r|| \tag{26}$$

implying:

$$||r||^2 \leq \frac{||\tilde{A}r - (U'U)r||^2}{(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2} \leq \frac{||S - L||^2}{(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2} \tag{27}$$

With high probability:

$$(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2 \geq \frac{\sigma_u^4}{2}N^2 \tag{28}$$

The second inequality is a result of Pythagorean theorem.

To show that r is bounded w.h.p., it is left to show that $||S - L||^2$ also grows as N^2. I use Chebychev's inequality on $||S - L||^2$:

$$
\begin{aligned}
E(||S-L||^2|U) &= E\left(\sum_i (S_i - L_i)^2 \middle| U\right) \\
&= E\left(\sum_i \left(\sum_j V_{ij}U_j\right)^2 \middle| U\right) \\
&= E\left(\sum_i U_i^2 V_{ii}^2 + \sum_i \left(\sum_{j\neq i} V_{ij}U_j\right)^2 + 2\sum_i U_i V_{ii} \sum_{j\neq i} V_{ij}U_j \middle| U\right) \\
&= \sum_i U_i^2 V_{ii}^2 + \sigma_v^2 \sum_i \sum_{j\neq i} U_j^2 \\
&= \sum_i U_i^2 (E(U_1)^2 - U_i^2)^2 + \sigma_v^2 (N-1) \sum_i U_i^2
\end{aligned}
$$

so

$$
\frac{E(||S-L||^2|U)}{N^2} \to \sigma_v^2 E(U_1^2) \text{ almost surely.} \tag{29}
$$

Also:

$$
\begin{aligned}
Var(||S-L||^2|U) &= Var\left(\sum_i \left(\sum_j V_{ij}U_j\right)^2 \middle| U\right) \\
&= Var\left(\sum_i U_i^2 V_{ii}^2 + \sum_i \left(\sum_{j\neq i} V_{ij}U_j\right)^2 + 2\sum_i U_i V_{ii} \sum_{j\neq i} V_{ij}U_j \middle| U\right) \\
&= Var\left(\sum_i \left[\left(\sum_{j\neq i} V_{ij}U_j\right)^2 + 2U_i V_{ii} \sum_{j\neq i} V_{ij}U_j\right] \middle| U\right) \\
&= \sum_{i,l} Cov\left(\left(\sum_{j\neq i} V_{ij}U_j\right)^2 + 2U_i V_{ii} \sum_{j\neq i} V_{ij}U_j, \left(\sum_{j\neq l} V_{lj}U_j\right)^2 + 2U_l V_{ll} \sum_{j\neq l} V_{lj}U_j \middle| U\right) \\
&= \sum_{i,l} Cov\left(\left(\sum_{j\neq i} V_{ij}U_j\right)^2, \left(\sum_{j\neq l} V_{lj}U_j\right)^2 \middle| U\right) \\
&+ 4\sum_{i,l} U_i V_{ii} U_l V_{ll} Cov\left(\sum_{j\neq i} V_{ij}U_j, \sum_{j\neq l} V_{lj}U_j \middle| U\right)
\end{aligned}
$$

$$+ 4\sum_{i,l} U_i V_{ii} Cov\left(\left(\sum_{j\neq i} V_{ij}U_j\right)^2, \sum_{j\neq l} V_{lj}U_j \middle| U\right)$$

Hence:

$$\begin{aligned}
Var(||S-L||^2|U) &= \sum_{i,l}\sum_{j_1,j_2\neq i}\sum_{k_1,k_2\neq l} U_{j_1}U_{j_2}U_{k_1}U_{k_2} Cov\left(V_{ij_1}V_{ij_2}, V_{lk_1}V_{lk_2}\right)\\
&+ 4\sum_{i,l}\sum_{j\neq i}\sum_{k\neq l} U_iV_{ii}U_lV_{ll}U_jU_k Cov(V_{ij}, V_{lk})\\
&+ 4\sum_{i,l}\sum_{j_1,j_2\neq i}\sum_{k\neq l} U_iV_{ii}U_{j_1}U_{j_2}U_k Cov\left(V_{ij_1}V_{ij_2}, V_{lk}\right)\\
&= 2\sigma_v^4\sum_i\sum_{j,k\neq i,k\neq j} U_j^2U_k^2 + Var(V_{ij}^2)\sum_i\sum_{j\neq i} U_j^4 + Var(V_{ij}^2)\sum_i\sum_{j\neq i} U_j^2U_i^2\\
&+ 4\sigma_v^2\sum_i\sum_{j\neq i} U_i^2U_j^2V_{ii}^2 + 4\sigma_v^2\sum_i\sum_{j\neq i} U_iV_{ii}U_jV_{jj}U_jU_i\\
&+ 4E(V_{12}^3)\sum_i\sum_{j\neq i} U_i^4V_{ii} + 4E(V_{12}^3)\sum_i\sum_{j\neq i} U_i^2V_{ii}U_j^2
\end{aligned}$$

so there exists a constant $c_1 \geq 0$ such that

$$\frac{Var(||S-L||^2|U)}{N^3} \to c_1 \text{ almost surely.}$$

By Chebychev's inequality:

$$\mathbb{P}\left(\left| ||S-L||^2 - E(||S-L||^2|U)\right| \geq \sqrt{Var(||S-L||^2|U)}N^{1/3}\right) \leq \frac{1}{N^{2/3}} \tag{30}$$

By (29) and (30), with high probability:

$$||S-L||^2 \leq 2N^2E(U_1^2)\sigma_v^2 \tag{31}$$

Combining (27),(28) and (31), with high probability:

$$||r||^2 \leq \frac{4E(U_1^2)}{\sigma_v^2} \tag{32}$$

Now note that:

$$\frac{\sum_i S_i^2}{\sum_i S_i U_i} = \frac{S'S}{S'U} = \frac{\lambda_1^2 ||v||^2 + ||\tilde{A}r||^2}{\lambda_1 ||v||^2 + r'\tilde{A}r} = \lambda_1 + \frac{||\tilde{A}r||^2 - \lambda_1 r'\tilde{A}r}{\sum_i S_i U_i} \tag{33}$$

let's now show that $\left| \frac{||\tilde{A}r||^2 - \lambda_1 r \tilde{A}r}{\sum_i S_i U_i} \right| = O\left(\frac{1}{\sqrt{N}}\right)$. From (25), w.h.p.:

$$||\tilde{A}r||^2 \leq 9\sigma_v^2 N ||r||^2$$

then by (32)

$$||\tilde{A}r||^2 \leq 9\sigma_v^2 \frac{4E(U_1^2)}{\sigma_v^2} N = 36E(U_1^2)N$$

then:

$$|r'\tilde{A}r| \leq ||r|| \times ||\tilde{A}r|| \leq \frac{2\sqrt{E(U_1^2)}}{\sigma_v} \times 6\sqrt{E(U_1^2)}\sqrt{N} = 12\frac{E(U_1^2)}{\sigma_v}\sqrt{N}$$

To bound $\lambda_1(\tilde{A})$, note that $\tilde{A}v = \lambda_1 v$. So $|\lambda_1(\tilde{A})||v_i| = |\sum_{j \neq i} a_{ij} v_j - E(U_1)^2 v_i| \leq \max_j |v_j|(E(U_1)^2 + \sum_{j \neq i} |a_{ij}|$. Taking a max over the i's: $|\lambda_1(\tilde{A})| \max_i |v_i| \leq \max_j |v_j| \times \max_i \sum_j |a_{ij}|$, therefore: $|\lambda_1| \leq E(U_1)^2 + \max_i \sum_{j \neq i} |a_{ij}|$. For any $\eta > 0$, Markov's inequality shows that $\max_i \sum_j |a_{ij}| = o_p(N^{1+\eta})$
Finally:

$$\sum_i S_i U_i = S'U = (U'U)^2 + U'VU = (\sum_i U_i^2)^2 + \sum_{i \neq j} U_i U_j V_{ij} + \sum_i U_i^2 (U_i^2 - \sigma_u^2)$$

so, almost surely,

$$\frac{1}{N^2} \sum_i S_i U_i = E(U_1^2)^2 + o_p(1) \tag{34}$$

implying that:

$$\begin{aligned} \lambda_1 &= \frac{\sum_i S_i^2}{\sum_i S_i U_i} + o_p(1) \\ &= \frac{(U'U)^3 + 2(U'U)U'VU + U'V^2U}{(U'U)^2 + U'VU} + o_p(1) \\ &= U'U + \frac{U'VU}{U'U} + \frac{(U'U)U'V^2U - (U'VU)^2}{(U'U)((U'U)^2 + U'VU)} + o_p(1) \end{aligned} \tag{35}$$

Note that, by the CLT $U'VU = O_p(N)$, and note that $U'V^2U = O_p(N^2)$ so

$$\lambda_1(\tilde{A}) = U'U + \frac{U'VU}{U'U} + \frac{U'V^2U}{(U'U)^2} + o_p(1)$$

or

$$\lambda_1(A) = \lambda_1(\tilde{A}) - E(U_1)^2 = U'U + \frac{U'VU}{U'U} + \frac{U'V^2U}{(U'U)^2} - E(U_1)^2 + o_p(1)$$

□

Proposition 9. *Fix some vector $\mu_0 \in \mathbb{R}^L$. For all μ, denote $M(\mu)$ the matrix:*

$$M(\mu) := X(\mu_0 - \mu) + UU' + V - E(U_1^2)I_N$$

where U and V are defined as in Proposition (8), and X is a linear function of the vector $(\mu_0 - \mu)$: $X = \sum_{l=1}^{L}(\mu_{0,l} - \mu_l)X_l$, with L a fixed, known number, X_l are symmetric matrices with zeros on the diagonal and such that $\lambda_1(X) := \max_{l=1..L} \lambda_1(X_l) = O_p(N)$.
Let $\lambda_1(\mu) > \lambda_2(\mu)... > \lambda_N(\mu)$ be the eignevalues of $M(\mu)$, then:

$$\lambda_1(\tilde{\mu}) = U'U + \frac{\sum_k(\mu_{0,k} - \tilde{\mu}_k)U'X_kU}{U'U} + O_p(1)$$

Moreover, define $v(\mu)$ and $r(\mu)$ the vectors such that:

1. $U = v(\mu) + r(\mu)$
2. $v(\mu)'r(\mu) = 0$
3. $M(\mu)v(\mu) = \lambda_1(\mu)v(\mu)$

Let $\tilde{\mu}$ be an estimator for μ_0 such that $||\tilde{\mu} - \mu_0|| = O_p\left(\frac{1}{\sqrt{N}}\right)$. Then

1. $$||U - v(\tilde{\mu})|| = O_p(1)$$
2. $$||U||^2 - ||v(\tilde{\mu})||^2 = ||r(\tilde{\mu})||^2 = O_p(1)$$
3. *for any $l, l' = 1..K$:*
$$v(\tilde{\mu})'X_{l'}X_lv(\tilde{\mu}) = U'X_{l'}X_lU + O_p\left(N^2\sqrt{N}\right)$$

Proof. Note that

$$||M(\tilde{\mu})r(\tilde{\mu})|| \le |\lambda_2(\tilde{\mu})| \times ||r(\tilde{\mu})||$$

and: for all $i = 2...N$

$$\lambda_i(M(\tilde{\mu}) - UU') \le \lambda_i(M(\tilde{\mu})) \le \lambda_{i-1}(M(\tilde{\mu}) - UU')$$

and by Weyl's inequalities:

$$-||\mu_0-\tilde{\mu}||\times|\lambda_1(X)|+\lambda_i(V-E(U_1^2)I_N) \leq \lambda_i(M(\tilde{\mu})-UU') \leq ||\mu_0-\tilde{\mu}||\times|\lambda_1(X)|+\lambda_i(V-E(U_1^2)I_N)$$

so:

$$|\lambda_2(M(\tilde{\mu})| \leq \max\{\lambda_1(V), |\lambda_N(V)|\} + E(U_1^2) + ||\mu_0 - \tilde{\mu}|| \times |\lambda_1(X)|$$

by Theorem 2 in Füredi and Komlós (1981), almost surely:

$$\max\{|\lambda_N(V)|, \lambda_1(V)\} = 2\sigma_v\sqrt{N} + o\left(N^{\frac{1}{2}}\right)$$

so:

$$|\lambda_2(M(\tilde{\mu})| = O_p(\sqrt{N})$$

as in the proof of proposition (8), with high probability:

$$||M(\tilde{\mu})r(\tilde{\mu}) - (U'U)r(\tilde{\mu})|| \geq (U'U)||r(\tilde{\mu})|| - ||M(\tilde{\mu})r(\tilde{\mu})|| \geq ((U'U) - |\lambda_2(M(\tilde{\mu}))|)||r(\tilde{\mu})||$$

so with high probability:

$$\begin{aligned}
||r(\tilde{\mu})||^2 &\leq \frac{||M(\tilde{\mu})r(\tilde{\mu}) - (U'U)r(\tilde{\mu})||^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2} \\
&\leq \frac{||M(\tilde{\mu})U - (U'U)U||^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2} \\
&\leq \frac{\left(||M(\tilde{\mu})U - M(\mu_0)U|| + ||M(\mu_0)U - (U'U)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2} \\
&= \frac{\left(||\sum_l(\mu_{0,l} - \tilde{\mu}_l)X_lU|| + ||M(\mu_0)U - (U'U)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2} \\
&\leq \frac{\left(\sum_l |\mu_{0,l} - \tilde{\mu}_l| \times |\lambda_1(X_l)| \times ||U|| + ||VU - E(U_1^2)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2} \\
&= \frac{\left(\sum_l |\mu_{0,l} - \tilde{\mu}_l| \times |\lambda_1(X_l)| \times ||U|| + ||S - L|| + E(U_1^2)||U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\mu}))|)^2}
\end{aligned}$$

where S and L are defined in the proof for equation (8). By equation (31), with high probability:

$$||S - L|| \leq \sqrt{2}N\sqrt{E(U_1^2)}\sigma_v$$

so

$$||r(\tilde{\mu})|| = O_p(1)$$

which proves the first result:

$$||U - v(\tilde{\mu})|| = O_p(1)$$

Also, as in equation (35):

$$\begin{aligned}
\lambda_1(\tilde{\mu}) &= \frac{U'M(\tilde{\mu})'M(\tilde{\mu})U}{U'M(\tilde{\mu})U} + o_p\left(1\right) \\
&= \frac{\sum_{k=1}^{K}\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)(\mu_{0,k}-\tilde{\mu}_l)U'X_kX_lU + (U'U)\sum_{k=1}^{K}(\mu_{0,k}-\tilde{\mu}_l)U'X_kU + \sum_{k=1}^{K}(\mu_{0,k}-\tilde{\mu}_l)U'X_kVU}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{-E(U_1^2)\sum_{k=1}^{K}(\mu_{0,k}-\tilde{\mu}_l)U'X_kU}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{(U'U)\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)U'X_lU + (U'U)^3 + (U'U)U'VU - E(U_1^2)(U'U)^2}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)U'VX_lU + (U'U)U'VU + U'V^2U - E(U_1^2)U'VU}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{-E(U_1^2)\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)U'X_lU - E(U_1^2)(U'U)^2 - E(U_1^2)U'VU + E(U_1^2)^2U'U}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} + o_p(1) \\
&= \frac{(U'U)\left(\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U\right)}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\left(\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU/U'U\right)\left(\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U\right)}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\sum_{k=1}^{K}\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)(\mu_{0,k}-\tilde{\mu}_l)U'X_kX_lU + (U'U)(U'VU)}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&\frac{-E(U_1^2)(U'U)^2 - \left(\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU\right)^2/U'U + U'V^2U}{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU + (U'U)^2 + U'VU - E(U_i^2)U'U} + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= U'U + \frac{\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU}{U'U} \\
&+ \sum_{k=1}^{K}\sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)(\mu_{0,k}-\tilde{\mu}_l)\frac{U'X_kX_lU}{(U'U)^2} + \frac{U'VU}{U'U} - E(U_1^2) - \frac{\left(\sum_k(\mu_{0,k}-\tilde{\mu}_k)U'X_kU\right)^2}{(U'U)^3} + \frac{U'V^2U}{(U'U)^2} + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned} \tag{36}$$

as desired.

For the third part of the proposition, note that:

$$M(\tilde{\mu})U = \sum_{l=1}^{K}(\mu_{0,l}-\tilde{\mu}_l)X_lU + (U'U)U + VU - E(U_1^2)U$$

so

$$M(\tilde{\mu})U - \lambda_1(M(\tilde{\mu}))U = VU + (U'U - \lambda_1(M(\tilde{\mu})))U + \sum_{l=1}^{K}(\mu_{0,l} - \tilde{\mu}_l)X_l U - E(U_1^2)U$$

remember:

$$M(\tilde{\mu})U = M(\tilde{\mu})r + \lambda_1(M(\tilde{\mu}))v$$

hence:

$$\begin{aligned}\lambda_1(M(\tilde{\mu}))(U - v(\tilde{\mu})) &= M(\tilde{\mu})r - VU + (\lambda_1(M(\tilde{\mu})) - U'U)U + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})X_l U + E(U_1^2)U \\ &= M(\tilde{\mu})r - VU + O_p(1)U + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})X_l U + E(U_1^2)U - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}U\end{aligned} \tag{37}$$

fix some l in $1..L$, multiplying both sides by $\iota' X_l'$:

$$\begin{aligned}\lambda_1(M(\tilde{\mu}))\iota' X_l'(v(\tilde{\mu}) - U) = -\iota' X_l' M(\tilde{\mu})r(\tilde{\mu}) + \iota' X_l' VU + (U'U - \lambda_1(M(\tilde{\mu})))\iota' X_l' U \\ + \sum_{l'=1}^{L}(\mu_{0,l'} - \tilde{\mu}_{l'})\iota' X_l' X_{l'} U - E(U_i^2)\iota' X_l' U\end{aligned}$$

The the proposition's second result, remember that $r(\tilde{\mu})$ and $v(\tilde{\mu})$ are orthogonal and that $U = v(\tilde{\mu}) + r(\tilde{\mu})$, so by the Pythagorean theorem:

$$||U||^2 = ||v(\tilde{\mu})||^2 + ||r(\tilde{\mu})||^2$$

as desired.

Finally, remember that

$$\lambda_1(M(\tilde{\mu}))(v(\tilde{\mu}) - U) = -M(\tilde{\mu})r + VU + (U'U - \lambda_1(M(\tilde{\mu})))U + \sum_{l=1}^{L}(\mu_{0,l} - \tilde{\mu}_l)X_l U - E(U_1^2)U =: \Delta$$

so

$$X_l v(\tilde{\mu}) = X_l U + \frac{1}{\lambda_1(M(\tilde{\mu}))}X_l \Delta$$

and

$$v(\tilde{\mu})'X_{l'}X_l v(\tilde{\mu}) = U'X_{l'}X_l U + \frac{1}{\lambda_1(M(\tilde{\mu}))^2}\Delta' X_{l'}X_l\Delta + \frac{1}{\lambda_1(M(\tilde{\mu}))}\Delta' X_{l'}X_l U + \frac{1}{\lambda_1(M(\tilde{\mu}))}UX_{l'}X_l\Delta'$$

Note that $||\Delta|| = O_p(N)$, $\lambda_1(M(\tilde{\mu})) = O_p(N)$ and $||X_l\Delta|| = O_p(N^2)$ since by assumption:$\lambda_{\max}(X_l)$, $\lambda_{\min}(X_l)$ $O_p(N)$ (by lemma 2), also

$$||X_lU|| \leq \lambda_{\max}(X_l)||U|| = O_p(N\sqrt{N})$$

so

$$v(\tilde{\mu})'X_{l'}X_l v(\tilde{\mu}) = U'X_{l'}X_l U + O_p(N^2\sqrt{N})$$

□

A.3 Proof of Lemma 2

Assume X satisfies the lemma's assumptions. Let λ be one of X's eigenvalues and let x be a corresponding eigenvector. Then

$$\begin{aligned} |\lambda|||x||_2 &= ||\lambda x||_2 \\ &= ||Xx||_2 \\ &\leq ||X||_2||x||_2 \end{aligned}$$

where $||.||_2$ designates the Euclidean norm for vectors and the spectral norm for matrices. Hence

$$|\lambda| \leq ||X||_2$$

but we know that the spectral normal is smaller than the Forbenius norm for any matrix. Therefore:

$$|\lambda| \leq \sqrt{\sum_{i,j} X_{ij}^2}$$

It is left to show that $\sum_{i,j} X_{ij}^2 = O_p(N^2)$. Decompose

$$\sum_{ij} X_{ij}^2 = \sum_{ij} E(X_{ij}^2|X_i, X_j) + \sum_{ij} X_{ij}^2 - E(X_{ij}^2|X_i, X_j)$$

First, by a U-statistic law of large numbers (e.g. theorem 3.1.3. in Korolyuk and Borovskich (2013)), $\sum_{ij} E(X_{ij}^2|X_i, X_j) = O_p(N^2)$. For the second term in the decomposition, it is enough to note:

$$Var\left(\frac{1}{N^2}\sum_{ij}X_{ij}^2\middle|(X_i)_{i=1}^{\infty}\right)\to 0,\quad \text{almost surely.}$$

A.4 Proof of lemma 1

Proof. First, note that, given the assumption $E(X_{12}'X_{12})$ is invertible. By a standard law of large numbers, the matrix $\frac{1}{N}\sum_{k=1}^{N}X_{2k,2k+1}'X_{2k,2k+1}$ converges almost surely to $E(X_{12}'X_{12})$, then with probability 1, $\frac{1}{N}\sum_{k=1}^{N}X_{2k,2k+1}'X_{2k,2k+1}$ is invertible for N large enough. Under the condition that $\frac{1}{N}\sum_{k=1}^{N}X_{2k,2k+1}'X_{2k,2k+1}$ is invertible:

Write

$$\begin{aligned}
\arg\min_{\mu\in\mathbb{R}^L}\sum_{i=2}^{N}\lambda_i\left(M(\mu)\right)^2 &= arg\min_{\mu}\sum_{i\neq j}\left(Y_{ij}-X_{ij}\mu\right)^2-\max_{\nu:||\nu||=1}\nu' M(\mu)^2\nu\\
&= arg\min_{\mu}\min_{\nu:||\nu||=1}\sum_{i\neq j}\left(Y_{ij}-X_{ij}\mu\right)^2-\sum_{i\neq j,k\neq i,j}\nu_i\nu_j\left(Y_{ik}-X_{ik}\mu\right)\left(Y_{kj}-X_{kj}\mu\right)
\end{aligned}$$

For a fixed ν in the unit sphere, the function that associates each μ to $\sum_{i\neq j}\left(Y_{ij}-X_{ij}\mu\right)^2-\sum_{i\neq j,k\neq i,j}\nu_i\nu_j\left(Y_{ik}-X_{ik}\mu\right)\left(Y_{kj}-X_{kj}\mu\right)$ is twice continuously differentiable with a Hessian equal to:

$$H:=2\left(\sum_{ij}X_{ij}'X_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i\nu_jX_{ik}'X_{jk}\right)$$

let's show that H is definite positive. Fix $\alpha\neq 0$ in $\mathbb{R}^L$, denote: $x_{ij}:=\sqrt{2}X_{ij}\alpha$ and X the matrix with entries x_{ij}. Because X is symmetric, represent ν in an orthonormal basis of eigenvector of X: $\nu=\sum_{i=1}^{N}m_ie_i$, where e_i is a normalized eigenvector of X. Note

$$\begin{aligned}
\alpha' H\alpha &= \sum_{ij}x_{ij}^2-\sum_{i\neq j,k\neq i,j}\nu_i\nu_jx_{ik}x_{jk}\\
&= Trace(X^2)-(X\nu)'(X\nu)+\sum_{i\neq k}\nu_i^2x_{ik}^2\\
&= \sum_{i=1}^{N}\lambda_i(X)^2-\sum_{i=1}^{N}m_i^2\lambda_i(X)^2+\sum_{i\neq k}\nu_i^2x_{ik}^2\\
&= \sum_{i=1}^{N}(1-m_i^2)\lambda_i(X)^2+\sum_{i\neq k}\nu_i^2x_{ik}^2>0
\end{aligned}$$

since $\sum_{i=1}^{N}(1-m_i^2)\lambda_i(X)^2=0$ implies that X is of rank at most 1 and ν is its unique eigenvector (up to a normalization) corresponding to a non null eigenvalue, if X is rank

1. Along with $\nu_i x_{ik} = 0$, this implies that $X = 0$, so $X_{ij}\alpha = 0$ for all i, j. Therefore $\alpha' \frac{1}{N} \sum_{k=1}^{N} X'_{2k,2k+1} X_{2k,2k+1} \alpha = 0$ and the matrix $\frac{1}{N} \sum_{k=1}^{N} X'_{2k,2k+1} X_{2k,2k+1}$ is not invertible; a contradiction.

