Title

Essays in Network Econometrics

Author
Sbai Sassi, Yassine

Publication Date
2024

Peer reviewed|Thesis/dissertation

Essays in Network Econometrics

by

Yassine Sbai Sassi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bryan S. Graham, Chair
Professor Michael Jansson
Associate Professor Demian Pouzo

Spring 2024

Essays in Network Econometrics

Abstract

Essays in Network Econometrics

by

Yassine Sbai Sassi

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Bryan S. Graham, Chair

This dissertation studies estimation and inference on models with dyadic dependence, that is models for double indexed observations where observations are correlated whenever they share an index. Data exhibiting this form of dependence are commonplace: from international trade (e.g. Rose (2004)) to sales on online platforms (e.g. Bajari et al. (2023)) or social networks (Fafchamps and Gubert (2007)). Because of the particular dependence structure, very little is known about efficiency in these models. For instance, for parametric models, only a handful of examples have likelihood functions or maximum likelihood estimators that can be expressed in closed form or that are computationally feasible. The analyst is forced to sacrifice efficiency for computational ease and tractability. Unfortunately, unlike cross-sectional models, efficiency losses in dyadic models can manifest as drops in rates of convergence rather than just asymptotic variance, immensely impacting the precision of estimation.

The dissertation explores new estimation methods for different dyadic models, with a particular attention to efficiency and computational feasibility. Each of The three chapters in this dissertation studies a set of dyadic models and estimators for those models. The first and last chapters present efficiency results.

In the first chapter I 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 panel data 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.

Chapter 2 proposes a general procedure to construct estimators for exchangeable network models. For any network model, consider an auxiliary *i.i.d.* model where each observation has the same distribution as any observation in the original model. The procedure returns estimators for the original model whenever valid estimators are known in the auxiliary *i.i.d.* model. The chapter then studies the asymptotic behavior of the *"the average MLE"*, the estimators returned by the procedure for parametric binomial network models. I show that the *average MLE* behaves asymptotically like the composite maximum likelihood estimator. Interestingly, the *average MLE* does not require the entire network to be observed. For instance, I show that for a balanced bipartite graph, observing almost any sub-graph with more than $N^{\frac{3}{2}+\epsilon}$ edges for some $\epsilon > 0$ (out of the total N^2 edges) is enough for the asymptotic result to hold. These results are readily extendable beyond the binomial model.

The final chapter studies the properties of the maximum likelihood estimator (MLE) for exponential families of distributions on network data. I show that, under some conditions, the MLE is asymptotically normally distributed with an asymptotic variance equal to the inverse of the information matrix. I also show that under those same conditions, the MLE is efficient compared to regular estimators with the same rate of convergence. This extends well known results on MLE for *i.i.d.* models.

Contents

Contents **i**

1 A linear regression model for non-oriented dyadic data with interactive individual effects **1**

1.1 Minimizing the least eigenvalues: definitions and main results 3
1.2 Some intuition 10
1.3 Single iteration analysis 13
1.4 Variance estimation and bias correction 22
1.5 The no interaction ($\delta = 0$) specification 24
1.6 Empirical illustration and simulation study 26
1.7 Conclusion 31
1.8 Proofs and intermediary results 31

2 A general estimation procedure for exchangeable random graph models **120**

2.1 Introduction 120
2.2 The model and the main results 121
2.3 The average estimator for networks with missing data 126
2.4 Concluding remarks 128
2.5 Proofs 129

3 Asymptotic efficiency of the maximum likelihood estimator for exponential families **144**

3.1 The model 146
3.2 The asymptotic distribution of the maximum likelihood estimator 148
3.3 Examples 150
3.4 The parametric efficiency of MLE for model (3.3) 158
3.5 An example of model (3.4): a simple binomial 162
3.6 Proofs 168

Bibliography **177**

Acknowledgments

I would like to thank my advisors, Bryan Graham and Michael Jansson, for their invaluable guidance and unwavering support. I also thank Demian Pouzo for the countless hours spent discussing my ideas and questions. I am grateful to my friend and fellow PhD student, Kevin Dano. Kevin's light spirit and intellectual curiosity made the journey all the more enjoyable. I would like to thank my parents, Khalid and Laila, for their unconditional support throughout the years, and my life partner, Khaoula, for her continued encouragement.

Chapter 1

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

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.

[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.

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 $\{1,2\}$, $\{3,4\}$, ..., $\{N-1,N\}$, 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 1.3 discusses the estimator's theoretical properties and numerical implementation. Section 3 proposes estimators for the covariance matrix. Section 1.5 examines the asymptotic distribution of the alternative estimator in a specification without interactive effects. Finally, section 1.6 shows the results from Monte Carlo simulations and from an empirical illustration and the last section concludes. All proofs are deferred to the end of the paper (appendix 1.8).

1.1 Minimizing the least eigenvalues: definitions and main results

Consider the model:

$$Y_{ij} = X'_{ij}\beta_0 + \gamma(A_i + A_j) + \delta(A_i \times A_j) + V_{ij} \tag{1.1}$$

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 $i \neq j$).

When $\delta \neq 0$, the model in equation (1.1) can also be re-expressed:

$$Y_{ij} = \sum_{l=1}^{L} \beta_{0,l} X_{ij,l} - \delta\gamma^2 + \delta(A_i + \gamma)(A_j + \gamma) + V_{ij} \tag{1.2}$$

[2] γ is set to be positive because the model (1.1) 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 (1.1) to that of:

$$Y_{ij} = \sum_{l=1}^{L} \pi_{0,l} X_{ij,l} + \delta U_i U_j + V_{ij} \tag{1.3}$$

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 (1.1) to (1.2) (or (1.3)), only the intercept is altered by the correction term"$-\delta\gamma^2$" in equation (1.2). Therefore, any "good" estimators for the parameters of the model (1.3) also provide good estimators for the parameters in models (1.1) and (1.2), except perhaps for their intercepts. Because the intercept is shifted by $-\delta\gamma^2$ when we move from the original model (1.1) to model (1.3), 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 $l = 1 \dots L$. 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 1.3 in a compact matrix form :

$$Y = \sum_{l=1}^{L} \pi_{0,l} X_l + \delta U U' + V - \delta E(U_1^2) I_N \tag{1.4}$$

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

$$M(\pi) := Y - \sum_{l=1}^{L} \pi_l X_l = \sum_{l=1}^{L} (\pi_{0,l} - \pi_l) X_l + \delta U U' + V - \delta E(U_1^2) I_N \tag{1.5}$$

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

$$g_N(\pi) := \sum_{i=2}^{N} \lambda_i \left(M(\pi)^2 \right) \tag{1.6}$$

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

$$\begin{aligned} g_N(\pi) &= \sum_{i=2}^{N} \lambda_i \left(M(\pi)^2 \right) = Tr\left(M(\pi)^2 \right) - \lambda_1 \left(M(\pi)^2 \right) \\ &= \sum_{i,j} \left(Y_{ij} - \sum_{l=1}^{L} \pi_l X_{ij,l} \right)^2 - \lambda_1 \left(M(\pi)^2 \right) \end{aligned}$$

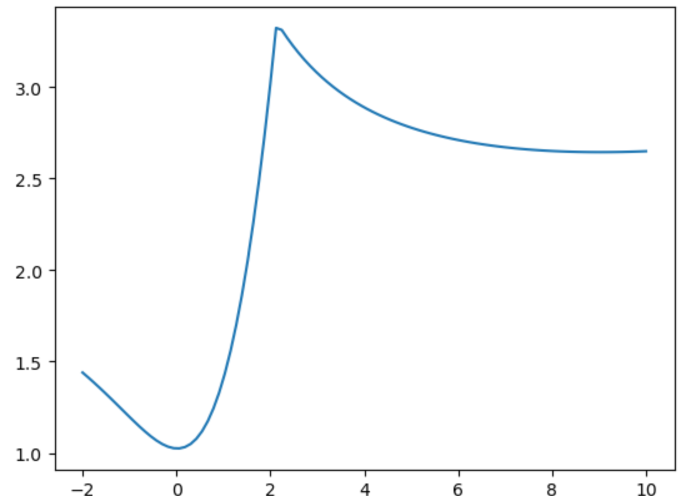

Figure 1.1: *The graph of the function* g_N *for the model* $Y_{ij} = \pi_0 + U_i U_j + V_{ij}$*;* $\sigma_U = \sigma_V = 1$*,* $E(U) = 1$ *,* $N = 100$ *and* $\pi_0 = 1$*. The values of* π *are on the X-axis, and the corresponding* $f_N(\pi)$ *is on the Y-axis.*

and note that the OLS estimator minimizes the sum of squared residuals

$$\sum_{i,j} \left(Y_{ij} - \sum_{l=1}^{L} \pi_l X_{ij,l} \right)^2 .$$

The OLS estimator is efficient when the interactive term $U_i U_j$ is absent from equation (1.3), and as we discuss further later in this paper, the interactive term mostly impacts $M(\pi)$'s largest eigenvalue, once π is *close enough* to the true π_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 1.2 for a detailed discussion).

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 (1.6), 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 (1.1) shows a plot of the function g_N for the model $Y_{ij} = \pi_0 + U_i U_j + V_{ij}$; $\sigma_U = \sigma_V = 1$, $E(U) = 1$, $N = 100$ and $\pi_0 = 1$.

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

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

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

$$\min_{\pi \in \mathbb{R}^L} \sum_{i=2}^{N} \lambda_i \left(M(\pi)^2 \right) \tag{1.7}$$

admits a solution for N large enough. Moreover, π^* is a minimizer of (1.6) if and only if it is a solution to the fixed point problem:

$$\pi = f_N(\pi) \tag{1.8}$$

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

Proof. See section 1.8. □

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 (1.6). Intuitively, equation (1.8) is a first order condition of a minimization problem that is equivalent to (1.6). Let $\pi^* \in \arg\min_{\pi \in \mathbb{R}^L} \sum_{i=2}^{N} \lambda_i \left(M(\pi)\right)^2$ and note

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

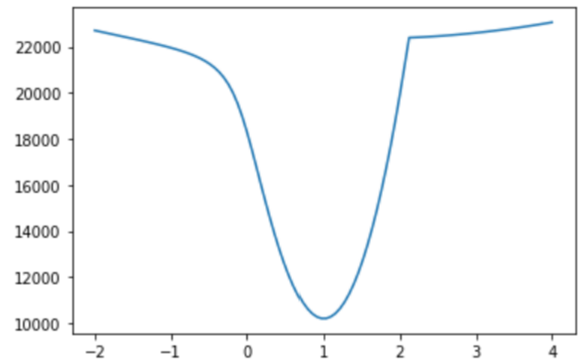

Figure 1.2: *The graph of the function* f_N *for the model* $Y_{ij} = \pi + U_iU_j + V_{ij}$*;* $\sigma_U = \sigma_V = 1$ *,* $N = 100$*. The values of* π *are on the X-axis, and the corresponding* $f_N(\pi)$ *is on the Y-axis.*

The last equality allows for the expression of π^* 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):

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

the first order condition results in the fixed point problem (1.8). The proof in section 1.8 closely follows this sketch.

Lemma 1 does not guarantee the uniqueness of the solution to the minimization problem (1.6). The iteration process just described, when it converges, could converge to one of many potential fixed point of (1.8) (solutions to (1.6)). 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 1.2 illustrates f_N's behavior for the simplest model nested in model (1.3): $Y_{ij} = \pi_0 + U_iU_j + V_{ij}$ for $\sigma_U = \sigma_V = 1$, $\pi_0 = 1$ 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 1.2), that is close to the true parameter $\pi = \pi_0 = 1$. 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 π_0; π_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 (1.6) by studying single successive iterations on the fixed point problem (1.8). 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 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 (1.8), that is, by plugging some "reasonable" initial estimator in the right hand side of (1.8) 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.

Before stating the main result, let's summarize the assumptions we have used in the previous Lemma and intuitions.

Assumption A. - *$X_{ij} = X_{ji} = \phi(X_i, X_j, W_{ij})$ has at least 4 finite moments.*

- *$E(X_{12}X_{12}')$ is full rank.*

Assumption B. - *The V_{ij}'s are i.i.d. across pairs, and have at least 2 finite moments.*

- *The U_i 's are i.i.d. across individuals and have at least 4 finite moments.*

- $Var(U) =: \sigma_U^2 \neq 0$

We will need to introduce a final assumption ensuring that the matrix of covariates does not include the multiplicative individual effects U_i:

Assumption C. *The vector of covariates is not perfectly collinear with the individual errors, that is: for any vector $\lambda \in \mathbb{R}^L$, $\mathbb{P}(\lambda' X_{12} = U_1U_2) < 1$.*

This is a basic identifying assumption, without which the "unobserved" effects term U_iU_j is in fact a linear combination of observable features. We are ready to state our main result:

Theorem 1. *Let the assumptions A, B and C hold. Let $\tilde{\pi}$ be an estimator such that $\tilde{\pi} - \pi_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$. Define the sequence $\hat{\pi}_m$ by: $\hat{\pi}_0 := \tilde{\pi}$ and $\hat{\pi}_{m+1} := f_N(\hat{\pi}_m)$, and let $\hat{\pi}^* := \limsup_m \hat{\pi}_m$. Then with probability approaching 1 $\hat{\pi}^* = \lim_{m\to+\infty} \hat{\pi}_m$ and $\hat{\pi}^*$ is a solution to (1.6). Moreover:*

$$N(\hat{\pi}^* - \pi_0) \to_d \delta\Sigma^{-1}\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2}E(U_1U_2X_{12})\right) + \mathcal{N}\left(0, 2\sigma_V^2\Sigma^{-1}\right) \quad (1.9)$$

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 Corollary 2 and Proposition 4. □

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 is noteworthy. If the individual effects 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 as $\hat{\pi}^*$. 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, $\hat{\pi}^*$ 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" $\hat{\pi}^*$. 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 $\tilde{\pi}$, define the matrix

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

then follow the three steps:

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

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 1.3. We conclude this section by stating the paper's second main theorem:

Theorem 2. *Under the assumptions of theorem 1*

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

Proof. Immediately follows from Corollaries 3 and 4. □

When $\delta = 0$ in model (1.1), however, the iterations become explosive (see proposition 6) and do not converge. Therefore, the limit distribution of * is not well defined. The alternative estimator remains consistent for all slope parameters, excluding the interecept, when the individual effects A in equation (1.1) are independent of the regressors X. When the individual effects and the regressors are correlated, the iterations introduce bias to all parameters, whence the need for the following assumption:

Assumption D. *Assume that for all i, j, the individual effects A_i, A_j are independent from the regressor X_{ij}.*

Theorem 3. *Under the assumptions A and D, and $\delta = 0$, and assuming $\tilde{\pi}$ is $\sqrt{N}$-consistent for π_0. Then:*

$$\begin{aligned}\sqrt{N} diag(0,1,\ldots,1)(\breve{\pi}_2 - \pi_0) &= diag(0,1,\ldots,1)\left(\Gamma_1(c)\sqrt{N}(\tilde{\pi}-\pi_0) + \Gamma_2(c)\frac{\sqrt{N}}{N^2}\sum_{ij} X_{ij}A_j\right) \\ &\quad + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= O_p(1)\end{aligned}$$

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 1.5) and Lemma 5 in Appendix 1.8. □

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 1.2 details how the interactive term in equation (1.3) affects the spectrum of the matrix $M(\pi)$ and why minimizing the function g_N is a sensible choice. Section 1.3 outlines the theoretical results starting from the behavior of a single iteration estimator and leading up to the construction of the alternative estimator $\breve{\pi}_2$.

1.2 Some intuition

Consider the ordinary least squares estimator on model 1.3, defined by

$$\begin{aligned}\hat{\pi}_{OLS} &:= \arg\min_{\pi\in\mathbb{R}^L} \sum_{i,j}\left(Y_{ij} - \sum_{l=1}^{L}\pi_l X_{ij,l}\right)^2 \\ &= \arg\min_{\pi\in\mathbb{R}^L} Tr\left(\left(Y - \sum_{l=1}^{L}\pi_l X_l\right)^2\right) \\ &= \arg\min_{\pi\in\mathbb{R}^L} \sum_{i=1}^{N}\left(\lambda_i\left(Y - \sum_{l=1}^{L}\pi_l X_l\right)^2\right) \\ &=: \arg\min_{\pi\in\mathbb{R}^L} \sum_{i=1}^{N}\lambda_i\left(M(\pi)^2\right)\end{aligned} \tag{1.10}$$

Equation (1.10) indicates that the OLS estimator can also be defined as a minimizer of the average squared eigenvalues of the matrix $M(\pi)$. Let's examine the distribution of $M(\pi)$'s eigenvalues for values of π that are "close" to the true value π_0, assuming $\delta = 1$ (the treatment for $\delta = -1$ is similar). Begin with the value $\pi = \pi_0$, that is, let's look at the distribution of the eigenvalues of the matrix $UU' + V$. [3] Figure 1.3 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 $N = 1000$.

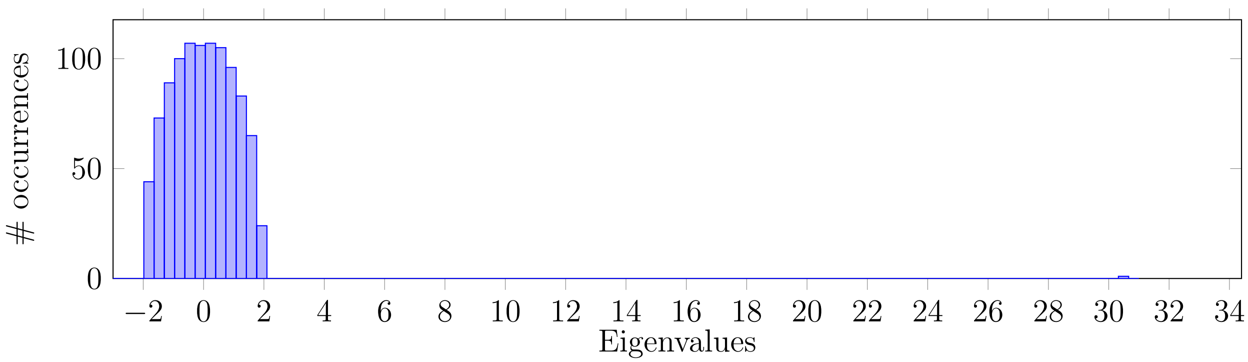

Figure 1.3: Histogram for $\frac{1}{\sqrt{N}}M(\pi_0)$'s eigenvalues; $\sigma_u = \sigma_v = 1$, $N = 1000$

The histogram in figure 1.3 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 ~ 31. 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 1.4 .

To rationalize the shape of Figure 1.3, let's examine the eigenvalues of each of the terms composing $M(\pi_0)$. 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 1.5 shows the histogram of V's eigenvalues. The two histograms in 1.4 and 1.5 are seemingly identical. Only $M(\pi_0)$'s outlier eigenvalue (the one approximately equal to 46) is absent from V's histogram. This should come as no surprise: the matrix $M(\pi_0)$ 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

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

[3] 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.

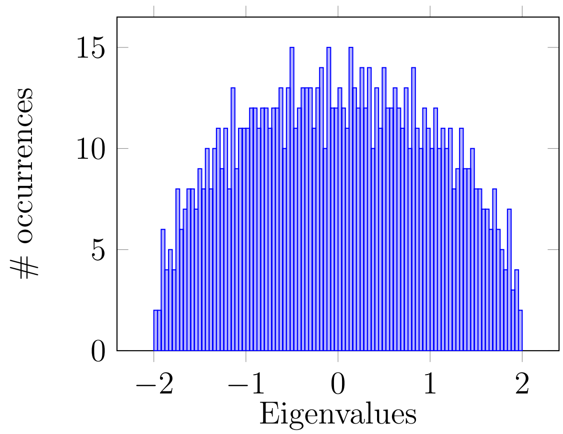

Figure 1.4: *A zoom into the semi circle (the left block in Figure 1.3)*

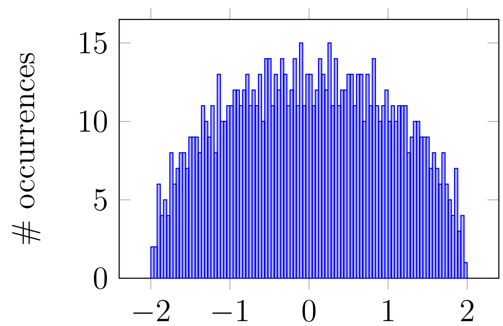

Figure 1.5: *A histogram for* $\frac{1}{\sqrt{N}}V$*'s eigenvalues;* $\sigma_v = 1$, $N = 1000$

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 1.4 and 1.5 are not visually distinguishible.