This proves that, almost surely, when N is large enough, $H(\nu)$ is definite positive for all ν.[8]

For any fixed ν, the function $\sum_{i \neq j} \left(Y_{ij} - X_{ij}\mu\right)^2 - \sum_{i \neq j, k \neq i,j} \nu_i \nu_j \left(Y_{ik} - X_{ik}\mu\right)\left(Y_{kj} - X_{kj}\mu\right)$ is minimized at $\mu^*(\nu)$ that is continuous in ν. So the problem of minimizing

$$\sum_{i \neq j} \left(Y_{ij} - \sum_{l=1}^{L} \mu_l X_{ij,l}\right)^2 - \sum_{i \neq j, k \neq i,j} \nu_i(\hat{\mu}^*) \nu_j(\hat{\mu}^*) \left(Y_{ik} - \sum_{l=1}^{L} \mu_l X_{ik,l}\right)\left(Y_{kj} - \sum_{l=1}^{L} \mu_l X_{kj,l}\right)$$

on the unit circle admits a solution (minimizing a continuous function on a compact).

So let (μ^*, ν^*) be a minimizer of the function $\sum_{i \neq j} \left(Y_{ij} - \sum_{l=1}^{L} \mu_l X_{ij,l}\right)^2 - \nu' M(\mu)^2 \nu$, then:

$$\mu^* = \arg\min \left(Y_{ij} - \sum_{l=1}^{L} \mu_l X_{ij,l}\right)^2 - \nu(\mu^*)' M(\mu)^2 \nu(\mu^*)$$

and

$$\nu^* = \arg \max_{||\nu||_2 = 1} \nu' M(\mu^*)^2 \nu$$

taking a first order condition for μ, we get that μ^* is a fixed point of f_N.

Conversely, let μ^* be a fixed point of f_N. Then μ^* satisfies the first order condition for the minimization of the function:

$$\mu \to \sum_{i \neq j} \left(Y_{ij} - X_{ij}\mu\right)^2 - \sum_{i \neq j, k \neq i,j} \nu_i(\mu^*) \nu_j(\mu^*) \left(Y_{ik} - X_{ik}\mu\right)\left(Y_{kj} - X_{kj}\mu\right)$$

we have shown that this function is strictly convex with probability approaching 1. Therefore μ^* is a minimizer, implying that μ^* minimizes the initial objective function $\mu \to \sum_{i=2}^{N} \lambda_i \left(M(\mu)\right)^2$

□

A.5 Proof of propositions 1 and 2

Proof. Note that the function f_N is symmetric as a function of the data, that is $f_N(Y_N, X_N; \mu) = f_N(-Y_N, -X_N; \mu) = f_N(\delta Y_N, \delta X_N; \mu)$ for all μ and for any sequence of data (Y_N, X_N). Therefore, an iteration using $f_N(Y_N, X_N; .)$ produces the exact same effect as an iteration using

[8] In fact, we have shown that almost surely, for N large enough, $\min_\nu \lambda_N(H(\nu)) > 0$.

the function $f_N(\delta Y_N, \delta X_N; .)$. In other words, given a first stage estimator $\tilde{\mu}$, the estimator $\hat{\mu}$ is numerically the same whether it is computed on the model

$$Y_{ij} = X_{ij}\mu_0 + \delta U_i U_j + V_{ij}$$

or

$$(\delta Y_{ij}) = (\delta X_{ij})\mu_0 + U_i U_j + \delta V_{ij}$$

To ease notation, I will prove the proposition for the case $\delta = 1$. The result for any $\delta \in \{-1, 1\}$ is easily derived through the previous observation.

First, note that:

$$\begin{aligned}(\hat{\mu} - \mu_0) = (1 + o_p(1)) &\left(\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \nu_i(\tilde{\mu}) X_{ij} \right)' \left(\sum_{i=1}^{N} \nu_i(\tilde{\mu}) X_{ij} \right) \right)^{-1} \\ &\times \left(\sum_{i \neq j} X_{ij}' \left(Y_{ij} - \sum_{l=1}^{K} \mu_{0,l} X_{ij,l} \right) - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\mu}) \nu_j(\tilde{\mu}) X_{jk}' \left(Y_{ik} - \sum_{l=1}^{K} \mu_{0,l} X_{ik,l} \right) \right)\end{aligned}$$

the (l, l') entry of the matrix $\sum_{j=1}^{N} \left(\sum_{i=1}^{N} \nu_i(\tilde{\mu}) X_{ij} \right)' \left(\sum_{i=1}^{N} \nu_i(\tilde{\mu}) X_{ij} \right)$ is given by:

$$\begin{aligned}\sum_{i,j,k} \nu_i(\tilde{\mu}) \nu_k(\tilde{\mu}) X_{ij,l} X_{kj,l'} &= \frac{1}{||v||^2} \sum_{i,j,k} v_i(\tilde{\mu}) v_k(\tilde{\mu}) X_{ij,l} X_{kj,l'} \\ &= \frac{1}{||v||^2} \sum_{i,j,k} U_i U_k X_{ij,l} X_{kj,l'} + O_p(N\sqrt{N})\end{aligned}$$

where the last inequality results from proposition 9. Using a U-statistic central limit theorem

$$\frac{1}{N^3} \sum_{i,j,k} U_i U_k X_{ij,l} X_{kj,l'} = E(U_1 U_3 X_{12,l} X_{32,l'}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and, by proposition 9:

$$||U - v(\tilde{\mu})|| = O_p(1)$$

implying

$$\Big| ||U|| - ||v(\tilde{\mu})|| \Big| \leq ||U - v(\tilde{\mu})|| = O_p(1)$$

hence

$$\frac{||v(\tilde{\mu})||^2}{N} = E(U_1^2) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

so

$$\frac{1}{N^2||v(\tilde{\mu})||^2}\sum_{i,j,k} U_i U_k X_{ij,l} X_{kj,l'} = \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,l} X_{32,l'}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and

$$\frac{1}{N^2}\sum_{i,j,k} \nu_i(\tilde{\mu})\nu_k(\tilde{\mu}) X_{ij,l} X_{kj,l'} = \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,l} X_{32,l'}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

implying:

$$\frac{1}{N^2}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\nu_i(\tilde{\mu})X_{ij}\right)'\left(\sum_{i=1}^{N}\nu_i(\tilde{\mu})X_{ij}\right) = \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12} X_{32}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and $\hat{\mu} - \mu_0$ has the same distribution as

$$\begin{aligned}
&\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \\
&\times \frac{1}{N^2}\left(\sum_{i\neq j} X_{ij}'\left(Y_{ij} - \sum_{l=1}^{K}\mu_{0,l}X_{ij,l}\right) - \sum_{i\neq j, k\neq i,j}\nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk}'\left(Y_{ik} - \sum_{l=1}^{K}\mu_{0,l}X_{ik,l}\right)\right) \\
&= \left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \\
&\times \frac{1}{N^2}\left(\sum_{i\neq j} X_{ij}'(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j}\nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk}'(U_iU_k + V_{ik})\right)
\end{aligned}$$

Now, define:

$$\begin{aligned}
\hat{\mu}^* := \mu_0 + &\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \\
&\times \frac{1}{N^2}\left(\sum_{i\neq j} X_{ij}'(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}'(U_iU_k + V_{ik})\right)
\end{aligned}$$

The proof proceeds in two steps. First, find the asymptotic distribution of $N(\hat{\mu}^* - \mu_0)$. Second, determine the asymptotic distribution of:

$$\hat{\mu}^* - \mu_0 - \left(E(X'_{12}X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \times \frac{1}{N^2}\left(\sum_{i\neq j} X'_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X'_{jk}(U_iU_k + V_{ik})\right)$$

combining the results of both steps allow to conclude.

Step 1: I will begin by assuming that $L = 1$, then generalize to an arbitrary but known L.

Let's determine the asymptotic distribution of $\hat{\mu}^* - \mu_0$, that is, of:

$$\sum_{i\neq j} X_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}(U_iU_k + V_{ik})$$

First, note:

$$\begin{aligned}
\sum_{i\neq j} X_{ij}U_iU_j - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}U_iU_k &= \sum_{i\neq j} X_{ij}U_iU_j - \sum_{i,j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}U_iU_k + \sum_{i,k\neq i} \frac{U_i^3}{||U||^2}X_{ik}U_k \\
&= \sum_{i\neq j} X_{ij}U_iU_j - \sum_{i,j,k\neq j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}U_iU_k \\
&+ \sum_{i,k\neq i} \frac{U_i^3}{||U||^2}X_{ik}U_k + \sum_{i,j} \frac{U_i^3}{||U||^2}X_{ij}U_j \\
&= 2\sum_{i,j} \frac{U_i^3}{||U||^2}X_{ij}U_j \\
&= N\left(2\frac{1}{E(U_1^2)}E(U_1^3X_{12}U_2) + O_p\left(\frac{1}{\sqrt{N}}\right)\right)
\end{aligned} \tag{38}$$

second:

$$\sum_{i\neq j} X_{ij}V_{ij} - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}V_{ik} = \sum_{i\neq j} V_{ij}\left(X_{ij} - \frac{U_i}{||U||^2}\sum_{k\neq i,j} U_k X_{jk}\right)$$

note that

$$\begin{aligned}
&Var\bigg(\sum_{i\neq j} V_{ij}\bigg(X_{ij} - \frac{U_i}{||U||^2}\sum_{k\neq i,j} U_k X_{jk}\bigg)\bigg| U, X\bigg) \\
&\qquad = Var\bigg(\sum_{i<j} V_{ij}\left(2X_{ij} - \frac{U_i}{||U||^2}\sum_{k\neq i,j} U_k X_{jk} - \frac{U_j}{||U||^2}\sum_{k\neq i,j} U_k X_{ik}\right)\bigg| U, X\bigg) \\
&\qquad = \sigma_V^2 \sum_{i<j}\left(2X_{ij} - \frac{U_i}{||U||^2}\sum_{k\neq i,j} U_k X_{jk} - \frac{U_j}{||U||^2}\sum_{k\neq i,j} U_k X_{ik}\right)^2 \\
&\qquad = \sigma_V^2 N^2 \left(2E(X_{12}^2) - 3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}) + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12})^2 + o_p(1)\right) \quad ;\text{ almost sur}
\end{aligned}$$

by a standard CLT:

$$\begin{aligned}
&\frac{1}{N}\left(\sum_{i\neq j} X_{ij}V_{ij} - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}V_{ik}\right) \\
&\rightarrow_d \mathcal{N}\left(0, \sigma_V^2\left(2E(X_{12}^2) - 3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}) + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12})^2\right)\right)
\end{aligned}$$

hence:

$$\begin{aligned}
&\frac{1}{N}\left(\sum_{i\neq j} X_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}(U_iU_k + V_{ik})\right) \\
&\rightarrow_d 2\frac{E(U_1^3)E(U_1)}{E(U_1^2)}E(X_{12}) + \mathcal{N}\left(0, \sigma_V^2\left(2E(X_{12}^2) - 3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}) + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12})^2\right)\right)
\end{aligned}$$

and by the Wold device, for a multivariate X:

$$\frac{1}{N}\left(\sum_{i\neq j} X_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}(U_iU_k + V_{ik})\right)$$

$$\rightarrow_d 2\frac{1}{E(U_1^2)}E(U_1^3U_2X_{12})$$

$$+\mathcal{N}\left(0,\sigma_V^2\left(2E(X_{12}X_{12}')-3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}')+\frac{1}{E(U_1^2)^2}E(U_1X_{12}U_2)E(U_1X_{12}'U_2)\right)\right)$$

therefore

$$N(\hat{\mu}^*-\mu_0)\rightarrow_d \left(E(X_{12}'X_{12})-\frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1}\times$$

$$\left(2\frac{1}{E(U_1^2)}E(U_1^3U_2X_{12})+\mathcal{N}\left(0,\sigma_V^2\left(2E(X_{12}X_{12}')-3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}')+\frac{1}{E(U_1^2)^2}E(U_1X_{12}U_2)E(U_1X_{12}'U_2\right.\right.\right.$$

Step 2:

Again, I will use the Wold device. Let $\eta\in\mathbb{R}^L$ and denote $X_{ij,\eta}=\eta X_{ij}'\in\mathbb{R}$ and $X_\eta:=(X_{ij,\eta})_{ij}\in\mathbb{R}^{N\times N}$.

Let's determine the asymptotic of

$$\begin{aligned}
&\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik})-\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik})\\
&=\nu(\tilde{\mu})'X_\eta M(\mu_0)\nu(\tilde{\mu})-\nu(\tilde{\mu})'diag\left(XM(\mu_0)\right)\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta M(\mu_0)U+\frac{1}{||U||^2}U'diag(X_\eta M(\mu_0))U\\
&=\nu(\tilde{\mu})'X_\eta M(\mu_0)\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta M(\mu_0)U-\left(\nu(\tilde{\mu})'diag\left(X_\eta M(\mu_0)\right)\nu(\tilde{\mu})-\frac{U'}{||U||}diag(X_\eta M)\frac{U'}{||U||}\right)
\end{aligned}\tag{39}$$

- <u>Case 1: $E(U_1)\neq 0$ and U_i and U_j are arbitrarily correlated with X_{ij}</u>
  On one side, note:[9]

$$\begin{aligned}
\left|\nu(\tilde{\mu})'diag\left(X_\eta M(\mu_0)\right)\nu(\tilde{\mu})-\frac{U'}{||U||}diag(X_\eta M)\frac{U'}{||U||}\right|&\leq\left|\left|\nu(\tilde{\mu})-\frac{U}{||U||}\right|\right|\max_k\left|\sum_i X_{ik,\eta}(U_iU_k+V_{ik})\right|\\
&\leq\left|\left|\nu(\tilde{\mu})-\frac{U}{||U||}\right|\right|\max_k\sum_i\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|-E\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|\right.\right.\\
&+N\left|\left|\nu(\tilde{\mu})-\frac{U}{||U||}\right|\right|E\left(\left|X_{ik.\eta}(U_iU_k+V_{ik})\right|\right)
\end{aligned}$$

[9]Remember that, by definition, $X_{ii,\eta}=0$ for all i.

I want to show that:

$$\max_k \sum_i \left(\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| - E\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| \right) = O_p\left(N\sqrt{N} \right)$$

Fix some $x > 0$ and by a union bound:

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{N\sqrt{N}} \max_k \sum_i \left(\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| - E\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| \right) \geq x \right) \\
&\leq \sum_k \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| - E\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| \right) \geq x \right) \\
&= N \times \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left| X_{i1,\eta}(U_i U_1 + V_{i1}) \right| - E\left| X_{i1,\eta}(U_i U_1 + V_{i1}) \right| \right) \geq x \right) \\
&\leq \frac{1}{N^2} \frac{Var\left(\sum_i \left(\left| X_{i1,\eta}(U_i U_1 + V_{i1}) \right| - E\left| X_{i1,\eta}(U_i U_1 + V_{i1}) \right| \right) \right)}{x^2} \\
&\leq \frac{1}{x^2} \left(Var\left(\left| X_{12,\eta}(U_2 U_1 + V_{12}) \right| \right) + Cov\left(\left| X_{12,\eta}(U_2 U_1 + V_{12}) \right|, \left| X_{13,\eta}(U_3 U_1 + V_{13}) \right| \right) \right)
\end{aligned}$$

where the second inequality is Markov's. This implies:

$$\max_k \sum_i \left(\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| - E\left| X_{ik,\eta}(U_i U_k + V_{ik}) \right| \right) = O_p\left(N\sqrt{N} \right)$$

as desired. Since:

$$\left|\left| \nu(\tilde{\mu}) - \frac{U}{||U||} \right|\right| \leq \frac{1}{||U||} \left|\left| v(\tilde{\mu}) - U \right|\right| + ||v(\tilde{\mu})|| \left| \frac{1}{||v(\tilde{\mu})||} - \frac{1}{||U||} \right| = O_p\left(\frac{1}{\sqrt{N}} \right)$$

then:

$$\nu(\tilde{\mu})' diag\left(X_\eta M(\mu_0) \right) \nu(\tilde{\mu}) - \frac{U'}{||U||} diag(X_\eta M(\mu_0)) \frac{U'}{||U||} = O_p(N)$$

and equation (39) becomes:

$$\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik})-\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik})$$
$$=\nu(\tilde{\mu})'X_\eta UU'\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta UU'U$$
$$+\nu(\tilde{\mu})'XV_\eta\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta VU+O_p(\sqrt{N})$$
$$=v(\tilde{\mu})'X_\eta U-U'X_\eta U+\nu(\tilde{\mu})'X_\eta V\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta VU+O_p(\tag{40}$$

On one side:

$$\left|\nu(\tilde{\mu})'X_\eta V\nu(\tilde{\mu})-\frac{1}{||U||^2}U'X_\eta VU\right|\leq\left|\left|\nu(\tilde{\mu})-\frac{U}{||U||}\right|\right|||X_\eta V||=O_p\left(N\right)$$

On the other side:

$$\begin{aligned}v(\tilde{\mu})'X_\eta U-U'X_\eta U&=U'X_\eta(v(\tilde{\mu})-U)\\&=-\frac{1}{\lambda_1(\tilde{\mu})}\sum_{l=1}^{L}(\tilde{\mu}_l-\mu_{0,l})U'X_\eta X_lU+\sum_{l=1}^{L}(\tilde{\mu}_l-\mu_{0,l})\frac{U'X_lU}{U'U}\frac{U'X_\eta U}{\lambda_1(\tilde{\mu})}+O_p(N)\\&=\sum_{l=1}^{L}(\tilde{\mu}_l-\mu_{0,l})\left(\frac{U'X_lU}{U'U}\frac{U'X_\eta U}{\lambda_1(\tilde{\mu})}-\frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta X_lU\right)+O_p(N)\end{aligned}$$

where the second equality is a consequence of equation (37) and from noting that $U'VX_\eta U=O_p(N^2)$ since, almost surely:

$$Var\left(U'X_\eta VU\middle|X,U\right)=\sigma_V^2\sum_{j,k}U_k^2\left(\sum_i X_{ij,\eta}U_i\right)^2=O(N^4)$$

Hence:

$$\frac{\sqrt{N}}{N^2}\left(\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik})-\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik})\right)$$
$$=\sqrt{N}\sum_{l=1}^{L}(\tilde{\mu}_l-\mu_{0,l})\left(\frac{U'X_lU}{NU'U}\frac{U'X_\eta U}{N\lambda_1(\tilde{\mu})}-\frac{N}{\lambda_1(\tilde{\mu})}\frac{1}{N^2}U'X_\eta X_lU\right)+O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= \sqrt{N}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l}) \left(\frac{1}{E(U_1^2)^2} E(U_1 X_{12,l} U_2) E(U_1 U_2 X_{12,\eta}) - \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,l}) \right) + O_p \Bigg($$

the previous equality holds for any fixed η, so:

$$\frac{\sqrt{N}}{N^2}\Bigg(\sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu}) X_{jk}(U_i U_k + V_{ik}) - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}(U_i U_k + V_{ik}) \Bigg)$$

$$= \sqrt{N}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l}) \left(\frac{1}{E(U_1^2)^2} E(U_1 X_{12,l} U_2) E(U_1 U_2 X_{12,\eta}) - \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,l}) \right) + O_p \Bigg($$

$$= \frac{1}{E(U_1^2)}\sqrt{N}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l}) \left(\frac{1}{E(U_1^2)} E(U_1 X_{12,l} U_2) E(U_1 U_2 X_{12,\eta}) - E(U_1 U_3 X_{12,\eta} X_{23,l}) \right) + O_p \Bigg($$

since, by step 1: $N(\hat{\mu}^* - \mu_0) = O_p(1)$, which allows to conclude:

$$\sqrt{N}(\hat{\mu} - \mu_0) = \left(E(X_{12}' X_{12}) - \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12}' X_{32}) \right)^{-1}$$

$$\times \frac{1}{E(U_1^2)}\sqrt{N}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l}) \left(E(U_1 U_3 X_{12,\eta} X_{23,l}) - \frac{1}{E(U_1^2)} E(U_1 X_{12,l} U_2) E(U_1 U_2 X_{12,\eta}) \right)$$

$$= \frac{1}{E(U_1^2)} \left(E(X_{12}' X_{12}) - \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12}' X_{32}) \right)^{-1}$$

$$\times \left(E(U_1 U_3 X_{12} X_{23}) - \frac{1}{E(U_1^2)} E(U_1 X_{12,l} U_2) E(U_1 U_2 X_{12,\eta}) \right) \sqrt{N}(\tilde{\mu} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

- <u>Case 2: $E(U_i) = 0$ and U is exogenous</u>

Let's prove the appropriate version of equation (40) for this case. On one side, note

$$\left| \nu(\tilde{\mu})' diag\left(X_\eta M(\mu_0)\right)\nu(\tilde{\mu}) - \frac{U'}{||U||} diag(X_\eta M)\frac{U'}{||U||} \right| \leq \left|\left| \nu(\tilde{\mu}) - \frac{U}{||U||} \right|\right| \max_k \Big| \sum_i X_{ik,\eta}(U_i U_k + V_{ik}) \Big|$$

$$\leq \left|\left| \nu(\tilde{\mu}) - \frac{U}{||U||} \right|\right| \max_k \Big| \sum_i X_{ik,\eta}(U_i U_k + V_{ik}) - E\left(X_{ik,\eta}(U_i U_k + V_{ik}) | X_k, U_k\right) \Big|$$

$$+ N \left|\left| \nu(\tilde{\mu}) - \frac{U}{||U||} \right|\right| \max_k \Big| E\left(X_{ik,\eta}(U_i U_k + V_{ik}) | X_k, U_k\right) \Big|$$

$$\leq \left|\left| \nu(\tilde{\mu}) - \frac{U}{||U||} \right|\right| \max_k \Big| \sum_i X_{ik,\eta}(U_i U_k + V_{ik}) - E\left(X_{ik,\eta}(U_i U_k + V_{ik}) | X_k, U_k\right) \Big|$$

because under $E(U_1) = 0$, $E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) = 0$. I want to show that:

$$\max_k \Big| \sum_i \left(X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right)\right) \Big| = O_p(N)$$

Fix some $x > 0$ and by a union bound:

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{N}\max_k \Big| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \Big| \geq x\right) \\
&\leq \sum_k \mathbb{P}\left(\frac{1}{N}\Big| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \Big| \geq x\right) \\
&= N \times \mathbb{P}\left(\frac{1}{N}\Big| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \Big| \geq x\right) \\
&\leq \frac{1}{N}\frac{Var\left(\sum_i \left(X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik.\eta}(U_iU_k + V_{ik})\big|X_k, U_k\right)\right)\right)}{x^2} \\
&= \frac{1}{x^2} Var\left(X_{12,\eta}(U_2U_1 + V_{12}) - E\left(X_{12.\eta}(U_1U_2 + V_{12})\big|X_1, U_1\right)\right)
\end{aligned}$$

where the second inequality is Markov's and the last equality results from the fact that for a fixed k, the terms $X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik.\eta}(U_iU_k + V_{ik})\big|X_k, U_k\right)$ are uncorrelated for different i's, because they are centered and independent conditionally on X_k, U_k.