The semi-circle in figure 1.4 is reminiscent of Wigner's semi-circle law in the random matrix literature (see for instance Benaych-Georges and Knowles (2016)). Wigner'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 1.3: $M(\pi_0)$'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 π that are different from the true parameter π_0. If π is too far from π_0, then the term $\sum_{l=1}^{L}(\pi_{0,l} - \pi_l)X_l$ in equation (1.5) can become dominant and dwarf the contributions of V and UU' in $M(\pi)$'s eigenvalue distribution. In the other extreme, when the candidate π is"very close" to π_0, then the contribution of the covariates' term becomes negligible and we obtain a histogram that is similar to the one in figure 1.3.

The values of π that are abberantly far from π_0 lead to the eigenvalues of $\sum_{l=1}^{L}(\pi_{0,l} - \pi_l)X_l$ 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 1.3. However, this rough discrimination strategy will be ineffective for values of π that return a term $\sum_{l=1}^{L}(\pi_{0,l} - \pi_l)X_l$ of order $\sqrt{N}$ or lower. In any case, for model (1.3), 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)).[4] The following lemma shows that any $\sqrt{N}$ estimator is in fact "close enough" for our purposes.

Lemma 2. *Assume that assumption A holds. We have that*

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

Proof. Refer to subsection 1.8. □

The lemma implies that any initial estimator $\tilde{\pi}$ that is $\sqrt{N}$ consistent - like the OLS estimator - would yield a covariates' term such that $\sum_{l=1}^{L}(\pi_{0,l} - \tilde{\pi}_l)X_l = O_p(\sqrt{N})$. It would produce an eigenvalue histogram for $M(\tilde{\pi})$ that is similar to figure 1.3 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 π is close enough to π_0, the largest eigenvalue of $M(\pi)$ is, at least up to a first order approximation, closely tied to the error term UU'. Notice that, absent the UU' from the model 1.3 (or the random effects A_i and A_j from the model (1.1)), 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 (1.1) or (1.3).

An appealing idea is then to modify the objective function in the matrix form definition of the OLS in (1.10) to remove the contribution of the random effects. Following the intuition laid down so far, this can for instance be done by removing $M(\pi)$'s largest eigenvalue from the sum of squared errors before minimizing. The new estimator would be a solution to the minimization problem

$$\min_{\pi \in \mathbb{R}^L} \sum_{i=2}^{N} \lambda_i \left(M(\pi)^2\right) = g_N(\pi) \tag{1.11}$$

1.3 Single iteration analysis

The first result examines a single iteration of the fixed point problem (1.8).

[4] In fact, if $E(U_1) \neq 0$, the OLS estimator of the intercept is biased. In proposition 7 (appendix 1.8), we show how that bias can be corrected to obtain a $\sqrt{N}$-consistent estimator.

Proposition 1. *Consider the model 1.3:*

$$Y_{ij} = \sum_{k=1}^{K} \pi_{0,k} X_{ij,k} + \delta U_i U_j + V_{ij} = X_{ij}\pi_0 + \delta U_i U_j + V_{ij}$$

where $\delta \in \{-1, +1\}$*. Under the assumptions A, B and C, given a first stage estimator* $\tilde{\pi}$ *such that* $||\tilde{\pi} - \pi_0|| = O_p\left(\frac{1}{\sqrt{N}}\right)$*, the single iteration estimator*

$$\hat{\pi} := \left(\sum_{i \neq j} X'_{ij} X_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\pi}) \nu_j(\tilde{\pi}) X'_{jk} X_{ik} \right)^{-1} \left(\sum_{i \neq j} X'_{ij} Y_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\pi}) \nu_j(\tilde{\pi}) X'_{jk} Y_{ik} \right) \tag{1.12}$$

satisfies

$$\sqrt{N}(\hat{\pi} - \pi_0) = K\sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{1.13}$$

for

$$K := \frac{1}{E(U_1^2)} \left(E(X'_{12}X_{12}) - \frac{1}{E(U_1^2)} E(U_1 U_3 X'_{12} X_{32}) \right)^{-1} \times \left(E(U_1 U_3 X_{12} X_{23}) - \frac{1}{E(U_1^2)} E(U_1 X_{12} U_2) E(U_1 U_2 X_{12}) \right)$$

A detailed proof is presented in section 1.8. Proposition 3 shows that

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

is invertible, ensuring that K is well defined.

Equation (1.13) 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 $\hat{\pi}$ is consistent and converges to π_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 $\tilde{\pi}$.

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 (1.13) becomes

$$\hat{\pi} - \pi_0 = O_p\left(\frac{1}{N}\right).$$

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 $\hat{\pi}$ 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 (1.13) and determine how it depends on the first stage estimator $\tilde{\pi}$ and/or how it behaves asymptotically. The next proposition and its proof in appendix 1.8 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*

$$N(\hat{\pi} - \pi_0) \to_d \mathcal{N}\left(0, 2\sigma_v^2 E(X_{12}X_{12}')^{-1}\right) \tag{1.14}$$

A "one step theorem" applies, one iteration is enough to achieve full efficiency. The argument proving the efficiency of $\hat{\pi}$ in proposition 2 is simple: consider the alternative model $Y_{ij} = \sum_{l=1}^{L} \pi_{0,l} X_{ij,l} + V_{ij} = X_{ij}\pi_0 + V_{ij}$ 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 (1.14) (see for instance Chamberlain (1987a) or Newey (1990)). Given that our model of interest (1.3) is noisier than the alternative model, the following corollary holds.

Corollary 1. *Under the assumptions of proposition 2, the single iteration estimator defined in (1.13) 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 $|K| < 1$, $\hat{\pi}$ is to a first order closer to π_0 than $\tilde{\pi}$. 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 $\hat{\pi}$ 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 positive definite and all the eigenvalues of the matrix

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

are positive and strictly smaller than 1.

Proof. cf. section 1.8 □

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 (1.12) 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 1 and 3 provide a rate N (a rate optimal) estimator.[5]

Another corollary of proposition 1 is that, if f_N has a fixed point $\hat{\pi}^*$ that is $\sqrt{N}$-consistent, then equation (1.13) yields:

$$(I - K)\sqrt{N}(\hat{\pi}^* - \pi_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

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

To establish the existence of a fixed point, notice that equation (1.13) 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 $\tilde{\pi}$ be an estimator such that $\tilde{\pi} - \pi_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$. Fix $\kappa \in (\lambda_1(K), 1)$ and some $C > 0$. Under the assumptions of proposition 1, with probability approaching 1:*

1. *The function f_N in equation (1.8) is differentiable in the closed ball $B(\pi_0, \frac{C}{\sqrt{N}})$ centered at π_0 and with radius $\frac{C}{\sqrt{N}}$.*
2. $\sup_{\pi \in B(\pi_0, \frac{C}{\sqrt{N}})} ||f_N'(\pi)|| \leq \kappa < 1$

[5] 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 $\sqrt{N}(\tilde{\pi} - \pi_0)$ 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.

Moreover, define the sequence $\hat{\pi}_m$ *by:* $\hat{\pi}_0 := \tilde{\pi}$ *and* $\hat{\pi}_{m+1} := f_N(\hat{\pi}_m)$*, and* $\hat{\pi}^* := \limsup_m \hat{\pi}_m$*. Then* $\hat{\pi}^* - \pi_0 = O_p\left(\frac{1}{\sqrt{N}}\right)$ *and with probability approaching 1* $\hat{\pi}^* = \lim_{m\to+\infty} \hat{\pi}_m$ *and* $\hat{\pi}^*$ *is a solution to (1.6).*

Proof. Cf. Section 1.8 . □

So $\hat{\pi}^*$ 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 (1.13) 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* $\tilde{\pi} - \pi_0 = O_p\left(\frac{1}{N}\right)$*, then*

$$N(\hat{\pi} - \pi_0) = KN(\tilde{\pi} - \pi_0) + R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{1.15}$$

with

$$R_N \to_d \left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32}')\right)^{-1}\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2}E(U_1U_2X_{12})\right) + \left(E(X_{12}X_{12}') - \frac{E(U_1)^2}{E(U_1^2)}E(X_{12}X_{32}')\right)^{-1}\mathcal{N}\left(0, 2\sigma_V^2\Sigma\right)$$

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 1.8 □

Because of the presence of the residual R_N in equation (1.15), the new expansion is fundamentally different from the previous one (equation (1.13) in Proposition 1). The effect of an iteration on the estimation quality is now ambiguous and depends on how the first stage estimator $\tilde{\pi}$ 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 $\tilde{\pi}$ and R_N, which is challenging even for a single iteration. However, we can see that because K has a smaller than one spectral radius (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 $\tilde{\pi}(=: \hat{\pi}_0)$, from (1.15):

$$N(\hat{\pi}_m - \pi_0) \approx K^m N(\tilde{\pi} - \pi_0) + \sum_{i=0}^{m-1} K^i R_N$$

$$\approx K^m N(\tilde{\pi} - \pi_0) + (I_L - K^m)(I_L - K)^{-1} R_N$$
$$\approx (I_L - K)^{-1} R_N; \text{ when } m \text{ is lage.}$$

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

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

$$N(\hat{\pi}^* - \pi_0) = (I - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and with probability approaching 1 $\hat{\pi}^*$ is a solution to (1.6). Therefore

$$N(\hat{\pi}^* - \pi_0) \to_d \delta \Sigma^{-1}\left(\frac{3}{E(U_1^2)} E(U_1^3 X_{12} U_2) - \frac{E(U_1^4)}{E(U_1^2)^2} E(U_1 U_2 X_{12})\right) + \mathcal{N}\left(0, 2\sigma_V^2 \Sigma^{-1}\right) \tag{1.16}$$

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 Proposition 4 and Corollary 2. □

Notice that $\hat{\pi}^*$ 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)...). Interestingly, the asymptotic bias of $\hat{\pi}^*$ has a particular structure when the individual effects are independent of the observable regressors. Notice that, in that case:

$$\Sigma(1, 0, ..., 0)' = \frac{Var(U_1)^2}{E(U_1^2)} E(X_{12})$$

so that:

$$\Sigma^{-1}\left(\frac{3}{E(U_1^2)} E(U_1^3 X_{12} U_2) - \frac{E(U_1^4)}{E(U_1^2)^2} E(U_1 U_2 X_{12})\right)$$
$$= \frac{E(U_1)}{Var(U_1)^2}\left(3E(U_1^2)E(U_1^3) - E(U_1)E(U_1^4)\right)(1, \ldots, 0)'$$

that is, the bias only affects the intercept. So, when the individual effects are independent of the regressors, and if the slope parameters are the only parameters of interest, no bias correction is needed.

In general however, all coefficients are affected by the asymptotic bias. In the next section, we offer a correction to this bias by proposing a consistent estimator for the bias term.

The corollary shows that if we initiate a sequence $\hat{\pi}_0 := \tilde{\pi}$, for some initial $\sqrt{N}$-consistent estimator $\tilde{\pi}$, and then we iterate "infinitely many" $\hat{\pi}_{m+1} := f_N(\hat{\pi}_m)$ as in corollary 2, then with high probability $\hat{\pi}_m$ approaches a fixed point $\hat{\pi}^*$. 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 (1.13) 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 (1.35), 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 $\hat{\pi}^*$ 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 $\hat{\pi}^*$. The alternative estimator requires exactly 2 iterations over the function f_N and is therefore computationally 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 $\tilde{\pi}$ be an initial $\sqrt{N}$ consistent estimator and let $\hat{\pi}_1$ be the estimator returned in the equation (1.12) after a single iteration. Write $\breve{\pi}_1^* := G\hat{\pi}_1 + (I_L - G)\tilde{\pi}$, 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 $\breve{\pi}_1^*$ converges at rate N. Write

$$\begin{aligned}
\sqrt{N}(\breve{\pi}_1^* - \pi_0) &= \sqrt{N}G(\hat{\pi}_1 - \pi_0) + \sqrt{N}(I_L - G)(\tilde{\pi} - \pi_0) \\
&= (GK + I_L - G)\sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= \big(I_L - G(I_L - K)\big)\sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

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 $\breve{\pi}_1 := \hat{G}\hat{\pi} + (I_L - \hat{G})\tilde{\pi}$. As for $\breve{\pi}^*$

$$\begin{aligned}\sqrt{N}(\breve{\pi}_1 - \pi_0) &= \sqrt{N}\hat{G}(\hat{\pi} - \pi_0) + \sqrt{N}(I_L - \hat{G})(\tilde{\pi} - \pi_0) \\ &= \left(I_L - (I_L - \hat{K})^{-1}(I_L - K)\right)\sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= (I_L - \hat{K})^{-1}\left(K - \hat{K}\right)\sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned} \tag{1.17}$$

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 $\breve{\pi}$ is rate optimal. The following proposition offers an example of a $\sqrt{N}$ consistent estimator for K.

Proposition 5. *Let the assumptions of Proposition 1 hold. Let $\tilde{\pi}$ be a $\sqrt{N}$-consistent estimator for π_0. Define:*

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

Then

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

Proof. c.f. section 1.8 □

Proposition 5 allows for the construction of an estimator that is rate optimal. However, studying the asymptotic distribution of $\breve{\pi}_1$ defined in (1.17) is challenging. It requires that we determine the joint asymptotic distribution of $\hat{K}$, $\sqrt{N}(\tilde{\pi} - \pi_0)$ and the residual of order $O_p\left(\frac{1}{\sqrt{N}}\right)$ in equation (1.17). However, as for the study of the fixed point $\hat{\pi}^*$, 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 $\hat{\pi}_2 := f_N(\breve{\pi}_1)$ and define $\breve{\pi}_2 := \hat{G}\hat{\pi}_2 + (I_L - \hat{G})\breve{\pi}_1$. Following the steps in equation (1.17),

$$\begin{aligned} N(\breve{\pi}_2 - \pi_0) &= (I_L - \hat{K})^{-1}\left(K - \hat{K}\right) N(\breve{\pi}_1 - \pi_0) + (I_L - \hat{K})^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= (I_L - K)^{-1} R_N + (I_L - \hat{K})^{-1}\left(K - \hat{K}\right) N(\breve{\pi}_1 - \pi_0) + (K - \hat{K})^{-1} R_N \\ &\quad + O_p\left(\frac{1}{\sqrt{N}}\right) \\ &= (I_L - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \end{aligned} \tag{1.18}$$

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

To summarize, the alternative estimation procedure follows these steps:

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

Corollary 4.

$$N(\breve{\pi}_2 - \pi_0) = (I_L - K)^{-1} R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \tag{1.19}$$

Proof. See the steps leading to equation (1.17). □

We conclude this section by highlighting important connections to the literature on the large N large T panel regression models. 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 (1.6) 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 $M(\pi)$ in (1.5).[6] 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 $\hat{\pi}^*$'s asymptotic variance and that we don't see

[6]The LS objective function: $S_N(\pi, U) := \frac{1}{N(N-1)}\sum_{i\neq j}(Y_{ij} - \sum_k \pi_k X_{ij,k} - U_i U_j)^2 = \frac{1}{N(N-1)}\sum_{i,j}(Y_{ij} - \sum_k \pi_k X_{ij,k} - U_i U_j)^2 - \frac{\sum_i U_i^4}{N(N-1)}$, and Moon and Weidner (2015) show that $\frac{g_N(\pi)}{N(N-1)} = arg\min_U \frac{1}{N(N-1)}\sum_{i,j}(Y_{ij} - \sum_k \pi_k X_{ij,k} - U_i U_j)^2$. 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.

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.

1.4 Variance estimation and bias correction

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) and a consistent estimator for the bias term: $\frac{3}{E(U_1^2)}E(U_1^3X_{12}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2}E(U_1U_2X_{12})$ (equation (1.9)).

The matrice $E(X_{12}X_{12}')$ the vector $E(X_{12})$ can be estimated through their sample analogues. δ, $E(U_1^2)$, $E(U_1^4)$, $E(U_1U_3X_{12}X_{32}')$, $E(U_1^3U_2X_{12})$ and $E(U_1U_2X_{12})$ are left to be estimated. Assume that $\delta = 1$, section 1.2 explained how the eigenvector corresponding to the largest eigenvalue of $M(\tilde{\pi})$ 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* $\hat{U}_i = \sqrt{\max_i |\lambda_i(\tilde{\pi})|}\nu_i(\tilde{\pi})$. *We have*

1. $\sum_i |\nu(\tilde{\pi})|^r = \frac{\sum_i |U_i|^r}{\max_i |\lambda_i(\tilde{\pi})|^{\frac{r}{2}}} + O_p\left(\frac{1}{N^{1/r}}\right)$, *for all* $r \geq 2$
2. $\frac{\sum_i U_i^2 - \hat{U}_i^2}{N} = O_p\left(\frac{1}{N}\right)$,
3. $\frac{1}{N^3}\sum_{i\neq j, k\neq i,j} \hat{U}_i\hat{U}_k X_{ij}X_{jk}' \to_p E(U_1U_3X_{12}X_{32}')$
4. $\frac{1}{N^2}\sum_{i\neq j} \hat{U}_i\hat{U}_j X_{ij} \to_p E(U_1U_2X_{12})$
5. $\frac{1}{N^2}\sum_{i\neq j} X_{ij}X_{ij}' \to_p E(X_{12}X_{12}')$
6. $\sum_{i,j} \nu_i(\tilde{\pi})^3 X_{ij}\nu_j(\tilde{\pi}) \to_p \frac{1}{E(U_1^2)^2}E(U_1^3X_{12}U_2)$

Proof. Cf Appendix 1.8. □

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 (1.2)

$$Y_{ij} = \sum_{l=1}^{L} \pi_{0,l}X_{ij,l} + \delta U_iU_j + V_{ij}$$

First, observe that

$$\frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij}-X_{ij}\breve{\pi}_2\right)^2 = \frac{1}{N^2}\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right)^2 + (\breve{\pi}_2-\pi_0)'\left(\sum_{i\neq j}X_{ij}'X_{ij}\right)(\breve{\pi}_2-\pi_0)$$
$$-2\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right)X_{ij}(\breve{\pi}_2-\pi_0)$$
$$=\frac{1}{N^2}\sum_{i\neq j}\left(\delta U_i U_j + V_{ij}\right)^2 + O_p\left(\frac{1}{N}\right)$$
$$=E(U_1^2)^2+\sigma_V^2+O_p\left(\frac{1}{\sqrt{N}}\right)$$

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

Corollary 5. *Define* $\hat{\sigma}_V^2 := \frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij}-X_{ij}\breve{\pi}_2\right)^2 - \left(\frac{\sum_i \hat{U}_i^2}{N}\right)^2$ *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 $\frac{1}{N^2}\sum_{i\neq j}\left(Y_{ij}-X_{ij}\breve{\pi}_2\right)^2 \to_p E(U_1^2)^2+\sigma_V^2$ and the convergence of $\frac{\hat{U}_i^2}{N}$ to $E(U_1^2)$ (Lemma 3). □

Therefore

$$2\hat{\sigma}_V^2\hat{\Sigma} := 2\hat{\sigma}_V^2(I-\hat{K})^{-1}$$
$$\times\left(\sum_{i\neq j}\nu_i(\tilde{\mu})X_{i,j}X_{j,k}'\nu_k(\tilde{\mu}) - \left(\sum_{i\neq j}\nu_i(\tilde{\mu})X_{i,j}\nu_j(\tilde{\mu})\right)\left(\sum_{i\neq j}\nu_i(\tilde{\mu})X_{i,j}\nu_j(\tilde{\mu})\right)'\right)^{-1}$$

is consistent for the asymptotic variance. For the bias term, define:

$$\hat{\delta} := \begin{cases} sign(\lambda_1(\tilde{\mu})) & \text{if } \lambda_1(\tilde{\mu}) := \max_i|\lambda_i(\tilde{\mu})| \\ sign(\lambda_N(\tilde{\mu})) & \text{if } |\lambda_N(\tilde{\mu})| = \max_i|\lambda_i(\tilde{\mu})| \end{cases}$$

In appendix 1.8 we show that $\hat{\delta}$ is consistent for δ. Therefore

$$\hat{\delta}\hat{\Sigma}^{-1}\max_i|\lambda(\tilde{\pi})|\left(\sum_{i,j}3\nu_i(\tilde{\pi})^3X_{ij}\nu_j(\tilde{\pi}) - \sum_i\nu_i(\tilde{\pi})^4\sum_{i,j}\nu_i(\tilde{\pi})_{ij}^{X}\nu_j(\tilde{\pi})\right)$$

is consistent for the bias term.