This implies:

$$\max_k \Big| \sum_i \left(X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right)\right) \Big| = O_p(N)$$

as desired. Since:

$$\left|\left|\nu(\tilde{\mu}) - \frac{U}{||U||}\right|\right| = O_p\left(\frac{1}{\sqrt{N}}\right)$$

then:

$$\nu(\tilde{\mu})' diag\left(X_\eta M(\mu_0)\right)\nu(\tilde{\mu}) - \frac{U'}{||U||} diag(X_\eta M(\mu_0))\frac{U'}{||U||} = O_p(\sqrt{N})$$

and equation (39) becomes:

$$\begin{aligned}
\sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik}) \\
&= \nu(\tilde{\mu})'X_\eta M(\mu_0)\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta M(\mu_0)U + O_p(\sqrt{N}) \\
&= \nu(\tilde{\mu})'X_\eta UU'\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta UU'U \\
&\quad + \nu(\tilde{\mu})'XV_\eta\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N}) \\
&= v(\tilde{\mu})'X_\eta U - U'X_\eta U + \nu(\tilde{\mu})'X_\eta V\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N})
\end{aligned} \tag{41}$$

Let's get back to the notation from earlier in step 2: Let $\eta \in \mathbb{R}^L$ and denote $X_{ij,\eta} = \eta X_{ij}' \in \mathbb{R}$ and $X_\eta := (X_{ij,\eta})_{ij} \in \mathbb{R}^{N\times N}$. From equation (41):

$$\begin{aligned}
\sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik}) \\
&= \nu(\tilde{\mu})'X_\eta M(\mu_0)\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta M(\mu_0)U + O_p(\sqrt{N}) \\
&= \nu(\tilde{\mu})'X_\eta M(\tilde{\mu})\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta M(\tilde{\mu})U \\
&\quad + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})\left(\nu(\tilde{\mu})'X_\eta X_l\nu(\tilde{\mu}) - \frac{U'}{||U||}X_\eta X_l\frac{U}{||U||}\right) + O_p(\sqrt{N})
\end{aligned}$$

I next show two useful results: for any random matrix $X \in \mathbb{R}^{N\times N}$ such that X's largest eigenvalue is at most of order N and $X_{ij} := g(X_i, X_j)$ for some fixed function g, then: 1) $||XU|| = O_p(N)$, and 2) $||Xv(\tilde{\mu})|| = ||XU|| + O_p(\sqrt{N})$.

Fix such a random matrix X, the proof of the two results goes as follows:

1. Note that $||XU||^2 = U'X^2U = \sum_{ijk} U_iX_{ij}X_{jk}U_k$, and that:

$$\begin{aligned}
Var\left(\sum_{ijk} U_iX_{ij}X_{jk}U_k \middle| X\right) &= \sum_{i_1,k_1,i_2,k_2} E(U_{i_1}U_{i_2}U_{k_1}U_{k_2}|X)\left(\sum_j X_{i_1j}X_{jk_1}\right)\left(\sum_j X_{i_2j}X_{jk_2}\right) \\
&= N^4(c+o(1));\ \text{alomost surely}
\end{aligned}$$

for some real number c. Then:

$$||XU||^2 = O_p(N^2)$$

as desired.

2. By the equation (37):

$$\lambda_1(\tilde{\mu})||XU - Xv(\tilde{\mu})|| \leq ||XM(\tilde{\mu})r(\tilde{\mu})|| + ||XVU|| + |\lambda_1(\tilde{\mu}) - U'U| \times ||XU|| + \sum_{l=1}^{L} |\tilde{\mu}_l - \mu_{0,l}| \times ||XX_lU|| + E(U_1^2)||XU||$$

let's show that each term in the right hand side is $O_p(N\sqrt{N})$.

(a) $||XM(\tilde{\mu})r(\tilde{\mu})|| \leq \lambda_1(X)||M(\tilde{\mu})r(\tilde{\mu})|| = O_p(N\sqrt{N})$

(b) for the term $||XVU||$, note that:

$$\begin{aligned} U'VX^2VU &= \sum_{i,j,k,l,m} U_iV_{ij}X_{jk}X_{kl}V_{lm}U_l \\ &= \sum_{i,j,k,l,m:\{i,j\}\neq\{l,m\}} U_iV_{ij}X_{jk}X_{kl}V_{lm}U_l \\ &\quad + \sum_{i,j,k} U_iV_{ij}X_{jk}X_{ki}V_{ij}U_j + \sum_{i,j,k} U_iV_{ij}X_{jk}X_{kj}V_{ij}U_i \\ &= \sum_{i,j,k,l,m,\{i,j\}\neq\{l,m\}} U_iV_{ij}X_{jk}X_{kl}V_{lm}U_l + O_p(N^3) \end{aligned}$$

almost surely:

$$Var\left(\sum_{i,j,k,l,m:\{i,j\}\neq\{l,m\}} U_iV_{ij}X_{jk}X_{kl}V_{lm}U_l \middle| X, U\right) = O(N^6)$$

so $||XVU||^2 = U'VX^2VU = O_p(N^3)$, or $||XVU|| = O_p(N\sqrt{N})$

(c) From proposition 9

$$\lambda(\tilde{\mu}) = U'U + \frac{\sum_{l=1}^{L}(\mu_{0,l} - \tilde{\mu}_l)U'X_lU}{U'U} + O_p(1)$$

when $E(U_1)=0$, then $\frac{\sum_{l=1}^{L}(\mu_{0,l}-\tilde{\mu}_l)U'X_lU}{U'U} = O_p(1)$, $|\lambda(\tilde{\mu})-U'U| = O_p(1)$, hence

$$|\lambda_1(\tilde{\mu}) - U'U| \times ||XU|| = O_p(N\sqrt{N})$$

(d) For every $l = 1..L$: $||XX_lU|| \leq \lambda_1(X)||X_lU|| = O_p(N^2)$, since $|\tilde{\mu}_l - \mu_{0,l}| = O_p\left(\frac{1}{\sqrt{N}}\right)$, then $\sum_{l=1}^{L} |\tilde{\mu}_l - \mu_{0,l}| \times ||XX_lU|| = O_p(N\sqrt{N})$

(e) $E(U_1^2)||XU|| = O_p(N)$

In conclusion:

$$||XU|| - ||Xv(\tilde{\mu})|| \leq ||XU - Xv(\tilde{\mu})|| = O_p(\sqrt{N})$$

so equation (40) becomes:

$$\begin{aligned}
\sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k+V_{ik}) - &\sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik}) \\
&= \nu(\tilde{\mu})'X_\eta M(\tilde{\mu})\nu(\tilde{\mu}) - \frac{1}{||U||^2}U'X_\eta M(\tilde{\mu})U + O_p(\sqrt{N}) \\
&= \frac{\lambda_1(\tilde{\mu})}{||v(\tilde{\mu})||^2}v(\tilde{\mu})'X_\eta v(\tilde{\mu}) - \frac{\lambda_1(\tilde{\mu})}{||U||^2}U'X_\eta v(\tilde{\mu}) \\
&- \frac{1}{||U||^2}U'X_\eta M(\tilde{\mu})r(\tilde{\mu}) + O_p(\sqrt{N}) \\
&= -\frac{\lambda_1(\tilde{\mu})}{||v(\tilde{\mu})||^2}v(\tilde{\mu})'X_\eta r(\tilde{\mu}) + \lambda_1(\tilde{\mu})\left(\frac{1}{||v(\tilde{\mu})||^2} - \frac{1}{||U||^2}\right)U'X_\eta v(\tilde{\mu}) \\
&+ O_p(\sqrt{N}) \\
&= -\frac{\lambda_1(\tilde{\mu})}{||v(\tilde{\mu})||^2}v(\tilde{\mu})'X_\eta r(\tilde{\mu}) + O_p(\sqrt{N})
\end{aligned}$$

when $E(U_1)=0$, the equation (37) yields:

$$\begin{aligned}
\lambda_1(M(\tilde{\mu}))v(\tilde{\mu})'X_\eta r(\tilde{\mu}) = v(\tilde{\mu})'X_\eta M(\tilde{\mu})r + v(\tilde{\mu})'X_\eta VU + O_p(1)v(\tilde{\mu})'X_\eta U \\
+ \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})v(\tilde{\mu})'X_\eta X_l U + E(U_1^2)v(\tilde{\mu})'X_\eta U
\end{aligned} \tag{42}$$

I want to show that each of the terms in the right hand side is of order $O_p(N\sqrt{N})$:

1. $|v(\tilde{\mu})'X_\eta M(\tilde{\mu})r| \leq ||v(\tilde{\mu})'X_\eta|| \times ||M(\tilde{\mu})r|| = O_p(N) \times O_p(\sqrt{N})$
2. for $||v(\tilde{\mu})'X_\eta VU||$, first write: $v(\tilde{\mu})'X_\eta VU = U'X_\eta VU - r(\tilde{\mu})'X_\eta VU$ and note that

$|r(\tilde{\mu})'X_\eta VU| \leq ||VU|| \times ||r(\tilde{\mu})'X_\eta|| = O_p(N) \times O_p(\sqrt{N})$, since $||VU|| \leq \lambda_1(V)||U||$ and $||r(\tilde{\mu})'X_\eta|| = O_p(1)$ by the proof in bulletpoint (e) above. So let's examine the term $U'X_\eta VU$:

$$\begin{aligned} Var(U'X_\eta VU|U,X) &= \sigma_V^2 \sum_{jk,\eta} \left(\sum_i U_i X_{ij,\eta} U_k \right)^2 \\ &= \sigma_V^2 \left(\sum_k U_k^2 \right) \sum_j \left(\sum_i U_i X_{ij,\eta} \right)^2 \\ &= \sigma_V^2 ||U||^2 ||X_\eta U||^2 \end{aligned}$$

so

$$\begin{aligned} Var\left(\frac{U'X_\eta VU}{||X_\eta U||} \middle| U,X \right) &= \sigma_V^2 ||U||^2 \\ &= N\sigma_V^2 (E(U_1^2) + o(1)); \quad \text{almost surely} \end{aligned}$$

and

$$\frac{U'X_\eta VU}{||X_\eta U||} = O_p(\sqrt{N})$$

since $||X_\eta U|| = O_p(N)$, then

$$U'X_\eta VU = O_p(\sqrt{N})$$

implying

$$||v(\tilde{\mu})'X_\eta VU|| = O_p(\sqrt{N})$$

3. $|v(\tilde{\mu})'X_\eta U| \leq ||v(\tilde{\mu})|| \times ||X_\eta U|| = O_p(N\sqrt{N})$
4. $|(\tilde{\mu}_l - \mu_{0,l})v(\tilde{\mu})'X_\eta X_l U| \leq |\tilde{\mu}_l - \mu_{0,l}| \times ||v(\tilde{\mu})'X_\eta|| \times ||X_l U|| = O_p(N\sqrt{N})$

this allows to conclude:

$$\lambda_1(M(\tilde{\mu}))v(\tilde{\mu})'X_\eta r(\tilde{\mu}) = O_p(N\sqrt{N})$$

and finally, the equation (40) becomes:

$$\sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) - \sum_{i \neq j, k \neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k + V_{ik}) = O_p(\sqrt{N})$$

so under the condition $E(U_1) = 0$:

$$N(\hat{\mu}^* - \hat{\mu}) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

and

$$N(\hat{\mu} - \mu_0) \to_d E(X'_{12}X_{12})^{-1} \times \mathcal{N}\left(0, 2\sigma_V^2 E(X_{12}X'_{12})\right)$$

$\square$

A.6 Proof of proposition 3

Proof. The case $E(U_1) = 0$ is straightforward, let's prove the proposition for $E(U_1) \neq 0$.

Note that:

$$\begin{aligned} K &= \frac{1}{E(U_1^2)}\left(E(X'_{12}X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X'_{12}X_{32})\right)^{-1}\left(E(U_1U_3X_{12}X_{23}) - \frac{1}{E(U_1^2)}E(U_1X_{12,l}U_2)E(U_1U_2X_1\right. \\ &= B^{-1}A \end{aligned}$$

with:

$$\begin{aligned} A &:= E(U_1^2)E(U_1U_3X_{12}X'_{23}) - E(U_1U_2X_{12})E(U_1U_2X'_{12}) \\ B &:= E(U_1^2)^2E(X_{12}X'_{12}) - E(U_1^2)E(U_1U_3X_{12}X'_{23}) \end{aligned}$$

I begin by showing that $B - A$ is semi-definite positive. Note:

$$\begin{aligned} &E\left(\left(E(U_1^2)X_{12} - U_2X_{13}U_3\right)\left(E(U_1^2)X'_{12} - U_1X'_{24}U_4\right)\right) \\ &= E(U_1^2)^2E(X_{12}X'_{12}) + E(U_1U_2X_{12})E(U_1U_2X'_{12}) - 2E(U_1^2)E(U_1U_3X_{12}X'_{23}) \\ &= B - A \end{aligned}$$

hence[10]

$$\begin{aligned} B - A &= E\left(\left(E(U_1^2)X_{12} - U_2X_{13}U_3\right)\left(E(U_1^2)X'_{12} - U_1X'_{24}U_4\right)\right) \\ &= E\left(E\left[(E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right]E\left[E(U_1^2)X'_{12} - U_1X'_{24}U_4|X_1, X_2, W_{12}, U_1, U_2\right]\right) \\ &= E\left(E\left[(E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right]E\left[(E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right]'\right. \end{aligned}$$

[10]Remember our notation $X_{ij} := \phi(X_i, X_j, W_{ij})$.

$$\geq 0$$

as desired. Moreover, let λ be a deterministic L-dimensional vector such that $\lambda'(B-A)\lambda = 0$, then, almost surely:

$$\lambda' E\left[(E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right] = 0$$

that is:

$$\lambda' E\left[(E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right] = E(U_1^2)\lambda' X_{12} - U_2 E(\lambda' X_{13}U_3|X_1)$$

so:

$$E(U_1^2)\lambda' X_{12} = U_2 E(\lambda' X_{13}U_3|X_1)$$

conditioning on X_2, U_2

$$E(U_1^2)E(\lambda' X_{12}|X_2, U_2) = U_2 E(\lambda' X_{13}U_3)$$

hence

$$E(U_1^2)\lambda' X_{12} = U_1 U_2 \frac{E(\lambda' X_{13}U_3)}{E(U_1^2)}$$

which contradicts our assumption that for any vector $\lambda \in \mathbb{R}^L$, $\mathbb{P}(\lambda' X_{12} = U_1U_2) < 1$. Therefore K has all its eigenvalues < 1.

Note that

$$\begin{aligned}
A &= E\left((U_1U_4X_{12} - U_1U_2X_{14})(U_3U_4X_{23} - U_2U_3X_{34})'\right) \\
&= E\Big(E\left[U_1U_4X_{12} - U_1U_2X_{14}|X_2, U_2, X_4, U_4\right] E\left[U_3U_4X_{23} - U_2U_3X_{34}|X_2, U_2, X_4, U_4\right]'\Big) \\
&= E\Big(E\left[U_1U_4X_{12} - U_1U_2X_{14}|X_2, U_2, X_4, U_4\right] E\left[U_1U_4X_{12} - U_1U_2X_{14}|X_2, U_2, X_4, U_4\right]'\Big) \\
&\geq 0
\end{aligned}$$

Since we have already shown that $B - A > 0$, then $B > 0$, so B is invertible. □

A.7 Proof of corollary 2

Proof. 1. The function $\mu \to |\lambda_1(M(\mu)^2) - \lambda_2(M(\mu)^2)|$ is continuous on the compact $B(\mu_0, \frac{C}{\sqrt{N}})$. Let μ_N be a minimizer on $B(\mu_0, \frac{C}{\sqrt{N}})$. We show in the proof of proposition 9 that $\lambda_1(M(\mu_N)^2) = O_p(N^2)$ and $\lambda_2(M(\mu_N)^2) = O_p(N)$. So $|\lambda_1(M(\mu_N)^2) -$

$\lambda_2(M(\mu_N)^2)| = O_p(N^2)$. So with probability approaching 1, the largest eigenvalue of $M(\mu_N)$ in absolute value is simple on all of $B(\mu_0, \frac{C}{\sqrt{N}})$. The compactness of $B(\mu_0, \frac{C}{\sqrt{N}})$along with theorem 1 in Magnus (1985) allows to conclude that $\mu \to \nu(\mu)$ is infinitely continuously differentiable on $B(\mu_0, \frac{C}{\sqrt{N}})$.

2. Following 1), assume f_N is continuously differentiable on $B(\mu_0, \frac{2C}{\sqrt{N}})$. Let $\mu_{\max}$ be a maximizer of $||f'_N(\mu)||$ on $B(\mu_0, \frac{C}{\sqrt{N}})$. By equation (13): $\sqrt{N}(f_N(\mu_{\max}) - \mu_0) = K\sqrt{N}(\mu_{\max} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right)$
also

$$\sqrt{N}(f_N(\mu_{\max} + \frac{1}{\sqrt{N}}) - \mu_0) = K\sqrt{N}(\mu_{\max} + \frac{1}{\sqrt{N}} - \mu_0) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

taking the difference of the two last equations:

$$\sqrt{N}\left(f_N\left(\mu_{\max} + \frac{1}{\sqrt{N}}\right) - f_N(\mu_{\max})\right) = K + O_p\left(\frac{1}{\sqrt{N}}\right)$$

on the other side, by a Taylor expansion:[11]

$$\sqrt{N}\left(f_N\left(\mu_{\max} + \frac{1}{\sqrt{N}}\right) - f_N(\mu_{\max})\right) = f'_N(\mu_{\max}) + o_p\left(1\right)$$

hence

$$f'_N(\mu_{\max}) - K = o_p(1)$$

with a probability approaching 1:

$$||f'(\mu_{\max})|| = \sup_{\mu \in B(\mu_0, \frac{C}{\sqrt{N}})} ||f'(\mu)|| \leq \kappa$$

for any $\kappa \in (\lambda_1(K), 1)$.

3. Fix some $\kappa \in (\lambda_1(K), 1)$ and $\epsilon > 0$. There exists $M > 0$ such that for N large enough, with probability at least $1 - \epsilon$, $\hat{\mu}_1, \hat{\mu}_0 \in B(\mu_0, \frac{M}{2\sqrt{N}})$ so that $||\hat{\mu}_1 - \hat{\mu}_0|| \leq \frac{M}{\sqrt{N}}$ (let this be event E_N) . Assume f_N is continuously differentiable on $B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$ (denote this event F_N) and that $\sup_{\mu \in B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)} ||f'(\mu)|| \leq \kappa$ (let this be event

[11]Building on the results in Magnus (1985), and after some tedious computations, we can show that $\sup_{\mu \in B(\mu_0, \frac{C}{\sqrt{N}})} ||\frac{\partial \nu(\mu)}{\partial \mu_l}|| = O_p(1)$ and $\sup_{\mu \in B(\mu_0, \frac{C}{\sqrt{N}})} ||\frac{\partial^2 \nu(\mu)}{\partial \mu_l \partial \mu_q}|| = O_p(1)$, implying $\sup_{\mu \in B(\mu_0, \frac{C}{\sqrt{N}})} ||f''_N(\mu)|| = O_p(1)$.

G_N). Then for any $\mu, \mu' \in B(\mu_0, \frac{M}{\sqrt{N}}(1+\frac{1}{1-\kappa}))$, we have:

$$||f_N(\mu) - f_N(\mu')|| \leq \kappa||\mu - \mu'||$$

By induction on m, assume $\hat{\mu}_0, ..., \hat{\mu}_m \in B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$

$$||\hat{\mu}_{m+1} - \hat{\mu}_m|| = ||f_N(\hat{\mu}_m) - f_N(\hat{\mu}_{m-1})|| \leq \kappa||\hat{\mu}_m - \hat{\mu}_{m-1}||$$

and

$$\begin{aligned}
||\hat{\mu}_{m+1} - \mu_0|| &\leq ||\hat{\mu}_{m+1} - \hat{\mu}_m|| + ||\hat{\mu}_m - \hat{\mu}_{m-1}|| + ... + ||\hat{\mu}_1 - \hat{\mu}_0|| + ||\hat{\mu}_0 - \mu_0|| \\
&\leq \sum_{i=0}^{m} \kappa^i ||\hat{\mu}_1 - \hat{\mu}_0|| + ||\hat{\mu}_0 - \mu_0|| \\
&\leq \frac{M}{\sqrt{N}}(1 + \sum_{i=0}^{m} \kappa^i) \\
&\leq \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)
\end{aligned}$$

so $\hat{\mu}_{m+1} \in B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$.

So even though f_N is not necessarily contracting on $B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$, because it may not preserve $B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$, we can follow the proof of the Banach fixed point theorem for the specific sequence $\hat{\mu}$ (or in fact any sequence initiated in a way that the first two first elements are in $B\left(\mu_0, \frac{M}{2\sqrt{N}}\right)$ and not just $B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$). First we show that the sequence $\hat{\mu}_m$ is a Cauchy sequence, let $p, q \in \mathbb{N}$, without loss of generality take $p > q$

$$||\hat{\mu}_p - \hat{\mu}_q|| \leq \frac{\kappa^q}{1-\kappa}||\hat{\mu}_1 - \hat{\mu}_0|| \to 0, \text{ as } q \to +\infty$$

so the sequence is a Cauchy sequence. Therefore, it has a limit in $B\left(\mu_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$ that can only be a fixed point of f_N. By lemma 1, the sequence converges to a minimizer. We have shown the following:

$$E_N, F_N \text{ and } G_N \Rightarrow \text{The sequence converges to a minimizer}$$

Which proves that with probability approching 1, the the sequence converges to a minimizer as desired.

Finally, the last result along with lemma 1 ensure that $\hat{\mu}^*$ is a solution to the minimization problem 7 and is $\sqrt{N}$-consistent. Equation (13) yields

$$(I-K)\sqrt{N}(\hat{\mu}^* - \mu_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

and finally

$$(\hat{\mu}^* - \mu_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

□

A.8 Proof of proposition 4

Proof. As for the proof of propositions 1 (section A.5), assume that $\delta = 1$. The result for an unknown $\delta \in \{-1, 1\}$ immediately follows as described in the the proof section A.5 .

Again, I use the Wold device. Let $\eta \in \mathbb{R}^L$ and denote $X_{ij,\eta} = \eta X'_{ij} \in \mathbb{R}$ and $X_\eta := (X_{ij,\eta})_{ij} \in \mathbb{R}^{N\times N}$.

Following equation (40):

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= \sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}U_iU_k \\
&+ \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}U_iU_k - \sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= \sum_{i\neq j} X_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}(U_iU_k + V_{ik}) \\
&+ U'X_\eta(U - v(\tilde{\mu})) - \nu(\tilde{\mu})'X_\eta V \nu(\tilde{\mu}) + \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N}) \\
&= N\left(\frac{2}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)\right) + \sum_{i\neq j} X_{ij}V_{ij} - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} \\
&+ U'X_\eta r(\tilde{\mu}) + \frac{1}{||v(\tilde{\mu}))||^2}\left(U'X_\eta VU - v(\tilde{\mu})'X_\eta V v(\tilde{\mu})\right) + U'X_\eta VU\frac{||v(\tilde{\mu})||^2 - ||U||^2}{||v(\tilde{\mu})||^2||U||^2} + O_p(\sqrt{N})
\end{aligned}$$

where the second equality results from equation (40) and the third from equation (38).