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

For this section only, assume that the individual effects are independent of the regressors. In model (1.1), when $\delta = 0$, the model becomes:

$$Y_{ij} = \sum_l X_{ij,l}\pi_l + A_i + A_j + V_{ij} \tag{1.20}$$

or, in matrix form:

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

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

$$\hat{\pi}_m - \pi_0 = -\sigma_A c_{m,N} \times (1, 0, 0, ..., 0)' + Z_{m,N} + O_p\left(\frac{1}{N}\right) \tag{1.21}$$

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 $Z_{0,N} = \tilde{\pi} - \pi_0$. The reason we allow for a bias in the intercept in equation (1.21) is that proposition 6 shows that even when we initiate with a consistent estimator under model (1.20), iterations introduce bias in the intercept.

Then

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

to be able to employ the ideas from the pure interaction model, we orthogonalize the symmetric matrix $\sigma_A c_{m,N}\iota\iota' + A\iota' + \iota A'$, to show that it is in fact a rank two matrix. Note that:[7],[8]

$$\sigma_A c_{m,N} \iota\iota' + A\iota' + \iota A' = e_{1,m}e'_{1,m} - e_{2,m}e'_{2,m} \tag{1.22}$$

where:

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

[7] See section 1.8 for a proof.

[8] the dependence on N is omitted in the notation for $e_{1,m}$ and $e_{2,m}$.

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

moreover:

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

Therefore, the error component of $M(\hat{\pi}_m)$ 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. Extending our past intuitions, by removing a single largest eigenvalue, one other eigenvalue, of a similar order of magnitude ($\sim N$) is left. We can state a result equivalent to Proposition 1 in the $\delta \neq 0$ specification:

Proposition 6. *Under the specification (1.20), assume iterations are initiated with an estimator* $\tilde{\pi} = \pi_0 - c\sigma_A(1, 0, 0, ..., 0)' + Z_N + O_p\left(\frac{1}{N}\right)$, *for* $Z_N = O_p\left(\frac{1}{\sqrt{N}}\right)$, *if* $\hat{\pi} = f_N(\tilde{\pi})$, *then:*

$$\begin{aligned}
\sqrt{N}(\hat{\pi} - \pi_0) = &-\sigma_A \left(h(c) + O_p\left(\frac{1}{\sqrt{N}}\right)\right)(1, 0, ..., 0)' \\
&+ M(c)^{-1}\left(E(X_{12}X_{23}') + A(c, c_{1,N})E(X_{12})E(X_{12}')\right)\sqrt{N}Z_{m,N} \\
&+ B(c, c_{1,N})\frac{1}{N\sqrt{N}}M(c)^{-1}\sum_{ij} X_{ij}A_j + O_p\left(\frac{1}{\sqrt{N}}\right) \\
= &-\sigma_A \left(h(c) + O_p\left(\frac{1}{\sqrt{N}}\right)\right)(1, 0, ..., 0)' + O_p(1)
\end{aligned}$$

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 1.8. □

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.

1.6 Empirical illustration and simulation study

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 1.2) 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 1.1). Comparing the two tables, as expected, the standard errors are lower for the least eigenvalues estimator than for the OLS.

Table 1.1: 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
Both in WTO	-0.479	0.072
One in WTO	-0.322	0.070
GSP	0.305	0.034
Log distance	-1.181	0.019
Log product real GDP	0.829	0.010
Log product real GDP p/c	-0.033	0.012
Regional FTA	0.679	0.081
Currency union	0.592	0.141
Common language	0.369	0.041
Land border	0.793	0.083
Number landlocked	-0.375	0.029
Number islands	0.017	0.036
Log product land area	-0.071	0.008
Common colonizer	0.863	0.055
Ever colony	1.246	0.105

Table 1.2: 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
Both in WTO	-0.269	0.096
One in WTO	-0.320	0.097
GSP	0.199	0.045
Log distance	-1.073	0.025
Log product real GDP	0.944	0.011
Log product real GDP p/c	-0.034	0.014
Regional FTA	0.946	0.108
Currency union	0.757	0.194
Common language	0.415	0.053
Land border	0.965	0.114
Number landlocked	-0.540	0.034
Number islands	0.022	0.041
Log product land area	-0.076	0.009
Common colonizer	0.966	0.074
Ever colony	1.113	0.141

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 $\gamma = 1$

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

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

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

for each of the two designs $X \sim Unif(0,1)$, $\beta_{0,1} = \beta_{0,2} = E(A_1^2) = E(V_{12}^2) = 1$.[9] The histograms for the estimated slope parameters $\beta_{0,2}$ are in the figures 1.6 to 1.9. In each graph, we show the histogram for the OLS estimator (in bleu) on the original model (1.1) as a benchmark, the estimator $\hat{\pi}_{EIG}$ 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 1.6 to 1.9 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 1.6 and 1.7) confirm the result in proposition 2. The histogram for the eigenvalue-corrected estimator (in green) is close to the oracle (orange). On 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 1.6 and 1.7, because the variance of A is not large enough (see figure 1.10 for a version of figure 1.7 with a $Var(A) = 100$ and where the skew is now obvious on the OLS estimator, whereas the eigenvalue corrected estimator is unaffected).

Figures 1.8 and 1.9 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.

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

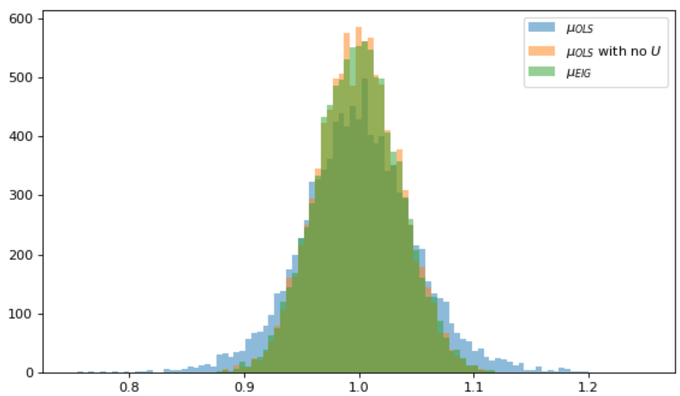

Figure 1.6: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_i A_j + V_{ij}$, and the "oracle" OLS estimator (orange) for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$.*

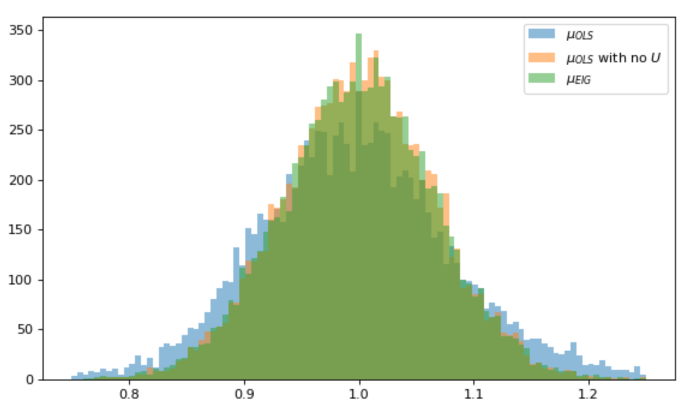

Figure 1.7: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2} X_i X_j + A_i A_j + V_{ij}$, and the "oracle" OLS estimator (orange) for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$.*

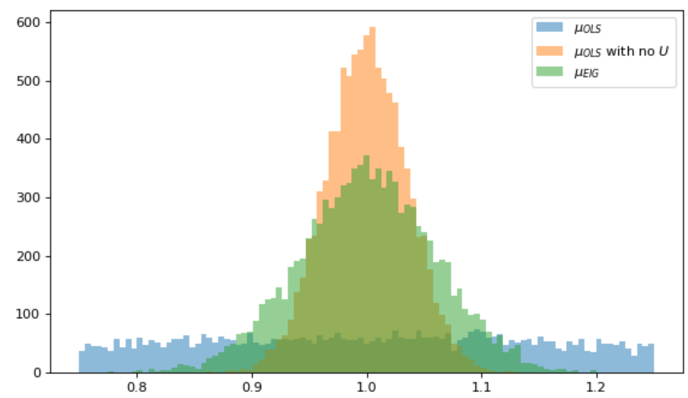

Figure 1.8: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + A_i + A_j + A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*.*

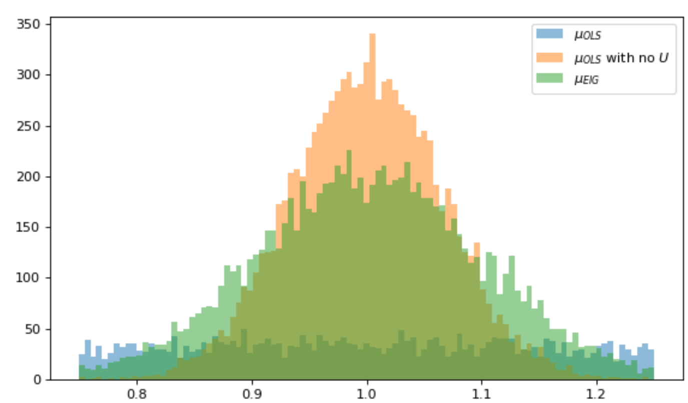

Figure 1.9: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + A_i + A_j + A_iA_j + V_{ij}$*, and the "oracle" OLS estimator (orange) for the slope parameter* $\beta_{0,1}$ *in the model* $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$*.*

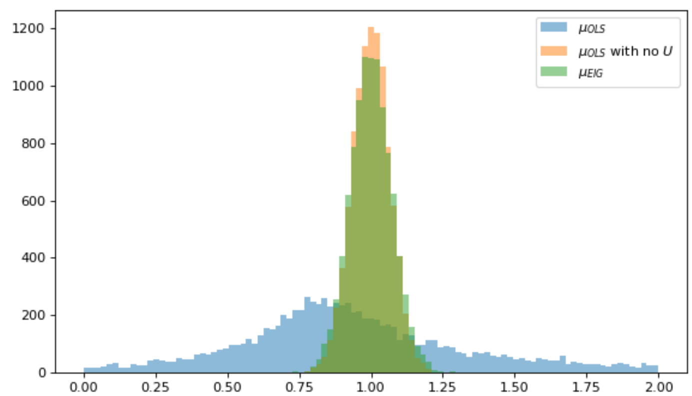

Figure 1.10: *OLS (blue) and eigenvalue-corrected (green) estimators for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}X_iX_j + 10 \times A_iA_j + V_{ij}$, and the "oracle" OLS estimator (orange) for the slope parameter $\beta_{0,1}$ in the model $Y_{ij} := \beta_{0,1} + \beta_{0,2}(X_i + X_j) + V_{ij}$. Note the 10 factor multiplying the A_iA_j term to amplify the skew of the OLS estimator.*

1.7 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 $M(\pi)$'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.

1.8 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 (1.1)) can be adjusted to obtain a $\sqrt{N}$- consistent estimator of the modified estimator (in model 1.2). Then we provide the technical ingredients (propositions 8 and 9) that our main results heavily rely on.

Adjustment to the intercept

Proposition 7. *Under model (1.1) 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 (1.1). Then* $\pi_{0,1}$*, the intercept in the modified model (1.2) is equal to:* $\pi_{0,1} = \beta_{0,1} - \delta\gamma^2$*. 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*

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

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

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

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

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

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

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

□

On the distribution of the largest eigenvalue

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

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

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:

$$||\tilde{A}r|| \leq \max\{|\lambda_N(V)|, \lambda_1(V)\} \leq 3\sigma_v\sqrt{N}||r|| \tag{1.23}$$

Thus:

$$||\tilde{A}r - (U'U)r|| \geq (U'U)||r|| - ||\tilde{A}r|| \geq (U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})||r|| \tag{1.24}$$

implying:

$$||r||^2 \leq \frac{||\tilde{A}r - (U'U)r||^2}{(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2} \leq \frac{||S - L||^2}{(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2} \tag{1.25}$$

With high probability:

$$(U'U - \max\{|\lambda_N(V)|, \lambda_1(V)\})^2 \geq \frac{\sigma_u^4}{2} N^2 \tag{1.26}$$

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

$$\frac{E(||S-L||^2|U)}{N^2} \to \sigma_v^2 E(U_1^2) \text{ almost surely.} \tag{1.27}$$

Also:

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

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

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

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

Hence:

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

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:

$$\mathbb{P}\left(\left| ||S-L||^2 - E(||S-L||^2|U)\right| \geq \sqrt{Var(||S-L||^2|U)}N^{1/3}\right) \leq \frac{1}{N^{2/3}} \tag{1.28}$$

By (1.27) and (1.28), with high probability:

$$||S-L||^2 \leq 2N^2E(U_1^2)\sigma_v^2 \tag{1.29}$$

Combining (1.25),(1.26) and (1.29), with high probability:

$$||r||^2 \leq \frac{4E(U_1^2)}{\sigma_v^2} \tag{1.30}$$

Now note that:

$$\frac{\sum_i S_i^2}{\sum_i S_i U_i} = \frac{S'S}{S'U} = \frac{\lambda_1^2||v||^2 + ||\tilde{A}r||^2}{\lambda_1||v||^2 + r'\tilde{A}r} = \lambda_1 + \frac{||\tilde{A}r||^2 - \lambda_1 r'\tilde{A}r}{\sum_i S_i U_i} \tag{1.31}$$

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 (1.23), w.h.p.:

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

then by (1.30)

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

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,

$$\frac{1}{N^2}\sum_i S_i U_i = E(U_1^2)^2 + o_p(1) \tag{1.32}$$

implying that:

$$\begin{aligned}\lambda_1 &= \frac{\sum_i S_i^2}{\sum_i S_i U_i} + o_p(1) \\ &= \frac{(U'U)^3 + 2(U'U)U'VU + U'V^2U}{(U'U)^2 + U'VU} + o_p(1) \\ &= U'U + \frac{U'VU}{U'U} + \frac{(U'U)U'V^2U - (U'VU)^2}{(U'U)((U'U)^2 + U'VU)} + o_p(1)\end{aligned} \tag{1.33}$$

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 $\pi_0 \in \mathbb{R}^L$. For all π, denote $M(\pi)$ the matrix:*

$$M(\pi) := X(\pi_0 - \pi) + UU' + V - E(U_1^2)I_N$$

where U and V are defined as in Proposition (8), and X is a linear function of the vector $(\pi_0 - \pi)$: $X = \sum_{l=1}^{L}(\pi_{0,l} - \pi_l)X_l$, 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 $\lambda_1(\pi) > \lambda_2(\pi)... > \lambda_N(\pi)$ be the eignevalues of $M(\pi)$, then:

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

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

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

Let $\tilde{\pi}$ be an estimator for π_0 such that $||\tilde{\pi} - \pi_0|| = O_p\left(\frac{1}{\sqrt{N}}\right)$. Then

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

Proof. Note that

$$||M(\tilde{\pi})r(\tilde{\pi})|| \leq |\lambda_2(\tilde{\pi})| \times ||r(\tilde{\pi})||$$

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

$$\lambda_i(M(\tilde{\pi}) - UU') \leq \lambda_i(M(\tilde{\pi})) \leq \lambda_{i-1}(M(\tilde{\pi}) - UU')$$

and by Weyl's inequalities:

$$-||\pi_0 - \tilde{\pi}|| \times |\lambda_1(X)| + \lambda_i(V - E(U_1^2)I_N) \leq \lambda_i(M(\tilde{\pi}) - UU')$$

and

$$\lambda_i(M(\tilde{\pi}) - UU') \leq ||\pi_0 - \tilde{\pi}|| \times |\lambda_1(X)| + \lambda_i(V - E(U_1^2)I_N)$$

so:

$$|\lambda_2(M(\tilde{\pi})| \leq \max\{\lambda_1(V), |\lambda_N(V)|\} + E(U_1^2) + ||\pi_0 - \tilde{\pi}|| \times |\lambda_1(X)|$$

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:

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

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

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

so with high probability:

$$\begin{aligned}
||r(\tilde{\pi})||^2 &\leq \frac{||M(\tilde{\pi})r(\tilde{\pi}) - (U'U)r(\tilde{\pi})||^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2} \\
&\leq \frac{||M(\tilde{\pi})U - (U'U)U||^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2} \\
&\leq \frac{\left(||M(\tilde{\pi})U - M(\pi_0)U|| + ||M(\pi_0)U - (U'U)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2} \\
&= \frac{\left(||\sum_l(\pi_{0,l} - \tilde{\pi}_l)X_lU|| + ||M(\pi_0)U - (U'U)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2} \\
&\leq \frac{\left(\sum_l |\pi_{0,l} - \tilde{\pi}_l| \times |\lambda_1(X_l)| \times ||U|| + ||VU - E(U_1^2)U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2} \\
&= \frac{\left(\sum_l |\pi_{0,l} - \tilde{\pi}_l| \times |\lambda_1(X_l)| \times ||U|| + ||S - L|| + E(U_1^2)||U||\right)^2}{(U'U - |\lambda_2(M(\tilde{\pi}))|)^2}
\end{aligned}$$

where S and L are defined in the proof for equation (8). By equation (1.29), with high probability:

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

so

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

which proves the first result:

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

Also, as in equation (1.33):

$$
\begin{aligned}
\lambda_1(\tilde{\pi}) &= \frac{U'M(\tilde{\pi})'M(\tilde{\pi})U}{U'M(\tilde{\pi})U} + o_p(1) \\
&= \frac{\sum_{k=1}^K \sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)(\pi_{0,k} - \tilde{\pi}_l)U'X_k X_l U + (U'U)\sum_{k=1}^K (\pi_{0,k} - \tilde{\pi}_l)U'X_k U}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\sum_{k=1}^K (\pi_{0,k} - \tilde{\pi}_l)U'X_k VU}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{-E(U_1^2)\sum_{k=1}^K (\pi_{0,k} - \tilde{\pi}_l)U'X_k U}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{(U'U)\sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)U'X_l U + (U'U)^3 + (U'U)U'VU - E(U_1^2)(U'U)^2}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)U'VX_l U + (U'U)U'VU + U'V^2U - E(U_1^2)U'VU}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{-E(U_1^2)\sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)U'X_l U - E(U_1^2)(U'U)^2 - E(U_1^2)U'VU + E(U_1^2)^2 U'U}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} + o_p(1) \\
&= \frac{(U'U)\left(\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U\right)}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\left(\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U/U'U\right)\left(\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U\right)}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&+ \frac{\sum_{k=1}^K \sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)(\pi_{0,k} - \tilde{\pi}_l)U'X_k X_l U + (U'U)(U'VU)}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} \\
&\frac{-E(U_1^2)(U'U)^2 - \left(\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U\right)^2 / U'U + U'V^2U}{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U + (U'U)^2 + U'VU - E(U_i^2)U'U} + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= U'U + \frac{\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U}{U'U} \\
&+ \sum_{k=1}^K \sum_{l=1}^K (\pi_{0,l} - \tilde{\pi}_l)(\pi_{0,k} - \tilde{\pi}_l)\frac{U'X_k X_l U}{(U'U)^2} + \frac{U'VU}{U'U} - E(U_1^2) - \frac{\left(\sum_k (\pi_{0,k} - \tilde{\pi}_k)U'X_k U\right)^2}{(U'U)^3} \\
&+ \frac{U'V^2U}{(U'U)^2} + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
\tag{1.34}
$$

as desired.

For the third part of the proposition, note that:

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

so

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

remember:

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

hence:

$$\begin{aligned}\lambda_1(M(\tilde{\pi}))(U - v(\tilde{\pi})) &= M(\tilde{\pi})r - VU + (\lambda_1(M(\tilde{\pi})) - U'U)U \\ &\quad + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})X_lU + E(U_1^2)U \\ &= M(\tilde{\pi})r - VU + O_p(1)U + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})X_lU + E(U_1^2)U \\ &\quad - \frac{\sum_k(\tilde{\pi}_k - \pi_{0,k})U'X_kU}{U'U}U\end{aligned} \tag{1.35}$$

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

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

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

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

as desired.