Note that

1.
$$U'X_\eta VU - v(\tilde{\mu})'X_\eta Vv(\tilde{\mu}) = -\frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta V^2U + O_p(N\sqrt{N})$$

to see that, observe that from equation (37):

$$\begin{aligned}
Vv(\tilde{\mu}) = VU - \frac{1}{\lambda_1(\tilde{\mu})}\Bigg(&VM(\tilde{\mu})r(\tilde{\mu}) - V^2U + O_p(1)VU + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})VX_lU + E(U_1^2)VU \\
&- \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}VU\Bigg) \\
X_\eta v(\tilde{\mu}) = X_\eta U - \frac{1}{\lambda_1(\tilde{\mu})}\Bigg(&X_\eta M(\tilde{\mu})r(\tilde{\mu}) - X_\eta VU + O_p(1)X_\eta U + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})X_\eta X_lU + E(U_1^2)X_\eta U \\
&- \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}X_\eta U\Bigg)
\end{aligned}$$

combining both identities:

$$\begin{aligned}
v(\tilde{\mu})'X_\eta Vv(\tilde{\mu}) = {} & U'X_\eta VU - \frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta\Bigg(VM(\tilde{\mu})r(\tilde{\mu}) - V^2U + O_p(1)VU \\
& + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})VX_lU + E(U_1^2)VU - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}VU\Bigg) \\
& - \frac{1}{\lambda_1(\tilde{\mu})}U'V\Bigg(X_\eta M(\tilde{\mu})r(\tilde{\mu}) - X_\eta VU + O_p(1)X_\eta U \\
& + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})X_\eta X_lU + E(U_1^2)X_\eta U - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}X_\eta U\Bigg) \\
& + \frac{1}{\lambda_1^2(\tilde{\mu})}\Bigg(X_\eta M(\tilde{\mu})r(\tilde{\mu}) - X_\eta VU + O_p(1)X_\eta U \\
& + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})X_\eta X_lU + E(U_1^2)X_\eta U - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}X_\eta U\Bigg)' \\
& \times \Bigg(VM(\tilde{\mu})r(\tilde{\mu}) - V^2U + O_p(1)VU + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})VX_lU + E(U_1^2)VU \\
& - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kU}{U'U}VU\Bigg) \\
= {} & U'X_\eta VU + \frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta V^2U + O_p(N\sqrt{N})
\end{aligned}$$

2. Remark that $Var(U'X_\eta VU|X,U) = \sigma_V^2 \sum_{jk}\left(\sum_i U_iX_{ij}U_k\right)^2 = O(N^4)$ almost surely,

hence

$$U'X_\eta VU = O_p(N^2)$$

so:

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= N\left(\frac{2}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)\right) + \sum_{i\neq j} X_{ij,\eta}V_{ij} - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} \\
&+ U'X_\eta r(\tilde{\mu}) - \frac{1}{||v(\tilde{\mu})||^2\lambda_1(\tilde{\mu})}U'X_\eta V^2U + O_p(\sqrt{N})
\end{aligned}$$

Let's determine the asymptotic distribution of $U'X_\eta r(\tilde{\mu})$. From the equation (37)

$$\lambda_1(M(\tilde{\mu}))U'X_\eta r(\tilde{\mu}) = U'X_\eta M(\tilde{\mu})r - U'X_\eta VU + (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta U + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_lU + E(U_1^2)U'X$$

Also

$$\begin{aligned}
\lambda_1(M(\tilde{\mu}))U'X_\eta M(\tilde{\mu})r(\tilde{\mu}) &= U'X_\eta M(\tilde{\mu})^2r - U'X_\eta M(\tilde{\mu})VU + (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta M(\tilde{\mu})U \\
&+ \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta M(\tilde{\mu})X_lU + E(U_1^2)U'X_\eta M(\tilde{\mu})U \\
&= -U'X_\eta UU'VU - U'XV^2U - \sum_{l=1}^{L}(\mu_{0,l} - \mu_l)U'X_\eta X_lVU + (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta UU' \\
&+ (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta VU + (\lambda_1(M(\tilde{\mu})) - U'U)\sum_{l=1}^{L}(\mu_0 - \tilde{\mu})U'X_\eta X_lU \\
&+ \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta UU'X_lU + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta VX_lU \\
&+ \sum_{k=1}^{L}(\mu_{0,k} - \tilde{\mu}_k)\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_kX_lU + E(U_1^2)U'X_\eta UU'U + E(U_1^2)U'X_\eta VU \\
&+ E(U_1^2)\sum_{l=1}^{L}(\mu_{0,l} - \tilde{\mu}_l)U'X_\eta X_lU + O_p(N^2\sqrt{N}) \\
&= -U'X_\eta UU'VU - U'X_\eta V^2U + (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta UU'U \\
&+ \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta UU'X_lU + E(U_1^2)U'X_\eta UU'U + O_p(N^2\sqrt{N})
\end{aligned}$$

By equation (36), when $\tilde{\mu} - \mu_0 = O_p\left(\frac{1}{N}\right)$ as we are assuming here, we get:

$$\lambda_1(\tilde{\mu}) = U'U + \frac{\sum_k(\mu_{0,k} - \tilde{\mu}_k)U'X_kU}{U'U} + \frac{U'VU}{U'U} - E(U_1^2) + \frac{U'V^2U}{(U'U)^2} + O_p\left(\frac{1}{\sqrt{N}}\right)$$

therefore:

$$\lambda_1(M(\tilde{\mu}))U'X_\eta M(\tilde{\mu})r(\tilde{\mu}) = -U'X_\eta V^2U + \frac{U'V^2UU'X_\eta U}{U'U} + O_p(N^2\sqrt{N})$$

plugging back in the expansion of $U'X_\eta r(\tilde{\mu})$:

$$\begin{aligned}
\lambda_1(M(\tilde{\mu}))U'X_\eta r(\tilde{\mu}) &= -\frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta V^2U + \frac{U'V^2UU'X_\eta U}{\lambda_1(\tilde{\mu})U'U} - U'X_\eta VU \\
&\quad + (\lambda_1(M(\tilde{\mu})) - U'U)U'X_\eta U + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_lU + E(U_1^2)U'X_\eta U + O_p(N\sqrt{N}) \\
&= \frac{1}{U'U}U'X_\eta UU'VU - \frac{1}{\lambda_1(\tilde{\mu})}U'X_\eta V^2U + \frac{U'V^2UU'X_\eta U}{\lambda_1(\tilde{\mu})U'U} - U'X_\eta VU \\
&\quad + \frac{U'V^2UU'X_\eta U}{(U'U)^2} - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kUU'X_\eta U}{U'U} + \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_lU + O_p(N\sqrt{N})
\end{aligned}$$

so

$$\begin{aligned}
U'X_\eta r(\tilde{\mu}) &= \frac{1}{(U'U)^2}U'X_\eta UU'VU - \frac{1}{(U'U)^2}U'X_\eta V^2U + 2\frac{U'V^2UU'X_\eta U}{(U'U)^3} - \frac{1}{U'U}U'X_\eta VU \\
&\quad - \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kUU'X_\eta U}{(U'U)^2} + \frac{1}{U'U}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_lU + O_p(\sqrt{N})
\end{aligned}$$

plugging:

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= N\left(\frac{2}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)\right) + \sum_{i\neq j} X_{ij,\eta}V_{ij} - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} \\
&+ \frac{1}{(U'U)^2}U'X_\eta UU'VU - \frac{1}{(U'U)^2}U'X_\eta V^2U + 2\frac{U'V^2UU'X_\eta U}{(U'U)^3} - \frac{1}{U'U}U'X_\eta VU \\
&- \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kUU'X_\eta U}{(U'U)^2} + \frac{1}{U'U}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_lU - \frac{1}{||v(\tilde{\mu})||^2\lambda_1(\tilde{\mu})}U'X_\eta V^2U + O_p(\sqrt{N})
\end{aligned}$$

Notice that

$$U'V^2U = \sum_{i,j} U_i^2 V_{ij}^2 + O_p(N\sqrt{N})$$

and

$$U'XV^2U = \sum_{i,j,k} U_i X_{ij} V_{jk}^2 U_j + O_p(N^2\sqrt{N})$$

so that

$$\frac{1}{||v(\tilde{\mu})||^2\lambda_1(\tilde{\mu})} U'X_\eta V^2 U - \frac{1}{(U'U)^2} U'X_\eta V^2 U = O_p\left(\frac{1}{\sqrt{N}}\right)$$

subsequently

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= N\left(\frac{2}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)\right) + \sum_{i\neq j} X_{ij,\eta}V_{ij} - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} \\
&+ \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU - \frac{1}{U'U}U'X_\eta VU \\
&- \frac{\sum_k(\tilde{\mu}_k - \mu_{0,k})U'X_kUU'X_\eta U}{(U'U)^2} + \frac{1}{U'U}\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})U'X_\eta X_l U + O_p(\sqrt{N}) \\
&= \sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})\left(\frac{U'X_\eta X_lU}{U'U} - \frac{U'X_lU}{U'U}\frac{U'X_\eta U}{U'U}\right) \\
&+ \frac{2N}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) \\
&+ \sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU + O_p(\sqrt{N}) \\
&= N^2\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})\left(\frac{1}{E(U_1^2)}E(U_1U_3X_{12,\eta}X_{23,l}) - \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,l})E(U_1U_2X_{12,\eta})\right) \\
&+ \frac{2N}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) \\
&+ \sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU + O_p(\sqrt{N}) \\
&= N^2\sum_{l=1}^{L}(\tilde{\mu}_l - \mu_{0,l})\left(\frac{1}{E(U_1^2)}E(U_1U_3X_{12,\eta}X_{23,l}) - \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,l})E(U_1U_2X_{12,\eta})\right) + R_{N,\eta} + O_p(\sqrt{N})
\end{aligned}$$

where the residual $R_{N,\eta}$ is of order $O_p(N)$ and is given by:

$$R_{N,\eta} := \frac{2N}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)$$

$$+\sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\frac{1}{NE(U_1^2)}\sum_{i\neq j,k\neq i,j} U_iU_jX_{jk,\eta}V_{ik} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU + O_p(\sqrt{N})$$

$$= \frac{2N}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)$$

$$+\sum_{ij} V_{ij}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right) + O_p(\sqrt{N})$$

we get:

$$\begin{aligned}
Var(R_{N,\eta}|X,U) &= \sigma_V^2\sum_{i<j}\Bigg(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j \\
&\quad + X_{ij,\eta} - 2U_j\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{ik,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\Bigg)^2 \\
&= \sigma_V^2\sum_{i<j}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right)^2 \\
&+ \sigma_V^2\sum_{i<j}\left(X_{ij,\eta} - 2U_j\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{ik,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right)^2 \\
&+ 2\sigma_V^2\sum_{i<j}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right) \\
&\quad \times\left(X_{ij,\eta} - 2U_j\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{ik,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right) \\
&= \sigma_V^2\sum_{i\neq j}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right)^2 \\
&+ 2\sigma_V^2\sum_{i<j}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right) \\
&\quad \times\left(X_{ij,\eta} - 2U_j\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{ik,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right) \\
&= \sigma_V^2\sum_{i\neq j}X_{ij,\eta}^2 + 4U_i^2\frac{1}{N^2E(U_1^2)^2}\left(\sum_{k\neq i,j}U_kX_{jk,\eta}\right)^2 + \frac{1}{E(U_1^2)^4}E(U_1U_2X_{12,\eta})^2U_i^2U_j^2 \\
&- \sigma_V^2\frac{4}{NE(U_1^2)}\sum_{i\neq j,k\neq i,j}X_{ij,\eta}U_iU_kX_{jk,\eta} + 2\frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})\sum_{i\neq j}X_{ij,\eta}U_iU_j \\
&- 4\sigma_V^2\frac{1}{NE(U_1^2)^3}E(U_1U_2X_{12,\eta})\sum_{i\neq j,k\neq i,j}U_i^2U_jU_kX_{jk,\eta}
\end{aligned}$$

$$+ \sigma_V^2 \sum_{i\neq j} \Bigg(X_{ij,\eta}^2 - 2U_j \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{ik,\eta} X_{ij,\eta} + \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta}) U_i U_j X_{ij,\eta} \Bigg)$$

$$+ \sigma_V^2 \sum_{i\neq j} \Bigg(- 2U_i \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{jk,\eta} X_{ij,\eta} + 4U_i U_j \frac{1}{N^2 E(U_1^2)^2} \sum_{k,l\neq i,j} U_k U_l X_{ik,\eta} X_{jl,\eta}$$

$$- 2U_i \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{jk,\eta} \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta}) U_i U_j \Bigg)$$

$$+ \sigma_V^2 \sum_{i\neq j} \Bigg(+ \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta}) U_i U_j X_{ij,\eta} - 2 \frac{1}{NE(U_1^2)^3} E(U_1 U_2 X_{12,\eta}) \sum_{k\neq i,j} U_i U_j^2 U_k X_{ik,\eta}$$

$$+ \frac{1}{E(U_1)^4} E(U_1 U_2 X_{12,\eta})^2 U_i^2 U_j^2 \Bigg)$$

$$= N^2 \sigma_V^2 \Bigg(E(X_{12,\eta}^2) + 4 \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,\eta}) + \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$- \frac{4}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,\eta}) + \frac{2}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2 - \frac{4}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$+ E(X_{12,\eta}^2) - 2 \frac{1}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,\eta}) + \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$- \frac{2}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,\eta}) + \frac{4}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$- \frac{2}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$+ \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2 - 2 \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2$$

$$+ \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2 + o(1) \Bigg) \text{ almost surely.}$$

$$= N^2 \sigma_V^2 \Bigg(2E(X_{12,\eta}^2) + \frac{2}{E(U_1^2)^2} E(U_1 U_2 X_{12,\eta})^2 - \frac{4}{E(U_1^2)} E(U_1 U_3 X_{12,\eta} X_{23,\eta}) + o(1) \Bigg) \text{ almost}$$

clearly, the Lyapunov condition is met and by the Lyapunov CLT

$$\frac{1}{N} R_{N,\eta} \to_d \mathcal{N}\left(0, \sigma_V^2 \eta \Sigma \eta'\right)$$

for

$$\Sigma := 2 \Bigg(E(X_{12} X_{12}') + \frac{1}{E(U_1^2)^2} E(U_1 U_2 X_{12}) E(U_1 U_2 X_{12})' - \frac{2}{E(U_1^2)} E(U_1 U_3 X_{12} X_{23}')) \Bigg)$$

Finally:

$$
\begin{aligned}
N(\hat{\mu}-\mu_0) &= \left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1} \\
&\times \frac{1}{N}\left(\sum_{i\neq j} X_{ij}'(U_iU_j+V_{ij}) - \sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\mu})\nu_j(\tilde{\mu})X_{jk}'(U_iU_k+V_{ik})\right) \\
&= \frac{1}{E(U_1^2)}\left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1} \\
&\times \left(E(U_1U_3X_{12}X_{23}') - \frac{1}{E(U_1^2)}E(U_1U_2X_{12})E(U_1U_2X_{12}')\right)N(\tilde{\mu}-\mu_0) \\
&+ R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= KN(\tilde{\mu}-\mu_0) + R_N + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

with

$$
K := \frac{1}{E(U_1^2)}\left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}')\right)^{-1}\left(E(U_1U_3X_{12}X_{23}') - \frac{1}{E(U_1^2)}E(U_1U_2X_{12})E(U_1U_2X
$$

$$
\begin{aligned}
R_N \to_d \frac{2}{E(U_1^2)}\left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32}')\right)^{-1}E(U_1^3U_2X_{12}) \\
+ \left(E(X_{12}X_{12}') - \frac{E(U_1)^2}{E(U_1^2)}E(X_{12}X_{32}')\right)^{-1}\mathcal{N}\left(0, \sigma_V^2\Sigma\right)
\end{aligned}
$$

□

A.9 Proof of proposition 5

Proof. Write:

$$
\begin{aligned}
K &= \frac{1}{E(U_1^2)}\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1}\left(E(U_1U_3X_{12}X_{23}) - \frac{1}{E(U_1^2)}E(U_1X_{12}U_2)E(U_1U_2X_{12}
\\
&=: F\left(E(U_1), E(X_{12}X_{12}'), E(U_1U_3X_{12}X_{23}'), E(U_1U_2X_{12})\right)
\end{aligned}
$$

for a function F that is continuously differentiable at $x := \left(E(U_1), E(X_{12}X_{12}'), E(U_1U_3X_{12}X_{23}'), E(U_1U_2X_{12})\right)$
For any estimator x_N of x:

$$
|F(x_N) - F(x)| \leq ||x_N - x|| \times \left|\left|\frac{\partial F}{\partial x}(\bar{x})\right|\right|
$$

where $||.||$ is the Euclidean norm and where $\bar{x}$ is a convex combination of x_N and x. So $||x_N - x|| = O_p\left(\frac{1}{\sqrt{N}}\right)$ implies $|F(x_N) - F(x)| = O_p\left(\frac{1}{\sqrt{N}}\right)$. Therefore, it is enough to propose $\sqrt{N}$ consistent estimators for each of the elements $E(U_1)$, $E(X_{12}X'_{12})$, $E(U_1U_3X_{12}X'_{23})$ and $E(U_1U_2X_{12})$.

Clearly, by the standard CLT: $\frac{\sum_{i=1 \leq N/2} X_{i,j}X'_{i,j}}{N(N-1)}$, is $\sqrt{N}$- consistent for $E(X_{12}X'_{12})$.

For the parameter $E(U_1^2)$, lemma 3 shows that the estimators $\frac{\sum_i \hat{U}_i}{N}$ (Cf. lemma 3 for the definitions) is enough for our purposes.

For any $l, q \in 1...L$:

$$\begin{aligned}
v(\tilde{\mu})'X_lX_qv(\tilde{\mu}) &= U'X_lX_qU + O_p(N^2\sqrt{N}) \\
&= N^3\left(E(U_1U_3X_{12,l}X_{23,q}) + O_p\left(\frac{1}{\sqrt{N}}\right)\right) \\
v(\tilde{\mu})'X_lv(\tilde{\mu}) &= U'X_lX_qU + O_p(N\sqrt{N}) \\
&= N^2\left(E(U_1U_2X_{12,l}) + O_p\left(\frac{1}{\sqrt{N}}\right)\right)
\end{aligned}$$

Plugging the five estimators in the function F yields the desired estimator:

$$\begin{aligned}
\hat{K}_N :=& \left(\frac{\sum_{i\neq j} X_{i,j}X'_{i,j}}{N^2} - \frac{\left[\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))\right]_{l,q}}{N^2}\right)^{-1} \\
&\times \left(\frac{\left[\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))\right]_{l,q}}{N^2} - \frac{\left[\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))\right]'_l}{N}\frac{\left[\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))\right]'_l}{N}\right) \\
&= \left(\sum_{i\neq j} X_{i,j}X'_{i,j} - \left[\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))\right]_{l,q}\right)^{-1}\left(\left[\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))\right]_{l,q} - \left[\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))\right]_l\left[\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))\right]'_l\right)
\end{aligned}$$

Where $\left[\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))\right]'_l$ and $\left[\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))\right]_{l,q}$ are the matrices of dimensions $L \times 1$ and $L \times L$, with entries $\nu(\tilde{\mu}'X_l\nu(\tilde{\mu}))$ and $\nu(\tilde{\mu})'X_lX_q\nu(\tilde{\mu}))$ respectively. □

A.10 Proof of lemma 3

Proof. When $\delta = 1$, by proposition 9, with probability approaching 1 $\hat{\delta} = \frac{\lambda_1(\tilde{\mu})}{|\lambda_1(\tilde{\mu})|}$ and $\frac{\lambda_1(\tilde{\mu})}{|\lambda_1(\tilde{\mu})|} = 1$.

When $\delta = -1$,

$$\hat{\delta} := \frac{-\lambda_N(-M(\tilde{\mu}))\mathbb{1}\{-\lambda_N(-M(\tilde{\mu})) > \lambda_1(-M(\tilde{\mu}))\} - \lambda_1(-M(\tilde{\mu}))\mathbb{1}\{-\lambda_N(-M(\tilde{\mu})) < \lambda_1(-M(\tilde{\mu}))\}}{\max_i |\lambda_i(\tilde{\mu})|}$$

so with probability approaching 1:

$$\hat{\delta} := -\frac{\lambda_N(-M(\tilde{\mu}))\mathbb{1}\{|\lambda_N(-M(\tilde{\mu}))| > \lambda_1(-M(\tilde{\mu})))\} + \lambda_1(-M(\tilde{\mu}))\mathbb{1}\{|\lambda_N(-M(\tilde{\mu}))| < \lambda_1(-M(\tilde{\mu}))\}}{\max_i |\lambda_i(\tilde{\mu})|}$$

by the same reasonning as for the case $\delta = 1$, $\frac{\lambda_N(-M(\tilde{\mu}))\mathbb{1}\{|\lambda_N(-M(\tilde{\mu}))| > \lambda_1(-M(\tilde{\mu})))\} + \lambda_1(-M(\tilde{\mu}))\mathbb{1}\{|\lambda_N(-M(\tilde{\mu}))| < \lambda_1(-M(\tilde{\mu}}{\max_i |\lambda_i(\tilde{\mu})|}$
1 with probability approaching 1. so $\hat{\delta} = -1$ with probability approaching 1.

I begin by proving the third approximation. Note that when $\delta = 1$:

$$\begin{aligned}
||U - \sqrt{\lambda_1(\tilde{\mu})}\nu(\tilde{\mu})||_2 &\leq ||U - v(\tilde{\mu})||_2 + ||v(\tilde{\mu}) - \sqrt{\lambda_1(\tilde{\mu})}\nu_i(\tilde{\mu})||_2 \\
&= ||U - v(\tilde{\mu})||_2 + |\sqrt{\lambda_1(\tilde{\mu})} - ||v(\tilde{\mu})||_2| \\
&= ||U - v(\tilde{\mu})||_2 + \frac{\lambda_1(\tilde{\mu}) - ||v(\tilde{\mu})||_2^2}{\sqrt{\lambda_1(\tilde{\mu})} + ||v(\tilde{\mu})||_2} \\
&= ||U - v(\tilde{\mu})||_2 + \frac{\lambda_1(\tilde{\mu}) - ||U||_2^2 + ||U||_2^2 - ||v(\tilde{\mu})||_2^2}{\sqrt{\lambda_1(\tilde{\mu})} + ||v(\tilde{\mu})||_2}
\end{aligned}$$

By proposition 9, $||U - v(\tilde{\mu})||_2 = O_p(1)$, $\lambda_1(\tilde{\mu}) - ||U||_2^2 = O_p(\sqrt{N})$ and $||U||_2^2 - ||v(\tilde{\mu})||_2^2 = (||U||_2 - ||v(\tilde{\mu})||_2)(||U||_2 + ||v(\tilde{\mu})||_2) = O_p(\sqrt{N})$, therefore $||U - \sqrt{\lambda_1(\tilde{\mu})}\nu(\tilde{\mu})||_2 = O_p(1)$. Likewise, when $\delta = -1$, the same reasonning applies to the matrix $-M(\tilde{\mu})$ and we get that $||U + \sqrt{|\lambda_N(\tilde{\mu})|}\nu_(\tilde{\mu})||_2 = O_p(1)$

Combining both cases, we establish that $||U - \delta\sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu})||_2 = O_p(1)$. Hence:

$$\begin{aligned}
||U - \hat{U}|| &\leq ||U - \delta\sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu})||_2 + ||\delta\sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu}) - \hat{\delta}\sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu})||_2 \\
&= ||U - \delta\sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu})||_2 + |\delta - \hat{\delta}| \times \sqrt{\max_i |\lambda_i(\tilde{\mu})|}\nu(\tilde{\mu})||_2 \\
&= O_p(1)
\end{aligned}$$

Fix $\eta \in \mathbb{R}^L$ and $X_{ij,\eta} = \eta' X_{ij} \in \mathbb{R}$:

$$\begin{aligned}
\left|\left|\frac{1}{N^2}\sum_{i\neq j,}\hat{U}_i\hat{U}_jX_{ij,\eta} - \frac{1}{N^2}\sum_{i\neq j}U_iU_jX_{ij,\eta}\right|\right| &= \left|\left|\frac{1}{N^2}\left(\sum_{i\neq j}(\hat{U}_i - U_i)\hat{U}_jX_{ij,\eta} + \sum_{i\neq j}U_i(\hat{U}_j - U_j)X_{ij,\eta}\right)\right|\right| \leq \frac{1}{N^2}\Bigg(\\
&= O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

The two remaining results are proved similarly. □

A.11 Proof of lemma 4:

Proof. First:

$$
\begin{aligned}
e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m} &= \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \iota\iota' - \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 \iota\iota \\
&\quad + \left(b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right) - b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)\right)(\iota A' + A\iota') \\
&= c_{m,N}\sigma_A\iota\iota' + \iota A' + A\iota'
\end{aligned}
$$

as desired. Second:

$$
\begin{aligned}
e'_{1,m}e_{2,m} &= \left(\frac{1}{4}b^2_{m,N}c^2_{m,N}\sigma^2_A - \frac{1}{4b^2_{m,N}}\right)N + b^2_{m,N}||A||^2 + b^2_{m,N}c_{m,N}\sigma_A\iota' A \\
&= \frac{1}{4b^2_{m,N}}\left(b^4_{m,N}\left(4||A||^2 + Nc^2_{m,N}\sigma^2_A + 4c_{m,N}\sigma_A\iota' A\right) - N\right) \\
&= 0
\end{aligned}
$$

□

A.12 Proof of proposition 6

In line with the proofs leading up to theorems 1 and 2, we beging by studying the behavior of the $M(\hat{\mu}_m)$'s largest eigenvalue ($\hat{\mu}_m$ defined in equation 21). First, decompose:

$$e_{1,m} = v_{11}(\hat{\mu}_m) + v_{12}(\hat{\mu}_m) + r_1(\hat{\mu}_m)$$

and likewise

$$e_{2,m} = v_{21}(\hat{\mu}_m) + v_{22}(\hat{\mu}_m) + r_2(\hat{\mu}_m)$$

where $v_{11}(\hat{\mu}_m)$ and $v_{12}(\hat{\mu}_m)$ are orthogonal projections of $e_{1,m}$ on $M(\hat{\mu}_m)'s$ eigen-spaces corresponding to $\lambda_1(\hat{\mu}_m)$ and $\lambda_N(\hat{\mu}_m)$ respectively. $v_{21}(\hat{\mu}_m)$ and $v_{22}(\hat{\mu}_m)$ are defined similarly. We have:

Lemma 5.

- $\lambda_1(\hat{\mu}_m) - \lambda_{1,m} = -\frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m} + O_p(1)$
- $\lambda_N(\hat{\mu}_m) - \lambda_{2,m} = -\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e'_{2,m}X_l e_{2,m} + O_p(1)$
- $||v_{11}(\hat{\mu}_m) - e_{1,m}|| = O_p(1)$, $v_{12}(\hat{\mu}_m) = O_p(1)$ *and* $r_1(\hat{\mu}_m) = O_p(1)$

- $||v_{21}(\hat{\mu}_m)|| = O_p(1)$, $||v_{22}(\hat{\mu}_m) - e_2|| = O_p(1)$ *and* $r_2(\hat{\mu}_m) = O_p(1)$
- $\lambda_1(\hat{\mu}_m)(v_{11}(\hat{\mu}_m) - e_{1,m}) = -\sum_l Z_{m,N,l} X_l v_{11}(\hat{\mu}_m) + (\lambda_{1,m} - \lambda_1(\hat{\mu}_m) + O_p(1))e_{1,m} + \Big(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2} \sum_l Z_{m,N,l} e_{2,m}$ $V v_{11}(\tilde{\mu}) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m)$
- $\lambda_N(\hat{\mu}_m)(v_{22}(\hat{\mu}_m) - e_{2,m}) = -\sum_l Z_{m,N,l} X_l v_{22}(\hat{\mu}_m) + (\lambda_{2,m} - \lambda_N(\hat{\mu}_m) + O_p(1))e_{2,m} + \Big(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2} \sum_l Z_{m,N,l} e_2$ $V v_{22}(\tilde{\mu}) - \sigma_A c_{m,N} v_{22}(\hat{\mu}_m)$

Proof.

- $\boldsymbol{||v_{11}(\hat{\mu}_m) - e_{1,m}|| = O_p(1)}$, $\boldsymbol{v_{12}(\hat{\mu}_m) = O_p(1)}$ **and** $\boldsymbol{r_1(\hat{\mu}_m) = O_p(1)}$ On one side:

$$M(\hat{\mu}_m)e_{1,m} := -\sum_l Z_{m,N,l} X_l e_{1,m} + V e_{1,m} + \lambda_{1,m} e_{1,m} - \sigma_A c_{m,N} e_{1,m}$$

on another side:

$$M(\hat{\mu}_m)e_{1,m} := M(\hat{\mu}_m) r_1(\hat{\mu}_m) + \lambda_1(\hat{\mu}_m) v_{11}(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) v_{12}(\hat{\mu}_m)$$

so

$$M(\hat{\mu}_m) r_1(\hat{\mu}_m) + \lambda_1(\hat{\mu}_m) v_{11}(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) v_{12}(\hat{\mu}_m) = -\sum_l Z_{m,N,l} X_l e_{1,m} + V e_1 + \lambda_{1,m} e_{1,m} - \sigma_A c_{m,N} e_{1,m}$$

multiplying by $v_{12}(\hat{\mu}_m)$ on both sides:

$$(\lambda_N(\hat{\mu}_m) - \lambda_{1,m})||v_{12}(\hat{\mu}_m)|| = O_p(N)$$

similarly

$$(\lambda_1(\hat{\mu}_m) - \lambda_{1,m})||v_{11}(\hat{\mu}_m)|| = O_p(N)$$

First, by the interlacement theorem (e.g. Bunch et al. (1978)), for all $i = 2..N$:

$$\lambda_i\left(M(\hat{\mu}_m) - e_{1,m}e'_{1,m}\right) \leq \lambda_i\left(M(\hat{\mu}_m)\right) \leq \lambda_{i-1}\left(M(\hat{\mu}_m) - e_{1,m}e'_{1,m}\right)$$

and for all $i = 1..N-1$:

$$\lambda_{i+1}\left(M(\hat{\mu}_m) - \left(e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m}\right)\right) \leq \lambda_i\left(M(\hat{\mu}_m) - e_{1,m}e'_{1,m}\right) \leq \lambda_i\left(M(\hat{\mu}_m) - \left(e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m}\right)\right)$$

therefore, for all $i = 2..N-1$

$$\lambda_{i+1}\left(M(\hat{\mu}_m) - \left(e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m}\right)\right) \leq \lambda_i\left(M(\hat{\mu}_m)\right) \leq \lambda_{i-1}\left(M(\hat{\mu}_m) - \left(e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m}\right)\right)$$

so

$$\max_{i=2..N-1} |\lambda_i\left(M(\hat{\mu}_m)\right)| = O_p(\sqrt{N})$$

also

$$\begin{aligned}\lambda_N\left(M(\hat{\mu}_m)\right) &\leq \lambda_{N-1}\left(M(\hat{\mu}_m) - e_{1,m}e'_{1,m}\right)\\ &\leq \lambda_{N-1}\left(M(\hat{\mu}_m) - \left(e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m}\right)\right)\end{aligned}$$

which implies that

$$||v_{12}(\hat{\mu}_m)||^2 = O_p(1)$$

to see that $||r_1(\hat{\mu}_m)||^2 = O_p(1)$, as for the proof of 8,

$$||M(\hat{\mu}_m)r(\hat{\mu}_m) - \lambda_{1,m}r_1(\hat{\mu}_m)|| \geq \lambda_{1,m}r(\hat{\mu}_m)||r_1(\hat{\mu}_m)|| - ||M(\hat{\mu}_m)r_1(\hat{\mu}_m)|| \geq (\lambda_{1,m} - \max_{i=2..N-1}|\lambda_i(M(\hat{\mu}_m))|)$$

and by the Pythagorean theorem:

$$||M(\hat{\mu}_m)r(\hat{\mu}_m) - \lambda_{1,m}r_1(\hat{\mu}_m)||^2 \leq ||M(\hat{\mu}_m)e_1 - \lambda_{1,m}e_1||^2 = ||-\sum_l Z_{m,N,l}X_le_1 + Ve_1 - \sigma_A c_{m,N}e_1||^2 = O_p(N^2)$$

in conclusion:

$$||r_1(\hat{\mu}_m)||^2 \leq \frac{||M(\hat{\mu}_m)e_{1,m} - \lambda_{1,m}e_{1.m}||^2}{\left(\lambda_{1,m} - \max_{i=2..N-1}|\lambda_i(M(\hat{\mu}_m))|\right)^2} = O_p(1)$$

- $\boldsymbol{\lambda_1(\hat{\mu}_m) - \lambda_{1,m} = -\frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_le_{1,m} + O_p(1)}$

  We established:

$$M(\hat{\mu}_m)r_1(\hat{\mu}_m) + \lambda_1(\hat{\mu}_m)v_{11}(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m)v_{12}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}X_le_{1,m} + Ve_{1,m} + \lambda_{1,m}e_{1,m} - \sigma_A c_{m,N}e_{1,m}$$

multiply by $v_{11}(\hat{\mu}_m)$ on both sides:

$$\begin{aligned}(\lambda_1(\hat{\mu}_m) - \lambda_{1,m})||v_{11}(\hat{\mu}_m)||^2 &= -\sum_l Z_{m,N,l}v_{11}(\hat{\mu}_m)X_le_{1,m} + v_{11}(\hat{\mu}_m)'Ve_{1,m} - \sigma_A c_{m,N}||v_{11}(\hat{\mu}_m)||^2\\ &= -\sum_l Z_{m,N,l}v_{11}(\hat{\mu}_m)X_le_{1,m} + e'_{1,m}Ve_{1,m} + O_p(N)\\ &= -\sum_l Z_{m,N,l}v_{11}(\hat{\mu}_m)X_le_{1,m} + O_p(N)\end{aligned}$$

so $\lambda_1(\hat{\mu}_m) - \lambda_{1,m} = -\frac{1}{||e_1||^2}\sum_l Z_{m,N,l}e'_{1,m}X_le_{1,m} + O_p(1)$

- $\boldsymbol{\lambda_1(\hat{\mu}_m)(v_{11}(\hat{\mu}_m) - e_{1,m}) = -\sum_l Z_{m,N,l} X_l v_{11}(\hat{\mu}_m) + (\lambda_{1,m} - \lambda_1(\hat{\mu}_m) + O_p(1)) e_{1,m} + \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2} \sum_l Z_{m,N,l} e_{2,m} X_l e_{1,m} + O_p(1)\right) e_{2,m} + V v_{11}(\tilde{\mu}) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m)}$

  Write:

$$\lambda_1(\hat{\mu}_m)v_{11}(\hat{\mu}_m) = M(\hat{\mu}_m)v_{11}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + Vv_{11}(\hat{\mu}_m) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m) + \left(||v_{11}(\hat{\mu}_m)||^2 e_{1,m} - (e'_{2,m}v_{11}(\hat{\mu}_m))e_{2,m}\right)$$

so

$$\begin{aligned}\lambda_1(\hat{\mu}_m)(v_{11}(\hat{\mu}_m) - e_{1,m}) &= -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + Vv_{11}(\hat{\mu}_m) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m)\\ &\quad + \left(||v_{11}(\hat{\mu}_m)||^2 - \lambda_1(\hat{\mu}_m)\right) e_{1,m} - (e'_{2,m}v_{11}(\hat{\mu}_m))e_{2,m}\\ &= -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + Vv_{11}(\hat{\mu}_m) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m)\\ &\quad + \left(\lambda_{1,m} - \lambda_1(\hat{\mu}_m) + O_p(1)\right) e_{1,m} - (e'_{2,m}v_{11}(\hat{\mu}_m))e_{2,m}\end{aligned}$$

Let's find an asymptotic approximation for $e'_{2,m}v_{11}(\tilde{\mu})$. We have shown:

$$\lambda_1(\hat{\mu}_m)v_{11}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + Vv_{11}(\hat{\mu}_m) - \sigma_A c_{m,N} v_{11}(\hat{\mu}_m) + \left(||v_{11}(\hat{\mu}_m)||^2 e_{1,m} - (e'_{2,m}v_{11}(\hat{\mu}_m))e_{2,m}\right)$$

so

$$\lambda_1(\hat{\mu}_m)e'_{2,m}v_{11}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}e_2X_l v_{11}(\hat{\mu}_m) + e_2Vv_{11}(\hat{\mu}_m) + \lambda_{2,m}(e'_{2,m}v_{11}(\hat{\mu}_m)) - \sigma_A c_{m,N}(e'_{2,m}v_{11}(\hat{\mu}_m))$$

implying

$$\begin{aligned}(\lambda_1(\hat{\mu}_m) - \lambda_{2,m})e'_{2,m}v_{11}(\hat{\mu}_m) &= -\sum_l Z_{m,N,l}e_{2,m}X_l v_{11}(\hat{\mu}_m) + e_{2,m}Vv_{11}(\hat{\mu}_m) - \sigma_A c_{m,N}(e'_{2,m}v_{11}(\hat{\mu}_m))\\ &= -\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m} + O_p(N)\end{aligned}$$

or

$$\left(\lambda_{1,m} - \lambda_{2,m} + O_p(\sqrt{N})\right) e'_2 v_{11}(\tilde{\mu}) = -\sum_l Z_{m,N,l}e_2X_l e_1 + O_p(N)$$

with

$$\lambda_{1,m} = ||e_{1,m}||^2$$
$$= \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 N + b_{m,N}^2||A||^2 + 2b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\iota' A$$

and

$$\lambda_{2,m} = -||e_{2,m}||^2$$
$$= -\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 N - b_{m,N}^2||A||^2 - 2b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)\iota' A$$

therefore:

$$e_{2,m}'v_{11}(\hat{\mu}_m) = -\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m} + O_p(N)$$

- $\boldsymbol{\lambda_2(\hat{\mu}_m)(v_{22}(\hat{\mu}_m) - e_{2,m}) = -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m) + (\lambda_{2,m} - \lambda_2(\hat{\mu}_m) + O_p(1))e_{2,m} + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m} + O_p(1)\right)e_{1,m} + V v_{22}(\tilde{\mu}) - \sigma_A c_{m,N} v_{22}(\hat{\mu}_m)}$

  Write:

$$\lambda_2(\hat{\mu}_m)v_{22}(\hat{\mu}_m) = M(\hat{\mu}_m)v_{22}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m) + V v_{22}(\hat{\mu}_m) - \sigma_A c_{m,N}v_{22}(\hat{\mu}_m)$$
$$+ \left((e_{1,m}'v_{22}(\hat{\mu}_m))e_{1,m} - ||v_{22}(\hat{\mu}_m)||^2 e_{2,m}\right)$$

so

$$\lambda_2(\hat{\mu}_m)(v_{22}(\hat{\mu}_m) - e_{2,m}) = -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m) + V v_{22}(\hat{\mu}_m) - \sigma_A c_{m,N}v_{22}(\hat{\mu}_m)$$
$$+ \left(-||v_{22}(\hat{\mu}_m)||^2 - \lambda_2(\hat{\mu}_m)\right)e_{2,m} + (e_{2,m}'v_{11}(\hat{\mu}_m))e_{1,m}$$
$$= -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m) + V v_{22}(\hat{\mu}_m) - \sigma_A c_{m,N}v_{22}(\hat{\mu}_m)$$
$$+ \left(\lambda_{2,m} - \lambda_2(\hat{\mu}_m) + O_p(1)\right)e_{2,m} - (e_{2,m}'v_{11}(\hat{\mu}_m))e_{1,m}$$

Let's find an asymptotic approximation for $e_{1,m}'v_{22}(\tilde{\mu})$. We have shown:

$$\lambda_2(\hat{\mu}_m)v_{22}(\hat{\mu}_m) = -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m) + V v_{22}(\hat{\mu}_m) - \sigma_A c_{m,N}v_{22}(\hat{\mu}_m)$$

$$+\left(-||v_{22}(\hat{\mu}_m)||^2 e_{2,m}+(e_{1,m}'v_{22}(\hat{\mu}_m))e_{1,m}\right)$$

so

$$\begin{aligned}\lambda_2(\hat{\mu}_m)e_{1,m}'v_{22}(\hat{\mu}_m) &= -\sum_l Z_{m,N,l}e_{1,m}X_l v_{22}(\hat{\mu}_m)+e_{1,m}Vv_{22}(\hat{\mu}_m)+\lambda_{1,m}(e_{2,m}'v_{11}(\hat{\mu}_m))\\ &\quad -\sigma_A c_{m,N}(e_{1,m}'v_{22}(\hat{\mu}_m))\end{aligned}$$

implying

$$\begin{aligned}(\lambda_2(\hat{\mu}_m)-\lambda_{1,m})e_{1,m}'v_{22}(\hat{\mu}_m) &= -\sum_l Z_{m,N,l}e_{1,m}X_l v_{22}(\hat{\mu}_m)+e_{1,m}Vv_{22}(\hat{\mu}_m)-\sigma_A c_{m,N}(e_{1,m}'v_{22}(\hat{\mu}_m))\\ &= -\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}+O_p(N)\end{aligned}$$

or

$$\left(\lambda_{2,m}-\lambda_{1,m}+O_p(\sqrt{N})\right)e_1'v_{22}(\tilde{\mu}) = -\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}+O_p(N)$$

therefore:

$$e_{2,m}'v_{11}(\hat{\mu}_m)=\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}+O_p(N)$$

□

Note that:

$$\hat{\mu}_{m+1}=\left(\sum_{i\neq j}X_{ij}'X_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'X_{ik}\right)^{-1}\left(\sum_{i\neq j}X_{ij}'Y_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'Y_{ik}\right)$$

where $\nu(\hat{\mu}_m)$ is the normalized eigenvector corresponding to the largest eigenvalue of $M(\hat{\mu}_m)^2$.
So

$$\begin{aligned}\hat{\mu}_{m+1}-\mu_0 &= \left(\sum_{i\neq j}X_{ij}'X_{ij}-\sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'X_{ik}\right)^{-1}\\ &\quad\times\left(\sum_{i\neq j}X_{ij}'(A_i+A_j+V_{ij})-\sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'(A_i+A_k+V_{ik})\right)\end{aligned}$$

First, note:

$$\sum_{i\neq j} X_{ij}'X_{ij} - \sum_{i\neq j,k\neq i,j} \nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'X_{ik} = \mathbb{1}(\lambda_1(\hat{\mu}_m)+\lambda_N(\hat{\mu}_m)\geq 0)\left(\sum_{i\neq j} X_{ij}'X_{ij} - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk}'X_{i\ldots}\right.$$

$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m)+\lambda_N(\hat{\mu}_m)< 0)\left(\sum_{i\neq j} X_{ij}'X_{ij} - \sum_{i\neq j,k\neq i,j} \frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk}'X_{i\ldots}\right.$$

We treat each of the two terms separately:

$$\sum_{i\neq j} X_{ij}'X_{ij} - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk}'X_{ik} = N^2\left(E(X_{12}X_{12}') - \frac{E(e_{1,m})^2}{||e_{1,m}||^2/N}E(X_{12}X_{23}) + o_p(1)\right)$$

$$= N^2\left(E(X_{12}X_{12}') - \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}E(X_{12}X_{23}) + o_p(1)\right)$$

likewise

$$\sum_{i\neq j} X_{ij}'X_{ij} - \sum_{i\neq j,k\neq i,j} \frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk}'X_{ik}$$

$$= N^2\left(E(X_{12}X_{12}') - \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}E(X_{12}X_{23}) + o_p(1)\right)$$

For the term $\left(\sum_{i\neq j} X_{ij}'(A_i+A_j+V_{ij}) - \sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'(A_i+A_k+V_{ik})\right)$, as for the proof of proposition 1, let η be some vector $\eta\in\mathbb{R}^L$, and define the matrix X_η with entries $X_{ij,\eta} := \eta' X_{ij}$.

$$\sum_{i\neq j} X_{ij}'(A_i+A_j+V_{ij}) - \sum_{i\neq j,k\neq i,j}\nu_i(\hat{\mu}_m)\nu_j(\hat{\mu}_m)X_{jk}'(A_i+A_k+V_{ik})$$

$$= \mathbb{1}(\lambda_1(\hat{\mu}_m)+\lambda_N(\hat{\mu}_m)\geq 0)\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\right)$$

$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m)+\lambda_N(\hat{\mu}_m)< 0)\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\right)$$

We have

$$\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)$$

$$= \sum_{i \neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i \neq j, k \neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} \frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} X_{jk,\eta} \left(M(\hat{\mu}_m)_{ik} - \sum_l (\mu_{0,l} - \hat{\mu}_{m,l}) X_{ik,l} \right)$$

$$= \sum_{i \neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \frac{1}{||v_{11}(\hat{\mu}_m)||^2} v_{11}(\hat{\mu}_m)' M(\hat{\mu}_m) X_\eta v_{11}(\hat{\mu}_m)$$

$$- \frac{1}{||v_{11}(\hat{\mu}_m)||^2} \sum_l (\mu_{0,l} - \hat{\mu}_{m,l}) v_{11}(\tilde{\mu})' X_l X_\eta v_{11}(\hat{\mu}_m) + \sum_{ik} \frac{v_{11,i}^2(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||^2} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) + O_p(N)$$

Note that:

$$\left| \sum_{ik} v_{11,i}^2 X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) - \sum_{ik} e_{1,i}^2 X_{ik,\eta}\left(A_i + A_k + V_{ik}\right) \right| \leq \sum_{ik} |v_{11,i}^2 - e_{1,i}^2| |X_{ik,\eta}\left(A_i + A_k + V_{ik}\right)|$$

$$\leq \max_k \sum_i |X_{ik,\eta}\left(A_i + A_k + V_{ik}\right)| \times \sum_i |v_{11,i}^2 - e_{1,i}^2|$$

$$\leq \max_k \sum_i \left(|X_{ik,\eta}\left(A_i + A_k + V_{ik}\right)| - E(|X_{ik,\eta}\left(A_i + A_k + V_{ik}\right)|) \right) \times ||v_{11} - e_1||_2 \times ||v_{11} + e_1||_2$$

$$+ NE(|X_{ik,\eta}\left(A_i + A_k + V_{ik}\right)|) \times ||v_{11} - e_1||_2 \times ||v_{11} + e_1||_2$$

Let's show that

$$\max_k \sum_i \left(\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| - E\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| \right) = O_p(N\sqrt{N})$$

Fix some $x > 0$ and by a union bound:

$$\mathbb{P}\left(\frac{1}{N\sqrt{N}} \max_k \sum_i \left(\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| - E\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| \right) \geq x \right)$$

$$\leq \sum_k \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| - E\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| \right) \geq x \right)$$

$$= N \times \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| - E\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| \right) \geq x \right)$$

$$\leq \frac{1}{N^2} \frac{Var\left(\sum_i \left(\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| - E\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| \right) \right)}{x^2}$$

$$\leq \frac{1}{x^2} \left(Var\left(\left|X_{12,\eta}(A_2 + A_1 + V_{12})\right| \right) + Cov\left(\left|X_{12,\eta}(A_2 + A_1 + V_{12})\right|, \left|X_{13,\eta}(A_3 + A_1 + V_{13})\right| \right) \right)$$

where the second inequality is Markov's. This implies:

$$\max_k \sum_i \left(\left| X_{ik,\eta}(A_i + A_k + V_{ik}) \right| - E\left| X_{ik,\eta}(A_i + A_k + V_{ik}) \right| \right) = O_p(N\sqrt{N})$$

we can infer

$$\sum_{i,k} \frac{v_{11,i}^2(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||^2} X_{ik,\eta} \left(A_i + A_k + V_{ik} \right) = O_p(N)$$

hence

$$\begin{aligned}
&\sum_{i \neq j} X_{ij,\eta} \left(A_i + A_j + V_{ij} \right) - \sum_{i \neq j, k \neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} \frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} X_{jk,\eta} \left(A_i + A_k + V_{ik} \right) \\
&= -\frac{1}{||v_{11}(\hat{\mu}_m)||^2} v_{11}(\hat{\mu}_m)' M(\hat{\mu}_m) X_\eta v_{11}(\hat{\mu}_m) + \frac{1}{||v_{11}(\hat{\mu}_m)||^2} \sum_l (\mu_{0,l} - \hat{\mu}_{m,l}) v_{11}(\hat{\mu}_m)' X_l X_\eta v_{11}(\hat{\mu}_m) + O_p(N\sqrt{N}) \\
&= -\frac{\lambda_1(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||^2} v_{11}(\hat{\mu}_m)' X_\eta v_{11}(\hat{\mu}_m) + \frac{1}{||v_{11}(\hat{\mu}_m)||^2} \sum_l (\mu_{0,l} - \hat{\mu}_{m,l}) v_{11}(\hat{\mu}_m)' X_l X_\eta v_{11}(\hat{\mu}_m) + O_p(N\sqrt{N}) \\
&= -\frac{\lambda_{1,m}}{||e_{1,m}||^2} e_{1,m}' X_\eta e_{1,m} + \frac{c_{m,N}}{||v_{11}(\hat{\mu}_m)||^2} \sigma_A v_{11}(\hat{\mu}_m)' \iota \iota' X_\eta v_{11}(\hat{\mu}_m) + O_p(N\sqrt{N}) \\
&= -e_{1,m}' X_\eta e_{1,m} + \frac{c_{m,N}}{||e_{1,m}||^2} \sigma_A e_{1,m}' \iota \iota' X_\eta e_{1,m} + O_p(N\sqrt{N})
\end{aligned}$$

Note that

$$\begin{aligned}
e_{1,m}' \iota &= \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right) N + b_{m,N} A' \iota; \\
e_{1,m}' X_\eta e_{1,m} &= \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right)^2 \sum_{ij} X_{ij,\eta} + 2 b_{m,N} \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right) \sum_{ij} X_{ij,\eta} A_j \\
&\quad + b_{m,N}^2 \sum_{ij} A_i A_j X_{ij,\eta} \\
&= N^2 \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right)^2 E(X_{ij,\eta}) + O_p(N\sqrt{N}); \\
\iota' X_\eta e_{1,m} &= \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right) \sum_{ij} X_{ij,\eta} + b_{m,N} \sum_{ij} X_{ij,\eta} A_j \\
&= N^2 \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2 b_{m,N}} \right) E(X_{ij,\eta}) + O_p(N\sqrt{N})
\end{aligned}$$

so

$$\sum_{i \neq j} X_{ij,\eta} \left(A_i + A_j + V_{ij} \right) - \sum_{i \neq j, k \neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} \frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} X_{jk,\eta} \left(A_i + A_k + V_{ik} \right)$$

$$= N^2 \left(-\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + \frac{Nc_{m,N}\sigma_A}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 N + b_{m,N}^2||A||^2} \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \right) E(X_{ij,\eta})$$

$$+ O_p(N\sqrt{N})$$

$$= N^2 \left(-1 + \frac{c_{m,N}\sigma_A}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2||A||^2/N} \right) \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

$$= -N^2 \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2||A||^2/N}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2||A||^2/N} \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

$$= -N^2 \frac{\left(\frac{1}{4}b_{m,N}^2c_{m,N}^2\sigma_A^2 - \frac{1}{4b_{m,N}^2}\right)^2 + b_{m,N}^2\sigma_A^2\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

$$= -N^2 \frac{b_{m,N}^4\sigma_A^4 + b_{m,N}^2\sigma_A^2\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

Similarly

$$\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right)$$

$$= -\frac{1}{||v_{22}(\hat{\mu}_m)||^2}v_{22}(\hat{\mu}_m)'M(\hat{\mu}_m)X_\eta v_{22}(\hat{\mu}_m) + \frac{1}{||v_{22}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l} - \hat{\mu}_{m,l})v_{22}(\hat{\mu}_m)'X_lX_\eta v_{22}(\hat{\mu}_m) + O_p(N\sqrt{N})$$

$$= -\frac{\lambda_2(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||^2}v_{22}(\hat{\mu}_m)'X_\eta v_{22}(\hat{\mu}_m) + \frac{1}{||v_{22}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l} - \hat{\mu}_{m,l})v_{22}(\hat{\mu}_m)'X_lX_\eta v_{22}(\hat{\mu}_m) + O_p(N\sqrt{N})$$

$$= -\frac{\lambda_{2,m}}{||e_{2,m}||^2}e_{2,m}'X_\eta e_{2,m} + \frac{c_{m,N}}{||v_{22}(\hat{\mu}_m)||^2}\sigma_A v_{22}(\hat{\mu}_m)'\iota\iota'X_\eta v_{22}(\hat{\mu}_m) + O_p(N\sqrt{N})$$

$$= e_{2,m}'X_\eta e_{2,m} + \frac{c_{m,N}}{||e_{2,m}||^2}\sigma_A e_{2,m}'\iota\iota'X_\eta e_{2,m} + O_p(N\sqrt{N})$$

$$= \left(\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + \frac{c_{m,N}\sigma_A}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 \right) E(X_{ij,\eta}) +$$

$$= \left(1 + \frac{c_{m,N}\sigma_A}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} \right) \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

$$= \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

$$= \frac{b_{m,N}^4\sigma_A^4 + b_{m,N}^2\sigma_A^2\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} E(X_{ij,\eta}) + O_p(N\sqrt{N})$$

Therefore

$$\hat{\mu}_{m+1} - \mu_0 = \Bigg[-\mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) \geq 0)\left(b_{m,N}^2\sigma_A^2 + \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2\right)$$
$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) < 0)\left(b_{m,N}^2\sigma_A^2 + \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2\right)\Bigg](1,0,0,...,0)' + O_p\left(\frac{1}{\sqrt{N}}\right.$$
$$= \Bigg[-\mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) \geq 0)\left(b_{m,N}^2\sigma_A^2 + \frac{1}{4}b_{m,N}^2\left(c_{m,N}\sigma_A + \frac{1}{b_{m,N}^2}\right)^2\right)$$
$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) < 0)\left(b_{m,N}^2\sigma_A^2 + \frac{1}{4}b_{m,N}^2\left(c_{m,N}\sigma_A - \frac{1}{b_{m,N}^2}\right)^2\right)\Bigg](1,0,0,...,0)' + O_p\left(\frac{1}{\sqrt{N}}\right)$$
$$= \Bigg[-\mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) \geq 0)\left(\frac{\sigma_A}{\sqrt{4+c_{m,N}^2}} + \frac{1}{4\sigma_A\sqrt{4+c_{m,N}^2}}\left(c_{m,N}\sigma_A + \sigma_A\sqrt{4+c_{m,N}^2}\right)^2\right)$$
$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) < 0)\left(\frac{\sigma_A}{\sqrt{4+c_{m,N}^2}} + \frac{1}{4\sigma_A\sqrt{4+c_{m,N}^2}}\left(c_{m,N}\sigma_A - \sigma_A\sqrt{4+c_{m,N}^2}\right)^2\right)\Bigg](1,0$$
$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$
$$= \frac{\sigma_A}{\sqrt{4+c_{m,N}^2}}\Bigg[-\mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) \geq 0)\left(1 + \frac{1}{4}\left(c_{m,N} + \sqrt{4+c_{m,N}^2}\right)^2\right)$$
$$+ \mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) < 0)\left(1 + \frac{1}{4}\left(c_{m,N} - \sqrt{4+c_{m,N}^2}\right)^2\right)\Bigg](1,0,0,...,0)' + O_p\left(\frac{1}{\sqrt{N}}\right)$$

Note that for any m, by lemma 5:

$$\frac{\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m)}{N} = \frac{\lambda_{1,m} + \lambda_{2,m}}{N} + O_p\left(\frac{1}{\sqrt{N}}\right)$$

Given that:

$$\lambda_{1,m} = \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 N + b_{m,N}^2||A||^2 + 2b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\iota' A$$

and

$$\lambda_{2,m} = -\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 N - b_{m,N}^2||A||^2 - 2b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)\iota' A$$

then, whenever $m > 1$:

$$\begin{aligned}\frac{\lambda_{1,m} + \lambda_{2,m}}{N} &= \frac{c_{m,N}\sigma_A N + 2\iota' A}{N} + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= c_{m,N}\sigma_A + O_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned}$$

so, for all $m > 1$:

$$\begin{aligned}\hat{\mu}_{m+1} - \mu_0 = \frac{\sigma_A}{\sqrt{4 + c_{m,N}^2}}\Bigg[&- \mathbb{1}(c_{m,N} \geq 0)\left(1 + \frac{1}{4}\left(c_{m,N} + \sqrt{4 + c_{m,N}^2}\right)^2\right) \\ &+ \mathbb{1}(c_{m,N} < 0)\left(1 + \frac{1}{4}\left(c_{m,N} - \sqrt{4 + c_{m,N}^2}\right)^2\right)\Bigg](1, 0, 0, ..., 0)' + O_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned}$$

$$\begin{aligned}c_{m+1,N} = \frac{1}{\sqrt{4 + c_{m,N}^2}}\Bigg[&\mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) \geq 0)\left(1 + \frac{1}{4}\left(c_{m,N} + \sqrt{4 + c_{m,N}^2}\right)^2\right) \\ &- \mathbb{1}(\lambda_1(\hat{\mu}_m) + \lambda_N(\hat{\mu}_m) < 0)\left(1 + \frac{1}{4}\left(c_{m,N} - \sqrt{4 + c_{m,N}^2}\right)^2\right)\Bigg]\end{aligned}$$

Proposition 10. *For all $m_0 \in \mathbb{N}$:*

$$\lim_N \mathbb{P}\left(\forall m \leq m_0 : c_{m+1,N} = \frac{1}{\sqrt{4 + c_{m,N}^2}}\left(1 + \frac{1}{4}\left(c_{m,N} + \sqrt{4 + c_{m,N}^2}\right)^2\right) \middle| c_{1,N} = 1\right) = 1$$

and

$$\lim_N \mathbb{P}\left(\forall m \le m_0 : c_{m+1,N} = -\frac{1}{\sqrt{4+c_{m,N}^2}}\left(1+\frac{1}{4}\left(c_{m,N}-\sqrt{4+c_{m,N}^2}\right)^2\right)\middle| c_{1,N}=-1\right)=1$$

Proof. Immediately follows from the computations above. □

Corollary 6. *Define the deterministic sequence c_m by:*

$$\begin{cases} c_1 &= 1 \\ c_{m+1} &= \frac{1}{\sqrt{4+c_m^2}}\left(1+\frac{1}{4}\left(c_m+\sqrt{4+c_m^2}\right)^2\right) \end{cases}$$

Then for all $m_0 \in \mathbb{N}$:

$$\lim_N \mathbb{P}\left(\forall m \le m_0 : c_{m,N} = c_m \middle| c_{1,N}=1\right)=1$$

and

$$\lim_N \mathbb{P}\left(\forall m \le m_0 : c_{m,N} = -c_m \middle| c_{1,N}=-1\right)=1$$

Proof. Direct consequence of proposition 10. □

Let's compute the second order (order $O_p\left(\frac{1}{\sqrt{N}}\right)$) term:

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) - \sum_{i\neq j,k\neq i,j}\frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\\
&= -\frac{1}{||v_{11}(\hat{\mu}_m)||^2}v_{11}(\hat{\mu}_m)'M(\hat{\mu}_m)X_\eta v_{11}(\hat{\mu}_m) + \frac{1}{||v_{11}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l}-\hat{\mu}_{m,l})v_{11}(\hat{\mu}_m)'X_lX_\eta v_{11}(\hat{\mu}_m)\\
&\quad + \sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) + O_p(N)\\
&= -\frac{\lambda_1(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||^2}v_{11}(\hat{\mu}_m)'X_\eta v_{11}(\hat{\mu}_m) + \frac{1}{||v_{11}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l}-\hat{\mu}_{m,l})v_{11}(\hat{\mu}_m)'X_lX_\eta v_{11}(\hat{\mu}_m)\\
&\quad + \sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) + O_p(N)\\
&= -\frac{\lambda_{1,m}}{||e_{1,m}||^2}e_{1,m}'X_\eta e_{1,m} + \frac{c_{m,N}}{||v_{11}(\hat{\mu}_m)||^2}\sigma_A v_{11}(\hat{\mu}_m)'\iota\iota'X_\eta v_{11}(\hat{\mu}_m)\\
&\quad + \sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right) + 2(e_{1,m}-v_{11}(\hat{\mu}_m))'X_\eta e_{1,m} - \frac{1}{||v_{11}(\hat{\mu}_m)||^2}\sum_l Z_{m,N,l}e_{1,m}'X_lX_\eta e_{1,m}\\
&\quad + \frac{\lambda_{1,m}-\lambda_1(\hat{\mu}_m)}{||e_{1,m}||^2}e_{1,m}'X_\eta e_{1,m} + O_p(N)\\
&= -e_{1,m}'X_\eta e_{1,m} + \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A e_{1,m}'\iota\iota'X_\eta e_{1,m}
\end{aligned}$$

$$
\begin{aligned}
&+ 2\sum_{i\neq j} X_{ij,\eta}A_i + 2(e_{1,m} - v_{11}(\hat{\mu}_m))' X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l X_\eta e_{1,m} \\
&+ \frac{\lambda_{1,m} - \lambda_1(\hat{\mu}_m)}{||e_{1,m}||^2} e'_{1,m} X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A \Big((e_{1,m} - v_{11}(\hat{\mu}_m))'\iota\iota' X_\eta e_{1,m} + e'_{1,m}\iota\iota' X_\eta (e_{1,m} - v_{11}(\hat{\mu}_m))\Big) + O_p(N) \\
= &-e'_{1,m} X_\eta e_{1,m} + \frac{N c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\iota' X_\eta e_{1,m} \\
&+ 2\sum_{i\neq j} X_{ij,\eta}A_i + 2(e_{1,m} - v_{11}(\hat{\mu}_m))' X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l X_\eta e_{1,m} + b_{m,N} A'\iota\iota' X_\eta e_{1,m} \\
&+ \frac{\lambda_{1,m} - \lambda_1(\hat{\mu}_m)}{||e_{1,m}||^2} e'_{1,m} X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A \Big((e_{1,m} - v_{11}(\hat{\mu}_m))'\iota\iota' X_\eta e_{1,m} + e'_{1,m}\iota\iota' X_\eta (e_{1,m} - v_{11}(\hat{\mu}_m))\Big) + O_p(N) \\
= &-e'_{1,m} X_\eta e_{1,m} + \frac{N c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\iota' X_\eta e_{1,m} \\
&+ 2\sum_{i\neq j} X_{ij,\eta}A_i + \frac{1}{||e_{1,m}||^4}\sum_l Z_{m,N,l} e'_{1.m} X_l e_{1,m} e'_{1,m} X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l X_\eta e_{1,m} + b_{m,N} A'\iota\iota' X_\eta e_{1,m} \\
&- \frac{1}{||e_{1,m}||^2}\Bigg(-\sum_l Z_{m,N,l} X_l e_{1,m} + \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l e_{1,m} e_{1,m} + \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l} e_{2,m} X_l e_{1,m}\right) e_{2,m} \\
&\times \left(2X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\iota\iota' X_\eta e_{1,m} + X_\eta \iota\iota' e_{1,m}\right)\right) + O_p(N) \\
= &-\left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \sum_{ij} X_{ij,\eta} - 2b_{m,N}\left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\sum_{ij} X_{ij,\eta}A_j \\
&+ \frac{N c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \sum_{ij} X_{ij,\eta} + \frac{N c_{m,N} b_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)\sum_{ij} X_{ij,\eta}A_j \\
&+ 2\sum_{i\neq j} X_{ij,\eta}A_i + \frac{1}{||e_{1,m}||^4}\sum_l Z_{m,N,l} e'_{1.m} X_l e_{1,m} e'_{1,m} X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l X_\eta e_{1,m} + b_{m,N} A'\iota\iota' X_\eta e_{1,m} \\
&- \frac{1}{||e_{1,m}||^2}\Bigg(-\sum_l Z_{m,N,l} X_l e_{1,m} + \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_\eta e_{1,m} e_{1,m} + \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l} e_{2,m} X_l e_{1,m}\right) e_{2,m} \\
&\times \left(2X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\iota\iota' X_\eta e_{1,m} + X_\eta \iota\iota' e_{1,m}\right)\right) + O_p(N) \\
= &\left(\frac{N c_{m,N}}{||e_{1,m}||^2}\sigma_A - 1\right)\left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \sum_{ij} X_{ij,\eta} \\
&+ \left(\left(\frac{N c_{m,N}}{||e_{1,m}||^2}\sigma_A - 2\right) b_{m,N}\left(\frac{1}{2} b_{m,N} c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right) + 2\right)\sum_{ij} X_{ij,\eta}A_j \\
&+ \frac{1}{||e_{1,m}||^4}\sum_l Z_{m,N,l} e'_{1.m} X_l e_{1,m} e'_{1,m} X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_l X_\eta e_{1,m} + b_{m,N} A'\iota\iota' X_\eta e_{1,m} \\
&- \frac{1}{||e_{1,m}||^2}\Bigg(-\sum_l Z_{m,N,l} X_l e_{1,m} + \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l} e'_{1,m} X_\eta e_{1,m} e_{1,m} + \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l} e_{2,m} X_l e_{1,m}\right) e_{2,m} \\
&\times \left(2X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A \left(\iota\iota' X_\eta e_{1,m} + X_\eta \iota\iota' e_{1,m}\right)\right) + O_p(N)
\end{aligned}
$$

$$
\begin{aligned}
&= N^2\left(\frac{Nc_{m,N}}{||e_{1,m}||^2}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \eta' E(X_{12}) + \eta' T_{1,m,N} \\
&\quad + \frac{1}{||e_{1,m}||^4}\sum_l Z_{m,N,l}e'_{1.m}X_l e_{1,m}e'_{1,m}X_\eta e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l X_\eta e_{1,m} \\
&\quad - \frac{1}{||e_{1,m}||^2}\left(- \sum_l Z_{m,N,l}X_l e_{1,m} + \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_\eta e_{1,m}e_{1,m} + \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right) e_{2,m} \right. \\
&\quad \times \left(2X_\eta e_{1,m} - \frac{c_{m,N}}{||e_{1,m}||^2}\sigma_A\left(\iota\iota' X_\eta e_{1,m} + X_\eta \iota\iota' e_{1,m}\right)\right) + O_p(N)
\end{aligned}
$$

where

$$
\begin{aligned}
T_{1,m,N} &:= N^2\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2\left(\frac{1}{N^2}\sum_{ij}X_{ij} - E(X_{12})\right) \\
&\quad + \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right) b_m\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right) + 2\right)\sum_{ij}X_{ij}A_j \\
&= N^2\left(\frac{Nc_{m,N}}{||e_{1,m}||^2}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2\left(\frac{1}{N^2}\sum_{ij}X_{ij} - E(X_{12})\right) \\
&\quad + \left(\left(\frac{Nc_{m,N}}{||e_{1,m}||^2}\sigma_A - 2\right) b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right) + 2\right)\sum_{ij}X_{ij}A_j \\
&= O_p\left(N\sqrt{N}\right)
\end{aligned}
$$

so:

$$
\begin{aligned}
&\frac{1}{N^2}\left(\sum_{i\neq j}X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j, k\neq i,j}\frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i + A_k + V_{ik}\right)\right) \\
&= \left(\frac{c_{m,N}}{E(e_{1,m}^2)}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \eta' E(X_{12}) + \frac{1}{N^2}\eta' T_{1,m,N} \\
&\quad + \left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}\right.\right. \\
&\quad \left.\left. - \frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)\eta' E(X_{12})E(X'_{12})Z_{m,N} \\
&\quad + \left(-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)} - \frac{1}{E(e_{1,m}^2)}\left(-2E(e_{1,m})^2 + \frac{c_{m,N}\sigma_A E(e_{1,m})^2}{E(e_{1,m}^2)}\right)\right)\eta' E(X_{12}X'_{23})Z_{m,N} + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{c_{m,N}}{E(e_{1,m}^2)}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \eta' E(X_{12}) + \frac{1}{N^2}\eta' T_{1,m,N} \\
&\quad + \left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N} E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}\right.\right. \\
&\quad \left.\left. - \frac{2\sigma_A c_{m,N} E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)\eta' E(X_{12}) E(X_{12}') Z_{m,N} \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}\left(1 - \frac{c_{m,N}\sigma_A}{E(e_{1,m}^2)}\right)\eta' E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

Note that

$$
\begin{aligned}
E(e_{1,m})^2 &= \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \\
E(e_{1,m}^2) &= \left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2
\end{aligned}
$$

Hence:

$$
\begin{aligned}
&\frac{1}{N^2}\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j, k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right)\right) \\
&= \left(\frac{c_{m,N}}{E(e_{1,m}^2)}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 \eta' E(X_{12}) + \frac{1}{N^2}\eta' T_{1,m,N} \\
&\quad + \left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N} E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}\right.\right. \\
&\quad \left.\left. - \frac{2\sigma_A c_{m,N} E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)\eta' E(X_{12}) E(X_{12}') Z_{m,N} \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}\eta' E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

Define

$$
\begin{aligned}
&\sum_{i\neq j} X_{ij}' X_{ij} - \sum_{i\neq j, k\neq i,j} \frac{v_{11,i}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||}\frac{v_{11,j}(\hat{\mu}_m)}{||v_{11}(\hat{\mu}_m)||} X_{jk}' X_{ik} \\
&\qquad = N^2\left(E(X_{12}X_{12}') - \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} E(X_{12}X_{23}) + R_{1,m,N} + O_p\left(\frac{1}{N}\right)\right)
\end{aligned}
$$

likewise

$$\sum_{i\neq j} X'_{ij}X_{ij} - \sum_{i\neq j, k\neq i,j} \frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||} X'_{jk}X_{ik}$$
$$= N^2\left(E(X_{12}X'_{12}) - \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b^2_{m,N}\sigma^2_A}E(X_{12}X_{23}) + R_{2,m,N} + O_p\left(\frac{1}{N}\right)\right)$$

Therefore:

$$\mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) > 0)\sqrt{N} diag(0,1,...,1)(\hat{\mu}_{m+1} - \mu_0)$$
$$= \frac{E(e_{1,m})^2}{E(e^2_{1,m})}\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b^2_{m,N}\sigma^2_A}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b^2_{m,N}\sigma^2_A} M^{-1}_{1,m,N}E(X_{12}X'_{23})\sqrt{N}Z_{m,N} + \frac{1}{\sqrt{N}}M^{-1}_{1,m,N}T_{1,m,N}$$
$$- \left(\frac{c_{m,N}}{E(e^2_{1,m})}\sigma_A - 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2\sqrt{N}M^{-1}_{1,m,N}R_{1,m,N}M^{-1}_{1,m,N}E(X_{12}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

with

$$M_{1,m,N} := E(X_{12}X'_{12}) - \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b^2_{m,N}\sigma^2_A}E(X_{12}X_{23})$$

so

$$\mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) > 0)\sqrt{N} diag(0,1,...,1)(\hat{\mu}_{m+1} - \mu_0)$$
$$= \frac{E(e_{1,m})^2}{E(e^2_{1,m})}\frac{\left(\frac{1}{2}b_m c_m\sigma_A - \frac{1}{2b_m}\right)^2 + b^2_m\sigma^2_A}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b^2_m\sigma^2_A} M^{-1}_m E(X_{12}X'_{23})\sqrt{N}Z_{m,N} + \frac{1}{\sqrt{N}}M^{-1}_m T_{1,m,N}$$
$$- \left(\frac{c_m}{E(e^2_{1,m})}\sigma_A - 1\right)\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_{m,N}}\right)^2\sqrt{N}M^{-1}_m R_{1,m,N}M^{-1}_m E(X_{12}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

with

$$M_m := E(X_{12}X'_{12}) - \frac{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b^2_m\sigma^2_A}E(X_{12}X_{23})$$

so that

$$
\begin{aligned}
&\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})>0)\sqrt{N}diag(0,1,...,1)(\hat{\mu}_{m+1}-\mu_0)\\
&= \frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}+\frac{1}{\sqrt{N}}M_m^{-1}T_{1,m,N}\\
&\quad-\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_{m,N}}\right)^2\sqrt{N}M_m^{-1}R_{1,m,N}M_m^{-1}E(X_{12})+O_p\left(\frac{1}{\sqrt{N}}\right.
\end{aligned}
$$

Generally, including the intercept:

$$
\begin{aligned}
&\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})>0)\sqrt{N}(\hat{\mu}_{m+1}-\mu_0)\\
&=-\sigma_A c_{m+1,N}(1,0,...,0)'+\left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\right.\right.\\
&\quad\left.\left.-\frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\right)\right)(1,0,...,0)'E(X_{12}')Z_{m,N}\\
&\quad+\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}+\frac{1}{\sqrt{N}}M_m^{-1}T_{1,m,N}\\
&\quad-\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_{m,N}}\right)^2\sqrt{N}M_m^{-1}R_{1,m,N}M_m^{-1}E(X_{12})+O_p\left(\frac{1}{\sqrt{N}}\right.
\end{aligned}
$$

We treat the term $\sum_{i\neq j}X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)$ in the same way:

$$
\begin{aligned}
&\sum_{i\neq j}X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\\
&=-\frac{1}{||v_{22}(\hat{\mu}_m)||^2}v_{22}(\hat{\mu}_m)'M(\hat{\mu}_m)X_\eta v_{22}(\hat{\mu}_m)+\frac{1}{||v_{22}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l}-\hat{\mu}_{m,l})v_{22}(\hat{\mu}_m)'X_lX_\eta v_{22}(\hat{\mu}_m)\\
&\quad+\sum_{i\neq j}X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)+O_p(N)\\
&=-\frac{\lambda_N(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||^2}v_{22}(\hat{\mu}_m)'X_\eta v_{22}(\hat{\mu}_m)+\frac{1}{||v_{22}(\hat{\mu}_m)||^2}\sum_l(\mu_{0,l}-\hat{\mu}_{m,l})v_{22}(\hat{\mu}_m)'X_lX_\eta v_{22}(\hat{\mu}_m)\\
&\quad+\sum_{i\neq j}X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)+O_p(N)
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{\lambda_{2,m}}{||e_{2,m}||^2} e'_{2,m} X_\eta e_{2,m} + \frac{c_{m,N}}{||v_{22}(\hat{\mu}_m)||^2} \sigma_A v_{22}(\hat{\mu}_m)' \iota \iota' X_\eta v_{22}(\hat{\mu}_m) \\
&\quad + \sum_{i \neq j} X_{ij,\eta} \left(A_i + A_j + V_{ij}\right) - 2(e_{2,m} - v_{22}(\hat{\mu}_m))' X_\eta e_{2,m} - \frac{1}{||v_{22}(\hat{\mu}_m)||^2} \sum_l Z_{m,N,l} e'_{2,m} X_l X_\eta e_{2,m} \\
&\quad + \frac{\lambda_{2,m} - \lambda_N(\hat{\mu}_m)}{||e_{2,m}||^2} e'_{2,m} X_\eta e_{2,m} + O_p(N) \\
&= e'_{2,m} X_\eta e_{2,m} + \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A e'_{2,m} \iota \iota' X_\eta e_{2,m} \\
&\quad + 2 \sum_{i \neq j} X_{ij,\eta} A_i - 2(e_{2,m} - v_{22}(\hat{\mu}_m))' X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_l X_\eta e_{2,m} \\
&\quad + \frac{\lambda_{2,m} - \lambda_N(\hat{\mu}_m)}{||e_{2,m}||^2} e'_{2,m} X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A \Big((e_{2,m} - v_{22}(\hat{\mu}_m))' \iota \iota' X_\eta e_{2,m} + e'_{2,m} \iota \iota' X_\eta (e_{2,m} - v_{22}(\hat{\mu}_m)) \Big) + O_p(N) \\
&= e'_{2,m} X_\eta e_{2,m} + \frac{N c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right) \iota' X_\eta e_{2,m} \\
&\quad + 2 \sum_{i \neq j} X_{ij,\eta} A_i - 2(e_{2,m} - v_{22}(\hat{\mu}_m))' X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_l X_\eta e_{2,m} + b_{m,N} A' \iota \iota' X_\eta e_{2,m} \\
&\quad + \frac{\lambda_{2,m} - \lambda_N(\hat{\mu}_m)}{||e_{2,m}||^2} e'_{2,m} X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A \Big((e_{2,m} - v_{22}(\hat{\mu}_m))' \iota \iota' X_\eta e_{2,m} + e'_{2,m} \iota \iota' X_\eta (e_{2,m} - v_{22}(\hat{\mu}_m)) \Big) + O_p(N) \\
&= e'_{2,m} X_\eta e_{2,m} + \frac{N c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right) \iota' X_\eta e_{2,m} \\
&\quad + 2 \sum_{i \neq j} X_{ij,\eta} A_i + \frac{1}{||e_{2,m}||^4} \sum_l Z_{m,N,l} e'_{2,m} X_l e_{2,m} e'_{2,m} X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_l X_\eta e_{2,m} + b_{m,N} A' \iota \iota' X_\eta e_2, \\
&\quad + \frac{1}{||e_{2,m}||^2} \Bigg(- \sum_l Z_{m,N,l} X_l e_{2,m} + \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_\eta e_{2,m} e_{2,m} + \Bigg(\frac{1}{||e_{2,m}||^2 + ||e_{1,m}||^2} \sum_l Z_{m,N,l} e_{1,m} X_l e_2 \\
&\times \left(-2 X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\iota \iota' X_\eta e_{2,m} + X_\eta \iota \iota' e_{2,m} \right) \right) + O_p(N) \\
&= \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right)^2 \sum_{ij} X_{ij,\eta} + 2 b_{m,N} \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right) \sum_{ij} X_{ij,\eta} A_j \\
&\quad + \frac{N c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right)^2 \sum_{ij} X_{ij,\eta} + \frac{N c_{m,N} b_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A - \frac{1}{2 b_{m,N}} \right) \sum_{ij} X_{ij,\eta} A_j \\
&\quad + 2 \sum_{i \neq j} X_{ij,\eta} A_i + \frac{1}{||e_{2,m}||^4} \sum_l Z_{m,N,l} e'_{2,m} X_l e_{2,m} e'_{2,m} X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_l X_\eta e_{2,m} + b_{m,N} A' \iota \iota' X_\eta e_2, \\
&\quad + \frac{1}{||e_{2,m}||^2} \Bigg(- \sum_l Z_{m,N,l} X_l e_{2,m} + \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e'_{2,m} X_\eta e_{2,m} e_{2,m} + \Bigg(\frac{1}{||e_{2,m}||^2 + ||e_{1,m}||^2} \sum_l Z_{m,N,l} e_{1,m} X_l e_2 \\
&\times \left(-2 X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\iota \iota' X_\eta e_{2,m} + X_\eta \iota \iota' e_{2,m} \right) \right) + O_p(N)
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{Nc_{m,N}}{||e_{2,m}||^2}\sigma_A + 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 \sum_{ij} X_{ij,\eta} \\
&\quad + \left(\left(\frac{Nc_{m,N}}{||e_{2,m}||^2}\sigma_A + 2\right) b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right) + 2\right)\sum_{ij} X_{ij,\eta}A_j \\
&\quad + \frac{1}{||e_{2,m}||^4}\sum_l Z_{m,N,l}e_{2,m}'X_l e_{2,m}e_{2,m}'X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_l X_\eta e_{2,m} + b_{m,N}A'\iota\iota'X_\eta e_{2,m} \\
&\quad + \frac{1}{||e_{2,m}||^2}\Bigg(-\sum_l Z_{m,N,l}X_l e_{2,m} + \frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_\eta e_{2,m}e_{2,m} + \Bigg(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_2 \\
&\quad \times \left(-2X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2}\sigma_A\left(\iota\iota'X_\eta e_{2,m} + X_\eta\iota\iota'e_{2,m}\right)\right) + O_p(N) \\
&= N^2\left(\frac{Nc_{m,N}}{||e_{2,m}||^2}\sigma_A + 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 \eta'E(X_{12}) + \eta'T_{2,m,N} \\
&\quad + \frac{1}{||e_{2,m}||^4}\sum_l Z_{m,N,l}e_{2,m}'X_l e_{2,m}e_{2,m}'X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_l X_\eta e_{2,m} \\
&\quad + \frac{1}{||e_{2,m}||^2}\Bigg(-\sum_l Z_{m,N,l}X_l e_{2,m} + \frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_\eta e_{2,m}e_{2,m} + \Bigg(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_2 \\
&\quad \times \left(-2X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2}\sigma_A\left(\iota\iota'X_\eta e_{2,m} + X_\eta\iota\iota'e_{2,m}\right)\right) + O_p(N)
\end{aligned}
$$

where

$$
\begin{aligned}
T_{2,m,N} :&= N^2\left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right)\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2\left(\frac{1}{N^2}\sum_{ij} X_{ij} - E(X_{12})\right) \\
&\quad + \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right) b_m\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right) + 2\right)\sum_{ij} X_{ij}A_j \\
&= N^2\left(\frac{Nc_{m,N}}{||e_{2,m}||^2}\sigma_A + 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2\left(\frac{1}{N^2}\sum_{ij} X_{ij} - E(X_{12})\right) \\
&\quad + \left(\left(\frac{Nc_{m,N}}{||e_{2,m}||^2}\sigma_A + 2\right) b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right) + 2\right)\sum_{ij} X_{ij}A_j + O_p(N) \\
&= O_p\left(N\sqrt{N}\right)
\end{aligned}
$$

so:

$$\begin{aligned}
&\frac{1}{N^2}\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\right)\\
&=\left(\frac{c_{m,N}}{E(e_{2,m}^2)}\sigma_A+1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2\eta' E(X_{12})+\frac{1}{N^2}\eta' T_{2,m,N}\\
&\quad+\Bigg(\frac{E(e_{2,m})^4}{E(e_{2,m}^2)^2}+\frac{1}{E(e_{2,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{2,m})^2}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^4}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)+E(e_{1,m}^2)}\\
&\quad-\frac{2\sigma_A c_{m,N}E(e_{2,m})^4}{E(e_{2,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)\Bigg)\eta' E(X_{12})E(X'_{12})Z_{m,N}\\
&\quad+\left(-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}+\frac{1}{E(e_{2,m}^2)}\left(2E(e_{2,m})^2+\frac{c_{m,N}\sigma_A E(e_{2,m})^2}{E(e_{2,m}^2)}\right)\right)\eta' E(X_{12}X'_{23})Z_{m,N}+O_p\left(\frac{1}{\sqrt{N}}\right)\\
&=\left(\frac{c_{m,N}}{E(e_{2,m}^2)}\sigma_A+1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2\eta' E(X_{12})+\frac{1}{N^2}\eta' T_{2,m,N}\\
&\quad+\Bigg(\frac{E(e_{2,m})^4}{E(e_{2,m}^2)^2}+\frac{1}{E(e_{2,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{2,m})^2}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^4}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)+E(e_{1,m}^2)}\\
&\quad-\frac{2\sigma_A c_{m,N}E(e_{2,m})^4}{E(e_{2,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)\Bigg)\eta' E(X_{12})E(X'_{12})Z_{m,N}\\
&\quad+\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\left(1+\frac{c_{m,N}\sigma_A}{E(e_{2,m}^2)}\right)\eta' E(X_{12}X'_{23})Z_{m,N}+O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

Note that

$$\begin{aligned}
E(e_{2,m})^2&=\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2\\
E(e_{2,m}^2)&=\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2
\end{aligned}$$

Hence:

$$\begin{aligned}
&\frac{1}{N^2}\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)\right)\\
&=\left(\frac{c_{m,N}}{E(e_{2,m}^2)}\sigma_A+1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2\eta' E(X_{12})+\frac{1}{N^2}\eta' T_{2,m,N}\\
&\quad+\Bigg(\frac{E(e_{2,m})^4}{E(e_{2,m}^2)^2}+\frac{1}{E(e_{2,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{2,m})^2}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^4}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)+E(e_{1,m}^2)}\\
&\quad-\frac{2\sigma_A c_{m,N}E(e_{2,m})^4}{E(e_{2,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)\Bigg)\eta' E(X_{12})E(X'_{12})Z_{m,N}
\end{aligned}$$

$$+\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A+\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}\eta' E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{\sqrt{N}}\right)$$

Define

$$\sum_{i\neq j}X_{ij}'X_{ij}-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk}'X_{ik}$$
$$=N^2\left(E(X_{12}X_{12}')-\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}E(X_{12}X_{23})+R_{2,m,N}\right)$$

where:

$$R_{2,m,N}:=\frac{1}{N^2}\left(\sum_{ij}X_{ij}X_{ij}'-E(X_{ij}X_{ij}')\right)-\left(\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk}'X_{ik}-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12}X_{23})\right)$$

Therefore:

$$\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)\sqrt{N}diag(0,1,...,1)(\hat{\mu}_{m+1}-\mu_0)$$
$$=\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)diag(0,1,...,1)\Bigg($$
$$\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A+\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}M_{2,m,N}^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}+\frac{1}{N\sqrt{N}}M_{2,m,N}^{-1}T_{2,m,N}$$
$$-\left(\frac{c_{m,N}}{E(e_{2,m}^2)}\sigma_A+1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2\sqrt{N}M_{2,m,N}^{-1}R_{2,m,N}M_{2,m,N}^{-1}E(X_{12})+O_p\left(\frac{1}{\sqrt{N}}\right)\Bigg)$$

with

$$M_{2,m,N}:=E(X_{12}X_{12}')-\frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A-\frac{1}{2b_{m,N}}\right)^2+b_{m,N}^2\sigma_A^2}E(X_{12}X_{23})$$

then

$$\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)\sqrt{N}diag(0,1,...,1)(\hat{\mu}_{m+1}-\mu_0)$$

$$
\begin{aligned}
= \quad & \frac{\left(-\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2}{\left(-\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(-\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(-\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N} + \frac{1}{N\sqrt{N}} M_m^{-1} T_{2,m} \\
& - \left(-\frac{c_m}{E(e_{2,m}^2)}\sigma_A + 1\right)\left(-\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 \sqrt{N} M_m^{-1} R_{2,m,N} M_m^{-1} E(X_{12}) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

with

$$
M_m := E(X_{12}X_{12}') - \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} E(X_{12}X_{23})
$$

and

$$
b_m = \left(\frac{1}{4\sigma_A^2 + c_m^2\sigma_A^2}\right)^{\frac{1}{4}}
$$

so

$$
\begin{aligned}
& \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0)\sqrt{N} diag(0,1,...,1)(\hat{\mu}_{m+1} - \mu_0) \\
= \quad & \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N} + \frac{1}{N\sqrt{N}} M_m^{-1} T_{2,m} \\
& - \left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right) \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{N} M_m^{-1} R \\
& + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

Again, including the intercept:

$$
\begin{aligned}
& \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0)\sqrt{N}(\hat{\mu}_{m+1} - \mu_0) = -c_{m+1,N}\sigma_A(1,0,0,...,0)' \\
& + \frac{E(e_{1,m}^2)}{E(e_{1,m}^2) - E(e_{1,m})^2}\left(\frac{E(e_{2,m})^4}{E(e_{2,m}^2)^2} + \frac{1}{E(e_{2,m}^2)}\left(\frac{\sigma_A c_{m,N} E(e_{2,m})^2}{E(e_{2,m}^2)} - 2\frac{E(e_{2,m})^4}{E(e_{2,m}^2)} - 2\frac{E(e_{2,m})^2 E(e_{1,m})^2}{E(e_{2,m}^2) + E(e_{1,m}^2)}\right.\right. \\
& \left.\left. - \frac{2\sigma_A c_{m,N} E(e_{2,m})^4}{E(e_{2,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{2,m})^2 E(e_{1,m})^2}{E(e_{2,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)(1,0,0,...0)' E(X_{12}') Z_{m,N}
\end{aligned}
$$

$$+\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}+\frac{1}{N\sqrt{N}}M_m^{-1}T_{2,m}$$

$$-\left(-\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A+1\right)\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R$$

$$+O_p\left(\frac{1}{\sqrt{N}}\right)$$

In conclusion:

$$\sqrt{N}diag(0,1,...,1)(\hat{\mu}_{m+1}-\mu_0)=diag(0,1,...,1)\times\Bigg(\tag{43}$$

$$\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N} \tag{44}$$

$$+\left(\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)+2\right)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j \tag{45}$$

$$+\left(\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})\geq 0)-\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)\right)\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2\sqrt{N}\left(\frac{1}{N^2}\right. \tag{46}$$

$$-\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})\geq 0)\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M \tag{47}$$

$$-\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)\left(-\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A+1\right)\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{} \tag{48}$$

$$+O_p\left(\frac{1}{\sqrt{N}}\right) \tag{49}$$

With the intercept:

$$\sqrt{N}(\hat{\mu}_{m+1}-\mu_0)=-\sigma_Ac_{m+1,N}(1,0,...,0)'$$
$$+\frac{E(e_{1,m}^2)}{E(e_{1,m}^2)-E(e_{1,m})^2}\left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_Ac_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\right.\right.$$

$$- \frac{2\sigma_A c_{m,N} E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)\Bigg)(1,0,...,0)' E(X_{12}') Z_{m,N}$$

$$\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N}$$

$$+ \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right) b_m \left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right) + 2\right) \frac{1}{N\sqrt{N}} M_m^{-1} \sum_{ij} X_{ij} A_j$$

$$+ \left(\mathbb{1}(\lambda_1(\hat\mu) + \lambda_N(\hat\mu) \geq 0) - \mathbb{1}(\lambda_1(\hat\mu) + \lambda_N(\hat\mu) < 0)\right) \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 \sqrt{N} \left(\frac{1}{N^2}\right.$$

$$- \mathbb{1}(\lambda_1(\hat\mu) + \lambda_N(\hat\mu) \geq 0) \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{N} M$$

$$- \mathbb{1}(\lambda_1(\hat\mu) + \lambda_N(\hat\mu) < 0) \left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right) \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{}$$

$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$

write

$$R_{1,m,N}(1,0,...,0)' = \frac{1}{N^2}\left(\sum_{ij} X_{ij} - E(X_{ij})\right) - \left(\frac{1}{N^2}\sum_{i\neq j, k\neq i,j} \frac{v_{11,i}(\hat\mu_m)}{||v_{11}(\hat\mu_m)||}\frac{v_{11,j}(\hat\mu_m)}{||v_{11}(\hat\mu_m)||} X_{jk}' - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12})\right)$$

and for any $\eta \in \mathbb{R}^L$:

$$\frac{1}{N^2}\sum_{i\neq j, k\neq i,j} \frac{v_{11,i}(\hat\mu_m)}{||v_{11}(\hat\mu_m)||}\frac{v_{11,j}(\hat\mu_m)}{||v_{11}(\hat\mu_m)||} X_{jk,\eta}' - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12,\eta})$$

$$= \frac{1}{N^2||v_{11}(\hat\mu_m)||^2}\left(\sum_{i,k\neq j} v_{11,i}(\hat\mu_m) v_{11,j}(\hat\mu_m) X_{jk,\eta}' - \sum_{j\neq k} v_{11,j}(\hat\mu_m)^2 X_{jk,\eta} - \sum_{j\neq k} v_{11,j}(\hat\mu_m) v_{11,k}(\hat\mu_m) X_{jk,\eta}\right)$$

$$- \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12,\eta})$$

$$= \frac{1}{N^2||v_{11}(\hat\mu_m)||^2}\left(v_{11}(\hat\mu_m)'\iota\right) v_{11}'(\hat\mu_m) X_\eta \iota - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12,\eta}) + O_p\left(\frac{1}{N}\right)$$

$$= \frac{1}{N^2||e_{1,m}||^2}\left(v_{11}(\hat\mu_m)'\iota\right) v_{11}'(\hat\mu_m) X_\eta \iota - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12,\eta}) + O_p\left(\frac{1}{N}\right)$$

$$
\begin{aligned}
&= \frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta}) \\
&\quad + \frac{1}{N^2||e_{1,m}||^2}\left((v_{11}(\hat{\mu}_m)-e_{1,m})'\iota\right)\iota' X_\eta e_{1,m} + \frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta (v_{11}(\hat{\mu}_m)-e_{1,m}) + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta}) \\
&\quad + \frac{1}{N^2||e_{1,m}||^2\lambda_1(\hat{\mu}_m)}\Bigg(- \sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + (\lambda_{1,m}-\lambda_1(\hat{\mu}_m))e_{1,m} \\
&\quad + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg)'\iota\iota' X_\eta e_{1,m} \\
&\quad + \frac{1}{N^2||e_{1,m}||^2\lambda_1(\hat{\mu}_m)}\left(e'_{1,m}\iota\right)\iota' X_\eta\Bigg(- \sum_l Z_{m,N,l}X_l v_{11}(\hat{\mu}_m) + (\lambda_{1,m}-\lambda_1(\hat{\mu}_m))e_{1,m} \\
&\quad + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg) + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta}) \\
&\quad + \frac{1}{N^2||e_{1,m}||^4}\Bigg(- \sum_l Z_{m,N,l}X_l e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m}e_{1,m}e_{1,m} \\
&\quad + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg)'\iota\iota' X_\eta e_{1,m} \\
&\quad + \frac{1}{N^2||e_{1,m}||^4}\left(e'_{1,m}\iota\right)\iota' X_\eta\Bigg(- \sum_l Z_{m,N,l}X_l e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m}e_{1,m} \\
&\quad + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg) + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta}) \\
&\quad + \frac{1}{N^2||e_{1,m}||^4}\Bigg(- \sum_l Z_{m,N,l}X_l e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m}e_{1,m}e_{1,m} \\
&\quad + \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg)'\iota\iota' X_\eta e_{1,m} \\
&\quad + \frac{1}{N^2||e_{1,m}||^4}\left(e'_{1,m}\iota\right)\iota' X_\eta\Bigg(- \sum_l Z_{m,N,l}X_l e_{1,m} - \frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m}e_{1,m}
\end{aligned}
$$

$$+\left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m}\right)e_{2,m}\Bigg)+O_p\left(\frac{1}{N}\right)$$

$$=\frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m}-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta})$$

$$+\frac{E(e_{1,m})}{E(e_{1,m}^2)^2}\Bigg(-E(e_{1,m})E(X_{12})Z_{m,N}-\frac{E(e_{1,m})^3}{E(e_{1,m}^2)}E(X_{12})Z_{m,N}$$

$$+\frac{E(e_{2,m})^2E(e_{1,m})}{E(e_{1,m}^2)+E(e_{2,m}^2)}E(X_{12})Z_{m,N}\Bigg)\eta' E(X_{12})$$

$$+\frac{E(e_{1,m})}{E(e_{1,m}^2)^2}\Bigg(-E(e_{1,m})\eta' E(X_{12}X_{23})Z_{m,N}-\frac{E(e_{1,m})^3}{E(e_{1,m}^2)}\eta' E(X_{12})E(X_{12})'Z_{m,N}$$

$$+\frac{E(e_{2,m})^2E(e_{1,m})}{E(e_{1,m}^2)+E(e_{2,m}^2)}\eta' E(X_{12})E(X_{12})'Z_{m,N}\Bigg)+O_p\left(\frac{1}{N}\right)$$

hence:

$$diag(0,1,1,...,1)M_m^{-1}R_{1,m,N}(1,0,...,0)'=\frac{1}{N^2}diag(0,1,1,...,1)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$

$$-diag(0,1,1,...,1)M_m^{-1}\left(\frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m}-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta})\right)$$

$$+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X'_{23})Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

$$=\frac{1}{N^2}diag(0,1,1,...,1)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$

$$-diag(0,1,1,...,1)M_m^{-1}\frac{1}{N^2||e_{1,m}||^2}\left(e'_{1,m}\iota\right)\iota' X_\eta e_{1,m}$$

$$+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X'_{23})Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

$$=\frac{1}{N^2}diag(0,1,1,...,1)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$

$$-diag(0,1,1,...,1)M_m^{-1}\frac{1}{N^2||e_{1,m}||^2}\Bigg(NE(e_{1,m})^2\sum_{ij}X_{ij,\eta}+E(e_{1,m})b_m\sum_i A_i\sum_{ij}X_{i}$$

$$+NE(e_{1,m})b_m\sum_{ij}A_iX_{ij,\eta}+b_m^2\sum_i A_i\sum_{ij}A_iX_{ij,\eta}\Bigg)$$

$$+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X'_{23})Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

$$
\begin{aligned}
&= \frac{1}{N^2} diag(0,1,1,...,1) M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - diag(0,1,1,...,1) M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \Bigg(N E(e_{1,m})^2 \sum_{ij} X_{ij,\eta} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{i} \\
&\quad + N E(e_{1,m}) b_m \sum_{ij} A_i X_{ij,\eta} \Bigg) \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} diag(0,1,1,...,1) M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} diag(0,1,1,...,1) M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - diag(0,1,1,...,1) M_m^{-1} \frac{E(e_{1,m})}{N^2 E(e_{1,m}^2)} \Bigg(E(e_{1,m}) \sum_{ij} \left(X_{ij,\eta} - E(X_{ij,\eta}) \right) + b_m \sum_{ij} A_i X_{ij,} \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} diag(0,1,1,...,1) M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} diag(0,1,1,...,1) \left(1 - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} \right) M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - diag(0,1,1,...,1) M_m^{-1} \frac{E(e_{1,m}) b_m}{N^2 E(e_{1,m}^2)} \sum_{ij} A_i X_{ij,\eta} \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} diag(0,1,1,...,1) M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right)
\end{aligned}
$$

Including the intercept:

$$
\begin{aligned}
M_m^{-1} R_{1,m,N} (1,0,...,0)' &= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) - M_m^{-1} \left(\frac{1}{N^2 ||e_{1,m}||^2} \left(e_{1,m}' \iota \right) \iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(\right. \\
&\quad + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&\quad + \frac{E(e_{1,m})}{E(e_{1,m}^2)^2} \left(E(e_{1,m}) + 2 \frac{E(e_{1,m})^3}{E(e_{1,m}^2)} - 2 \frac{E(e_{2,m})^2 E(e_{1,m})}{E(e_{1,m}^2) + E(e_{2,m}^2)} \right) M_m^{-1} E(X_{12}) E(X_{12})' Z_{m,N} \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) - M_m^{-1} \left(\frac{1}{N^2 ||e_{1,m}||^2} \left(e_{1,m}' \iota \right) \iota' X_\eta e_{1,m} - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} E(X_{12,\eta}) \right) \\
&\quad + \frac{E(e_{1,m})}{E(e_{1,m}^2) \left(E(e_{1,m}^2) - E(e_{1,m})^2 \right)} \left(E(e_{1,m}) + 2 \frac{E(e_{1,m})^3}{E(e_{1,m}^2)} - 2 \frac{E(e_{2,m})^2 E(e_{1,m})}{E(e_{1,m}^2) + E(e_{2,m}^2)} \right) (1,0,...,0)' E(X_{12})' Z
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} M_m^{-1} E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&- M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \left(N E(e_{1,m})^2 \sum_{ij} X_{ij} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij} \right. \\
&\left. + N E(e_{1,m}) b_m \sum_{ij} A_i X_{ij} + b_m^2 \sum_i A_i \sum_{ij} A_i X_{ij} \right) + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} M_m^{-1} E(X_{12}) \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2) b_m^2 \sigma_A^2} \left(1 + 2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)} - 2\frac{E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)} \right) (1,0,...,0)' E(X_{12})' Z_{m,N} \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} M_m^{-1} E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&- M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \left(N E(e_{1,m})^2 \sum_{ij} X_{ij} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij} \right. \\
&\left. + N E(e_{1,m}) b_m \sum_{ij} A_i X_{ij} \right) \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2) b_m^2 \sigma_A^2} \left(1 + 2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)} - 2\frac{E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)} \right) (1,0,...,0)' E(X_{12})' Z_{m,N} \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} M_m^{-1} E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&- M_m^{-1} \frac{E(e_{1,m})}{N^2 E(e_{1,m}^2)} \left(E(e_{1,m}) \sum_{ij} \left(X_{ij,\eta} - E(X_{ij}) \right) + b_m \sum_{ij} A_i X_{ij} + \frac{1}{N} b_m \sum_i A_i \sum_{ij} X_{ij} \right) \\
&+ \frac{E(e_1^2) - ||e_{1,m}||^2/N}{E(e_1^2)^2} M_m^{-1} E(X_{12}) \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2) b_m^2 \sigma_A^2} \left(1 + 2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)} - 2\frac{E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)} \right) (1,0,...,0)' E(X_{12})' Z_{m,N} \\
&+ \frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2} M_m^{-1} E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N^2} \left(1 - \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} \right) M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right)
\end{aligned}
$$

$$
\begin{aligned}
&- M_m^{-1}\frac{E(e_{1,m})b_m}{N^2E(e_{1,m}^2)}\sum_{ij}A_iX_{ij,\eta} \\
&+\left(\frac{E(e_{1,m}^2)-||e_{1,m}||^2/N}{E(e_{1,m}^2)^2}-\frac{E(e_{1,m})b_m}{E(e_{1,m}^2)^2}\frac{\sum_i A_i}{N}\right)M_m^{-1}E(X_{12}) \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)b_m^2\sigma_A^2}\left(1+2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}-2\frac{E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\right)(1,0,...,0)'E(X_{12})'Z_{m,N} \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right) \\
=&\frac{1}{N^2}\left(1-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}\right)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right) \\
&- M_m^{-1}\frac{E(e_{1,m})b_m}{N^2E(e_{1,m}^2)}\sum_{ij}A_iX_{ij,\eta} \\
&+\left(\frac{E(e_{1,m}^2)-||e_{1,m}||^2/N}{E(e_{1,m}^2)^2}-\frac{E(e_{1,m})b_m}{E(e_{1,m}^2)^2}\frac{\sum_i A_i}{N}\right)M_m^{-1}E(X_{12}) \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)b_m^2\sigma_A^2}\left(1+2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}-2\frac{E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\right)(1,0,...,0)'E(X_{12})'Z_{m,N} \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right) \\
=&\frac{1}{N^2}\left(1-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}\right)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right) \\
&- M_m^{-1}\frac{E(e_{1,m})b_m}{N^2E(e_{1,m}^2)}\sum_{ij}A_iX_{ij,\eta} \\
&+\frac{1}{E(e_{1,m}^2)b_m\sigma_A^2}\left(b_m\left(\sigma_A^2-\frac{\sum A_i^2}{N}\right)-3E(e_{1,m})\frac{\sum_i A_i}{N}\right)(1,0,...,0)' \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)b_m^2\sigma_A^2}\left(1+2\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}-2\frac{E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\right)(1,0,...,0)'E(X_{12})'Z_{m,N} \\
&+\frac{E(e_{1,m})^2}{E(e_{1,m}^2)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)
\end{aligned}
$$

and similarly for $R_{2,m,N}$, write

$$
R_{2,m,N}(1,0,...,0)'=\frac{1}{N^2}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)-\left(\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X_{jk}'-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12})\right.
$$

and for any $\eta \in \mathbb{R}^L$:

$$
\begin{aligned}
&\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}\frac{v_{22,j}(\hat{\mu}_m)}{||v_{22}(\hat{\mu}_m)||}X'_{jk,\eta}-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12,\eta})\\
&=\frac{1}{N^2||e_{2,m}||^2}\left(e'_{2,m}\iota\right)\iota' X_\eta e_{2,m}-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12,\eta})\\
&\quad+\frac{1}{N^2||e_{2,m}||^2\lambda_N(\hat{\mu}_m)}\Bigg(-\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m)+(\lambda_{2,m}-\lambda_N(\hat{\mu}_m))e_{2,m}\\
&\quad+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}\right)e_{1,m}\Bigg)'\iota\iota' X_\eta e_{2,m}\\
&\quad+\frac{1}{N^2||e_{2,m}||^2\lambda_N(\hat{\mu}_m)}\left(e'_{2,m}\iota\right)\iota' X_\eta\Bigg(-\sum_l Z_{m,N,l}X_l v_{22}(\hat{\mu}_m)+(\lambda_{2,m}-\lambda_N(\hat{\mu}_m))e_{2,m}\\
&\quad+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}\right)e_{1,m}\Bigg)+O_p\left(\frac{1}{N}\right)\\
&=\frac{1}{N^2||e_{2,m}||^2}\left(e'_{2,m}\iota\right)\iota' X_\eta e_{2,m}-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12,\eta})\\
&\quad-\frac{1}{N^2||e_{2,m}||^4}\Bigg(-\sum_l Z_{m,N,l}X_l e_{2,m}-\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e'_{2,m}X_l e_{2,m}e_{2,m}\\
&\quad+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}\right)e_{1,m}\Bigg)'\iota\iota' X_\eta e_{2,m}\\
&\quad-\frac{1}{N^2||e_{2,m}||^4}\left(e'_{2,m}\iota\right)\iota' X_\eta\Bigg(-\sum_l Z_{m,N,l}X_l e_{2,m}-\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e'_{2,m}X_l e_{2,m}e_{2,m}\\
&\quad+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_l e_{2,m}\right)e_{1,m}\Bigg)+O_p\left(\frac{1}{N}\right)\\
&=\frac{1}{N^2||e_{2,m}||^2}\left(e'_{2,m}\iota\right)\iota' X_\eta e_{2,m}-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12,\eta})\\
&\quad-\frac{E(e_{2,m})}{E(e_{2,m}^2)^2}\Bigg(-E(e_{2,m})E(X_{12})Z_{m,N}-\frac{E(e_{2,m})^3}{E(e_{2,m}^2)}E(X_{12})Z_{m,N}\\
&\quad+\frac{E(e_{1,m})^2E(e_{2,m})}{E(e_{2,m}^2)+E(e_{1,m}^2)}E(X_{12})Z_{m,N}\Bigg)\eta' E(X_{12})\\
&\quad-\frac{E(e_{2,m})}{E(e_{2,m}^2)^2}\Bigg(-E(e_{2,m})\eta' E(X_{12}X_{23})Z_{m,N}-\frac{E(e_{2,m})^3}{E(e_{2,m}^2)}\eta' E(X_{12})E(X_{12})' Z_{m,N}\\
&\quad+\frac{E(e_{1,m})^2E(e_{2,m})}{E(e_{2,m}^2)+E(e_{1,m}^2)}\eta' E(X_{12})E(X_{12})' Z_{m,N}\Bigg)+O_p\left(\frac{1}{N}\right)
\end{aligned}
$$

hence:

$$diag(0,1,1,...,1)M_m^{-1}R_{2,m,N}(1,0,...,0)' = \frac{1}{N^2}diag(0,1,1,...,1)\left(1-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\right)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$-diag(0,1,1,...,1)M_m^{-1}\frac{E(e_{2,m})b_m}{N^2E(e_{2,m}^2)}\sum_{ij}A_iX_{ij,\eta}$$
$$-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$
$$=\frac{1}{N^2}diag(0,1,1,...,1)\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$+diag(0,1,1,...,1)M_m^{-1}\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)b_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{1}{N^2}\sum_{ij}A_iX_{ij,\eta}$$
$$-\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

Including the intercept:

$$M_m^{-1}R_{2,m,N}(1,0,...,0)' = \frac{1}{N^2}\left(1-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\right)M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$-M_m^{-1}\frac{E(e_{2,m})b_m}{N^2E(e_{2,m}^2)}\sum_{ij}A_iX_{ij,\eta}$$
$$+\frac{E(e_{2,m})}{E(e_{2,m}^2)^2}\left(-E(e_{2,m})-2\frac{E(e_{2,m})^3}{E(e_{2,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})}{E(e_{2,m}^2)+E(e_{1,m}^2)}\right)M_m^{-1}E(X_{12})E(X_{12})'Z_{m,N}$$
$$+\left(\frac{E(e_{2,m}^2)-||e_{2,m}||^2/N}{E(e_{2,m}^2)^2}-\frac{E(e_{2,m})b_m}{E(e_{2,m}^2)^2}\frac{\sum_i A_i}{N}\right)M_m^{-1}E(X_{12})$$
$$-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$
$$=\frac{1}{N^2}\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$

$$
\begin{aligned}
&+ M_m^{-1} \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{1}{N^2} \sum_{ij} A_i X_{ij,\eta} \\
&+ \frac{E(e_{2,m})^2}{E(e_{2,m}^2) b_m^2 \sigma_A^2} \left(-1 - 2 \frac{E(e_{2,m})^2}{E(e_{2,m}^2)} + 2 \frac{E(e_{1,m})^2}{E(e_{2,m}^2) + E(e_{1,m}^2)} \right) (1, 0, ..., 0)' E(X_{12})' Z_{m,N} \\
&+ \frac{1}{E(e_{2,m}^2) b_m \sigma_A^2} \left(b_m \left(\sigma_A^2 - \frac{\sum A_i^2}{N} \right) - 3 E(e_{2,m}) \frac{\sum_i A_i}{N} \right) (1, 0, ..., 0)' \\
&- \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2\right)^2} M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
= &\frac{1}{N^2} \frac{b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&+ M_m^{-1} \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{1}{N^2} \sum_{ij} A_i X_{ij,\eta} \\
&+ \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2\right) b_m^2 \sigma_A^2} \Bigg(-1 - 2 \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} + \\
&\qquad + 2 \frac{\left(\frac{1}{2} b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + \left(\frac{1}{2} b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + 2 b_m^2 \sigma_A^2} \Bigg) (1, 0, ..., 0)' E(X_{12})' Z_{m,N} \\
&+ \frac{1}{E(e_{2,m}^2) b_m \sigma_A^2} \left(b_m \left(\sigma_A^2 - \frac{\sum A_i^2}{N} \right) + 3 \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m} \right) \frac{\sum_i A_i}{N} \right) (1, 0, ..., 0)' \\
&- \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2\right)^2} M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right) \\
= &\frac{1}{N^2} \frac{b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&+ M_m^{-1} \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{1}{N^2} \sum_{ij} A_i X_{ij,\eta}
\end{aligned}
$$

$$+\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)b_m^2\sigma_A^2}\Bigg(-1-2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}+$$

$$+2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2\Bigg)(1,0,...,0)'E(X_{12})'Z_{m,N}$$

$$+\frac{1}{E(e_{2,m}^2)b_m\sigma_A^2}\left(b_m\left(\sigma_A^2-\frac{\sum A_i^2}{N}\right)+3\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)\frac{\sum_i A_i}{N}\right)(1,0,...,0)'$$

$$-\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

where the last inequality results from the observation that:

$$\frac{1}{4}b_m^2c_m^2\sigma_A-\frac{1}{4b_m^2}+b_m^2\sigma_A^2=0$$

remember:

$$diag(0,1,1,...,1)M_m^{-1}R_{1,m,N}(1,0,...,0)'=\frac{1}{N^2}diag(0,1,1,...,1)\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-\right.$$

$$-diag(0,1,1,...,1)M_m^{-1}\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)b_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{1}{N^2}\sum_{ij}A_iX_{ij,\eta}$$

$$+\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right.$$

Including the intercept:

$$M_m^{-1}R_{1,m,N}(1,0,...,0)'=\frac{1}{N^2}\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$

$$
\begin{aligned}
&- M_m^{-1} \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{1}{N^2}\sum_{ij} A_i X_{ij,\eta} \\
&+ \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right) b_m^2\sigma_A^2} \left(1 + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} + \right. \\
&\qquad \left. - 2b_m^2\sigma_A^2 \left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2\right)(1,0,...,0)' E(X_{12})' Z_{m,N} \\
&+ \frac{1}{E(e_{1,m}^2) b_m \sigma_A^2}\left(b_m\left(\sigma_A^2 - \frac{\sum A_i^2}{N}\right) - 3\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)\frac{\sum_i A_i}{N}\right)(1,0,...,0)' \\
&+ \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)^2} M_m^{-1} E(X_{12}X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right)
\end{aligned}
$$

plugging in equation (43):

$$
\begin{aligned}
&\sqrt{N} diag(0,1,...,1)(\hat{\mu}_{m+1} - \mu_0) = diag(0,1,...,1) \times \Bigg(\\
&\quad \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N} \\
&\quad + \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right) b_m\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right) + 2\right)\frac{1}{N\sqrt{N}} M_m^{-1}\sum_{ij} X_{ij}A_j \\
&+ \left(\mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) \geq 0) - \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0)\right)\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right)^2 \sqrt{N}\left(\frac{1}{N^2}\right. \\
&- \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) \geq 0)\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)\frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M \\
&- \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0)\left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right)\frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N} \\
&\quad + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

$$= diag(0,1,...,1) \times \Bigg($$

$$\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N}$$

$$+ \Bigg[\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A - 2 \right) b_m \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m} \right) + 2 +$$

$$+ \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A - 1 \right) \frac{b_m \left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2} \Bigg] \frac{1}{N\sqrt{N}} M_m^{-1} \sum_{ij} X_{ij} A_j$$

$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$

Including the intercept

$$\sqrt{N}(\hat{\mu}_{m+1} - \mu_0) = -\sigma_A c_{m+1,N}(1,0,...,0)'$$

$$+ \frac{E(e_{1,m}^2)}{E(e_{1,m}^2) - E(e_{1,m})^2} \Bigg(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)} \Bigg(\frac{\sigma_A c_{m,N} E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}$$

$$- \frac{2\sigma_A c_{m,N} E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))} \Bigg) \Bigg) (1,0,...,0)' E(X_{12}') Z_{m,N}$$

$$\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N}$$

$$+ \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A - 2 \right) b_m \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m} \right) + 2 \right) \frac{1}{N\sqrt{N}} M_m^{-1} \sum_{ij} X_{ij} A_j$$

$$+ \left(\mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) \geq 0) - \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0)\right) \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A - 1 \right) \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m} \right)^2 \sqrt{N} \Bigg(\frac{1}{N^2}$$

$$- \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) \geq 0) \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A - 1 \right) \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{N} M$$

$$- \mathbb{1}(\lambda_1(\hat{\mu}) + \lambda_N(\hat{\mu}) < 0) \left(- \frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \sigma_A + 1 \right) \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{}$$

$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= -\sigma_A c_{m+1,N}(1,0,...,0)'$$

$$
\begin{aligned}
&+\frac{E(e_{1,m}^2)}{E(e_{1,m}^2)-E(e_{1,m})^2}\Bigg(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}\\
&-\frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)\Bigg)(1,0,...,0)'E(X_{12}')Z_{m,N}\\
&+\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}\\
&+\Bigg[\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)+2+\\
&\quad+\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2}\Bigg]\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j\\
&+\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^4}{b_m^4\sigma_A^4}\times\\
&\left(1+2\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2\right)(1,0,...,0)'\sqrt{N}E(X_{12})'Z_{m,N}\\
&+3\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^3\sigma_A^4}\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'\\
&+\left(\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})\geq 0)-\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)\right)\left(\frac{\sum A_i^2}{N}-\sigma_A^2\right)\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\times\\
&\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^4}\sqrt{N}(1,0,0,...,0)'\\
&+O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

Let's simplify the coefficient of the term $M_m^{-1}E(X_{12})E(X_{12})'Z_{m,N}$:

$$
\begin{aligned}
&\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}-\frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,}}{E(e_{1,m}^2)(E(e_{1,m}^2)+E(e_2}\\
&+\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^4}{b_m^2\sigma_A^2\left(\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b^2\sigma_A^2\right)}\\
&\quad\left(1+2\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)} + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)} + 2b_m^6\sigma_A^6 \\
&\quad - \frac{2\sigma_A c_{m,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} b_m^6 \sigma_A^6}{E(e_{1,m}^2)}\Bigg) \\
&+ \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{b_m^2\sigma_A^2\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b^2\sigma_A^2\right)} \\
&\quad \left(1 + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} - 2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2\right) \\
&= \frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)} + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)} + 2b_m^6\sigma_A^6 \\
&\quad - \frac{2\sigma_A c_{m,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} b_m^6 \sigma_A^6}{E(e_{1,m}^2)}\Bigg) \\
&+ \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{b_m^2\sigma_A^2\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b^2\sigma_A^2\right)} \\
&\quad \left(1 + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} - 2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2\right)
\end{aligned}
$$

To simplify notation, for every c, $c_{1,N}$ and σ_A denote

$$
\begin{aligned}
A(\sigma_A, c, c_{1,N}) := &\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_m c_{1,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)} + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)} + 2b_m^6\sigma_A^6 \\
&- \frac{2\sigma_A c_{m,N}\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} b_m^6 \sigma_A^6}{E(e_{1,m}^2)}\Bigg) \\
&+ \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^4}{b_m^2\sigma_A^2\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b^2\sigma_A^2\right)} \\
&\left(1 + 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} - 2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2\right)
\end{aligned}
$$

$$B(\sigma_A,c) := \left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)+2+$$

$$+\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2}$$

$$C(\sigma_A,c) := 3\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^3\sigma_A^4}$$

$$D(\sigma_A,c,c_{1,N}) := c_{1,N}\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^4}$$

(remember $c_{1,N} := \mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})\geq 0)-\mathbb{1}(\lambda_1(\hat{\mu})+\lambda_N(\hat{\mu})<0)$) so that:

$$\begin{aligned}
\sqrt{N}(\hat{\mu}_{m+1}-\mu_0) &= -\sigma_A c_{m+1,N}(1,0,...,0)'+A(\sigma_A,c_m)M(c_m)^{-1}E(X_{12})E(X_{12}')Z_{m,N}+M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}\\
&+B(\sigma_A,c_m)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j+C(\sigma_A,c_m)\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'+D(\sigma_A,c_m)\sqrt{N}\left(\frac{\sum A_i^2}{N}-\sigma_A^2\right)(1,0,0,...,0)'+O_p\,(\\
&= -\sigma_A c_{m+1,N}(1,0,...,0)'+M(c_m)^{-1}\left(E(X_{12}X_{23}')+A(\sigma_A,c_m)E(X_{12})E(X_{12}')\right)Z_{m,N}\\
&+B(\sigma_A,c_m)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j+C(\sigma_A,c_m)\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'+D(\sigma_A,c_m)\sqrt{N}\left(\frac{\sum A_i^2}{N}-\sigma_A^2\right)(1,0,0,...,0)'+O_p\,(
\end{aligned}$$

Since

$$b_m=\left(\frac{1}{4\sigma_A^2+c_m^2\sigma_A^2}\right)^{\frac{1}{4}}$$

Then:

$$\frac{1}{4}b_m^2\sigma_A^2c_m^2-\frac{1}{4b_m^2}=-b_m^2\sigma_A^2$$

implying

$$\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2\times\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2=\left(\frac{1}{4}b_m^2\sigma_A^2c_m^2-\frac{1}{4b_m^2}\right)^2=b_m^4\sigma_A^4$$

therefore:

$$\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}=1$$

and

$$\sqrt{N}diag(0,1,...,1)(\hat{\mu}_{m+1}-\mu_0) = diag(0,1,...,1)\times\Bigg(M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}$$

$$+2\left(\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)+1\right)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

$$+O_p\left(\frac{1}{\sqrt{N}}\right)$$

$$= diag(0,1,...,1)\times\Bigg(M_m^{-1}E(X_{12}X_{23}')\sqrt{N}diag(0,1,...,1)(\hat{\mu}_m-\mu_0)$$

$$+2\left(\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)+1\right)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

$$+O_p\left(\frac{1}{\sqrt{N}}\right)$$

A.13 Lemma for the proof of theorem 3

Lemma 6.

$$\hat{K}\to_p K_0$$

with

$$K_0 := \frac{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2}{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2+b_0^2\sigma_A^2}\left(E(X_{12}X_{12}')-\frac{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2}{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2+b_0^2\sigma_A^2}E(X_{12}X_{23}')\right)^{-1}$$

$$\times\left(E(X_{12}X_{23}')-\frac{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2}{\left(\frac{1}{2}b_0c_0\sigma_A+\frac{1}{2b_0}\right)^2+b_0^2\sigma_A^2}E(X_{12})E(X_{12}')\right)$$

$$=\frac{1}{2}\left(E(X_{12}X_{12}')-\frac{1}{2}E(X_{12}X_{23}')\right)^{-1}\left(E(X_{12}X_{23}')-\frac{1}{2}E(X_{12})E(X_{12}')\right)$$

Proof. Follows the same proof strategy as the proof of proposition 5 in appendix A.9. □