Finally, remember that

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

so

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

and

$$\begin{aligned}v(\tilde{\pi})' X_{l'} X_l v(\tilde{\pi}) &= U' X_{l'} X_l U + \frac{1}{\lambda_1(M(\tilde{\pi}))^2}\Delta' X_{l'} X_l \Delta \\ &\quad + \frac{1}{\lambda_1(M(\tilde{\pi}))}\Delta' X_{l'} X_l U + \frac{1}{\lambda_1(M(\tilde{\pi}))}U X_{l'} X_l \Delta'\end{aligned} \tag{1.36}$$

Note that $||\Delta|| = O_p(N)$, $\lambda_1(M(\tilde{\pi})) = O_p(N)$ 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

$$v(\tilde{\pi})'X_{l'}X_lv(\tilde{\pi}) = U'X_{l'}X_lU + O_p(N^2\sqrt{N})$$

□

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:

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

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

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

For a fixed ν in the unit sphere, the function that associates each π to $\sum_{i\neq j}\left(Y_{ij}-X_{ij}\pi\right)^2 - \sum_{i\neq j,k\neq i,j}\nu_i\nu_j\left(Y_{ik}-X_{ik}\pi\right)\left(Y_{kj}-X_{kj}\pi\right)$ 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

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

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 $\nu_ix_{ik}=0$, 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.[10]

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

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

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

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

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

and

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

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

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

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

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

□

Proof of Propositions 1 and 2

Proof. Note that the function f_N is symmetric as a function of the data, that is $f_N(Y_N,X_N;\pi)=f_N(-Y_N,-X_N;\pi)=f_N(\delta Y_N,\delta X_N;\pi)$ for all π 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 the function $f_N(\delta Y_N,\delta X_N;.)$. In other words, given a first stage estimator $\tilde{\pi}$, the estimator $\hat{\pi}$ is numerically the same whether it is computed on the model

$$Y_{ij}=X_{ij}\pi_0+\delta U_iU_j+V_{ij}$$

or

$$(\delta Y_{ij})=(\delta X_{ij})\pi_0+U_iU_j+\delta V_{ij}$$

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

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:

$$(\hat{\pi} - \pi_0) = (1 + o_p(1)) \left(\sum_{i \neq j} X'_{ij} X_{ij} - \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \nu_i(\tilde{\pi}) X_{ij} \right)' \left(\sum_{i=1}^{N} \nu_i(\tilde{\pi}) X_{ij} \right) \right)^{-1}$$

$$\times \left(\sum_{i \neq j} X'_{ij} \left(Y_{ij} - \sum_{l=1}^{K} \pi_{0,l} X_{ij,l} \right) - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\pi}) \nu_j(\tilde{\pi}) X'_{jk} \left(Y_{ik} - \sum_{l=1}^{K} \pi_{0,l} X_{ik,l} \right) \right)$$

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

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

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:

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

implying

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

hence

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

so

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

and

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

implying:

$$\frac{1}{N^2}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}\nu_i(\tilde{\pi})X_{ij}\right)'\left(\sum_{i=1}^{N}\nu_i(\tilde{\pi})X_{ij}\right) = \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32}) + O_p\left(\frac{1}{\sqrt{N}}\right)$$

and $\hat{\pi} - \pi_0$ has the same distribution as

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

Now, define:

$$\begin{aligned}
\hat{\pi}^* := \pi_0 + &\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \\
&\times \frac{1}{N^2}\left(\sum_{i\neq j}X_{ij}'(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}'(U_iU_k + V_{ik})\right)
\end{aligned}$$

The proof proceeds in two steps. First, find the asymptotic distribution of $N(\hat{\pi}^* - \pi_0)$. Second, determine the asymptotic distribution of:

$$\begin{aligned}
\hat{\pi}^* - \pi_0 - &\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{32})\right)^{-1} \\
&\times \frac{1}{N^2}\left(\sum_{i\neq j}X_{ij}'(U_iU_j + V_{ij}) - \sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk}'(U_iU_k + V_{ik})\right)
\end{aligned}$$

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 $\hat{\pi}^* - \pi_0$, 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:

$$\begin{aligned}
\sum_{i\neq j} X_{ij}U_iU_j - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}U_iU_k &= \sum_{i\neq j} X_{ij}U_iU_j - \sum_{i,j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}U_iU_k \\
&\quad + \sum_{i,k\neq i} \frac{U_i^3}{||U||^2} X_{ik}U_k \\
&= \sum_{i\neq j} X_{ij}U_iU_j - \sum_{i,j,k\neq j} \frac{U_i}{||U||}\frac{U_j}{||U||} X_{jk}U_iU_k \\
&\quad + \sum_{i,k\neq i} \frac{U_i^3}{||U||^2} X_{ik}U_k + \sum_{i,j} \frac{U_i^3}{||U||^2} X_{ij}U_j \\
&= 2\sum_{i,j} \frac{U_i^3}{||U||^2} X_{ij}U_j \\
&= N\left(2\frac{1}{E(U_1^2)}E(U_1^3X_{12}U_2) + O_p\left(\frac{1}{\sqrt{N}}\right)\right)
\end{aligned} \tag{1.37}$$

second:

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

note that

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

almost surely

by a standard CLT:

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

hence:

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

and by the Wold device, for a multivariate X:

$$\frac{1}{N}\left(\sum_{i\neq j} X_{ij}(U_iU_j+V_{ij}) - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}(U_iU_k+V_{ik})\right)$$
$$\to_d 2\frac{1}{E(U_1^2)}E(U_1^3U_2X_{12})$$
$$+\mathcal{N}\left(0,\sigma_V^2\left(2E(X_{12}X_{12}') - 3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}') + \frac{1}{E(U_1^2)^2}E(U_1X_{12}U_2)E(U_1X_{12}'U_2)\right)\right)$$

therefore

$$N(\hat{\pi}^* - \pi_0) \to_d$$
$$\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12}'X_{32})\right)^{-1} \times \left(2\frac{1}{E(U_1^2)}E(U_1^3U_2X_{12}) + \mathcal{N}\Big(0, AsV\Big)\right)$$

with the asymptotic variance

$$AsV := \sigma_V^2\left(2E(X_{12}X_{12}') - 3\frac{1}{E(U_1^2)}E(U_1U_3X_{12}X_{23}') + \frac{1}{E(U_1^2)^2}E(U_1X_{12}U_2)E(U_1X_{12}'U_2)\right)\Big)$$

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

$$\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k+V_{ik})-\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik})$$
$$=\nu(\tilde{\pi})'X_\eta M(\pi_0)\nu(\tilde{\pi})-\nu(\tilde{\pi})'diag\left(XM(\pi_0)\right)\nu(\tilde{\pi})-\frac{1}{||U||^2}U'X_\eta M(\pi_0)U$$
$$+\frac{1}{||U||^2}U'diag(X_\eta M(\pi_0))U \tag{1.38}$$
$$=\nu(\tilde{\pi})'X_\eta M(\pi_0)\nu(\tilde{\pi})-\frac{1}{||U||^2}U'X_\eta M(\pi_0)U$$
$$-\left(\nu(\tilde{\pi})'diag\left(X_\eta M(\pi_0)\right)\nu(\tilde{\pi})-\frac{U'}{||U||}diag(X_\eta M)\frac{U'}{||U||}\right)$$

- Case 1: $E(U_1)\neq 0$ and U_i and U_j are arbitrarily correlated with X_{ij}

  On one side, note:[11]

$$\left|\nu(\tilde{\pi})'diag\left(X_\eta M(\pi_0)\right)\nu(\tilde{\pi})-\frac{U'}{||U||}diag(X_\eta M)\frac{U'}{||U||}\right|$$
$$\leq\left|\left|\nu(\tilde{\pi})-\frac{U}{||U||}\right|\right|\max_k\big|\sum_i X_{ik,\eta}(U_iU_k+V_{ik})\big|$$
$$\leq\left|\left|\nu(\tilde{\pi})-\frac{U}{||U||}\right|\right|\max_k\sum_i\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|-E\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|\right)\right)$$
$$+N\left|\left|\nu(\tilde{\pi})-\frac{U}{||U||}\right|\right|E\left(\left|X_{ik.\eta}(U_iU_k+V_{ik})\right|\right)$$

  I want to show that:

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

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

$$\mathbb{P}\left(\frac{1}{N\sqrt{N}}\max_k\sum_i\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|-E\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|\right)\geq x\right)$$
$$\leq\sum_k\mathbb{P}\left(\frac{1}{N\sqrt{N}}\sum_i\left(\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|-E\left|X_{ik,\eta}(U_iU_k+V_{ik})\right|\right)\geq x\right)$$
$$=N\times\mathbb{P}\left(\frac{1}{N\sqrt{N}}\sum_i\left(\left|X_{i1,\eta}(U_iU_1+V_{i1})\right|-E\left|X_{i1,\eta}(U_iU_1+V_{i1})\right|\right)\geq x\right)$$

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

$$\leq \frac{1}{N^2} \frac{Var\left(\sum_i \left(\left|X_{i1,\eta}(U_iU_1 + V_{i1})\right| - E\left|X_{i1,\eta}(U_iU_1 + V_{i1})\right|\right)\right)}{x^2}$$

$$\leq \frac{1}{x^2}\Bigg(Var\left(\left|X_{12,\eta}(U_2U_1 + V_{12})\right|\right)$$

$$+ Cov\left(\left|X_{12,\eta}(U_2U_1 + V_{12})\right|, \left|X_{13,\eta}(U_3U_1 + V_{13})\right|\right)\Bigg)$$

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

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

as desired. Since:

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

then:

$$\nu(\tilde{\pi})' diag\left(X_\eta M(\pi_0)\right)\nu(\tilde{\pi}) - \frac{U'}{||U||} diag(X_\eta M(\pi_0))\frac{U'}{||U||} = O_p(N)$$

and equation (1.38) becomes:

$$\begin{aligned}\sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k + V_{ik}) - &\sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k + V_{ik}) \\ &= \nu(\tilde{\pi})'X_\eta UU'\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta UU'U \\ &\quad + \nu(\tilde{\pi})'XV_\eta\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N}) \\ &= v(\tilde{\pi})'X_\eta U - U'X_\eta U + \quad \nu(\tilde{\pi})'X_\eta V\nu(\tilde{\pi}) \\ &\quad - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N})\end{aligned} \tag{1.39}$$

On one side:

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

On the other side:

$$v(\tilde{\pi})'X_\eta U - U'X_\eta U = U'X_\eta(v(\tilde{\pi}) - U)$$

$$= -\frac{1}{\lambda_1(\tilde{\pi})}\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta X_l U + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\frac{U'X_lU}{U'U}\frac{U'X_\eta U}{\lambda_1(\tilde{\pi})} + O_p(N)$$
$$= \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{U'X_lU}{U'U}\frac{U'X_\eta U}{\lambda_1(\tilde{\pi})} - \frac{1}{\lambda_1(\tilde{\pi})}U'X_\eta X_l U\right) + O_p(N)$$

where the second equality is a consequence of equation (1.35) 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:

$$\frac{1}{N^2}\Bigg(\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k + V_{ik}) - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k + V_{ik})\Bigg)$$
$$= \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{U'X_lU}{NU'U}\frac{U'X_\eta U}{N\lambda_1(\tilde{\pi})} - \frac{N}{\lambda_1(\tilde{\pi})}\frac{1}{N^2}U'X_\eta X_l U\right) + O_p\left(\frac{1}{N}\right)$$
$$= \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{1}{E(U_1^2)^2}E(U_1X_{12,l}U_2)E(U_1U_2X_{12,\eta}) - \frac{1}{E(U_1^2)}E(U_1U_3X_{12,\eta}X_{23,l})\right)$$
$$+ O_p\left(\frac{1}{N}\right)$$

the previous equality holds for any fixed η, so:

$$\frac{\sqrt{N}}{N^2}\Bigg(\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk}(U_iU_k + V_{ik}) - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}(U_iU_k + V_{ik})\Bigg)$$
$$= \frac{1}{E(U_1^2)}\sqrt{N}\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{1}{E(U_1^2)}E(U_1X_{12,l}U_2)E(U_1U_2X_{12,\eta}) - E(U_1U_3X_{12,\eta}X_{23,l})\right)$$
$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$

since, by step 1: $N(\hat{\pi}^* - \pi_0) = O_p(1)$, we conclude:

$$\begin{aligned}
&\sqrt{N}(\hat{\pi} - \pi_0) \\
&= \left(E(X'_{12}X_{12}) - \frac{1}{E(U_1^2)} E(U_1U_3X'_{12}X_{32}) \right)^{-1} \\
&\times \frac{1}{E(U_1^2)} \sqrt{N} \sum_{l=1}^{L} (\tilde{\pi}_l - \pi_{0,l}) \left(E(U_1U_3X_{12,\eta}X_{23,l}) - \frac{1}{E(U_1^2)} E(U_1X_{12,l}U_2)E(U_1U_2X_{12,\eta}) \right) \\
&\quad + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= \frac{1}{E(U_1^2)} \left(E(X'_{12}X_{12}) - \frac{1}{E(U_1^2)} E(U_1U_3X'_{12}X_{32}) \right)^{-1} \\
&\times \left(E(U_1U_3X_{12}X_{23}) - \frac{1}{E(U_1^2)} E(U_1X_{12,l}U_2)E(U_1U_2X_{12,\eta}) \right) \sqrt{N}(\tilde{\pi} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

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

Let's prove the appropriate version of equation (1.39) for this case. On one side, note

$$\begin{aligned}
&\left| \nu(\tilde{\pi})' diag\left(X_\eta M(\pi_0)\right) \nu(\tilde{\pi}) - \frac{U'}{||U||} diag(X_\eta M) \frac{U'}{||U||} \right| \\
&\leq \left|\left| \nu(\tilde{\pi}) - \frac{U}{||U||} \right|\right| \max_k \left| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) \right| \\
&\leq \left|\left| \nu(\tilde{\pi}) - \frac{U}{||U||} \right|\right| \max_k \left| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \right| \\
&+ N \left|\left| \nu(\tilde{\pi}) - \frac{U}{||U||} \right|\right| \max_k \left| E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \right| \\
&\leq \left|\left| \nu(\tilde{\pi}) - \frac{U}{||U||} \right|\right| \max_k \left| \sum_i X_{ik,\eta}(U_iU_k + V_{ik}) - E\left(X_{ik,\eta}(U_iU_k + V_{ik})|X_k, U_k\right) \right|
\end{aligned}$$

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:

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

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

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

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

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:

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

as desired. Since:

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

then:

$$\nu(\tilde{\pi})'diag\left(X_\eta M(\pi_0)\right)\nu(\tilde{\pi}) - \frac{U'}{||U||}diag(X_\eta M(\pi_0))\frac{U'}{||U||} = O_p(\sqrt{N})$$

and equation (1.38) becomes:

$$\begin{aligned}
&\sum_{i\neq j,k\neq i,j}\nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k+V_{ik}) - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik}) \\
&= \nu(\tilde{\pi})'X_\eta M(\pi_0)\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta M(\pi_0)U + O_p(\sqrt{N}) \\
&= \nu(\tilde{\pi})'X_\eta UU'\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta UU'U \\
&+ \nu(\tilde{\pi})'XV_\eta\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N}) \\
&= v(\tilde{\pi})'X_\eta U - U'X_\eta U + \nu(\tilde{\pi})'X_\eta V\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta VU + O_p(\sqrt{N})
\end{aligned} \tag{1.40}$$

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 (1.40):

$$\begin{aligned}\sum_{i\neq j,k\neq i,j} \nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k+V_{ik}) - \sum_{i\neq j,k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}(U_iU_k+V_{ik})\\
&= \nu(\tilde{\pi})'X_\eta M(\pi_0)\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta M(\pi_0)U + O_p(\sqrt{N})\\
&= \nu(\tilde{\pi})'X_\eta M(\tilde{\pi})\nu(\tilde{\pi}) - \frac{1}{||U||^2}U'X_\eta M(\tilde{\pi})U\\
&+ \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\nu(\tilde{\pi})'X_\eta X_l \nu(\tilde{\pi}) - \frac{U'}{||U||}X_\eta X_l \frac{U}{||U||}\right) + O_p(\sqrt{N})\end{aligned}$$

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) $||Xv(\tilde{\pi})|| = ||XU|| + O_p(\sqrt{N})$.

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 (1.35):

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

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

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

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

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

almost surely:

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

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

c) From proposition 9

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

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

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

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

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

In conclusion:

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

so equation (1.39) becomes:

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

when $E(U_1) = 0$, the equation (1.35) yields:

$$\begin{aligned}
\lambda_1(M(\tilde{\pi}))v(\tilde{\pi})'X_\eta r(\tilde{\pi}) &= v(\tilde{\pi})'X_\eta M(\tilde{\pi})r + v(\tilde{\pi})'X_\eta VU + O_p(1)v(\tilde{\pi})'X_\eta U \\
&+ \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})v(\tilde{\pi})'X_\eta X_l U + E(U_1^2)v(\tilde{\pi})'X_\eta U
\end{aligned} \tag{1.41}$$

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

1. $|v(\tilde{\pi})'X_\eta M(\tilde{\pi})r| \leq ||v(\tilde{\pi})'X_\eta|| \times ||M(\tilde{\pi})r|| = O_p(N) \times O_p(\sqrt{N})$
2. for $||v(\tilde{\pi})'X_\eta VU||$, first write: $v(\tilde{\pi})'X_\eta VU = U'X_\eta VU - r(\tilde{\pi})'X_\eta VU$ and note that $|r(\tilde{\pi})'X_\eta VU| \leq ||VU|| \times ||r(\tilde{\pi})'X_\eta|| = O_p(N) \times O_p(\sqrt{N})$, since $||VU|| \leq \lambda_1(V)||U||$ and $||r(\tilde{\pi})'X_\eta|| = O_p(1)$ 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

$$Var\left(\frac{U'X_\eta VU}{||X_\eta U||}\middle| U, X\right) = \sigma_V^2 ||U||^2$$

$$= N\sigma_V^2(E(U_1^2) + o(1)); \quad \text{almost surely}$$

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

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

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

this allows to conclude:

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

and finally, the equation (1.39) becomes:

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

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

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

and

$$N(\hat{\pi} - \pi_0) \to_d E(X_{12}'X_{12})^{-1} \times \mathcal{N}\left(0, 2\sigma_V^2 E(X_{12}X_{12}')\right)$$

□

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:

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

$$= B^{-1}A$$

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:

$$\begin{aligned}&E\left(\left(E(U_1^2)X_{12} - U_2X_{13}U_3\right)\left(E(U_1^2)X_{12}' - U_1X_{24}'U_4\right)\right) \\ &\quad = E(U_1^2)^2E(X_{12}X_{12}') + E(U_1U_2X_{12})E(U_1U_2X_{12}') - 2E(U_1^2)E(U_1U_3X_{12}X_{23}') \\ &\quad = B - A\end{aligned}$$

hence[12]

$$\begin{aligned}B - A &= E\left(\left(E(U_1^2)X_{12} - U_2X_{13}U_3\right)\left(E(U_1^2)X_{12}' - U_1X_{24}'U_4\right)\right) \\ &= E\big(E\left[E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right] \\ &\qquad \times E\left[E(U_1^2)X_{12}' - U_1X_{24}'U_4|X_1, X_2, W_{12}, U_1, U_2\right]\big) \\ &= E\big(E\left[E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right] \\ &\qquad \times E\left[E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right]'\big) \\ &\geq 0\end{aligned}$$

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:

$$\lambda'E\left[E(U_1^2)X_{12} - U_2X_{13}U_3|X_1, X_2, W_{12}, U_1, U_2\right] = E(U_1^2)\lambda'X_{12} - U_2E(\lambda'X_{13}U_3|X_1)$$

so:

$$E(U_1^2)\lambda'X_{12} = U_2E(\lambda'X_{13}U_3|X_1)$$

conditioning on X_2, U_2

$$E(U_1^2)E(\lambda'X_{12}|X_2, U_2) = U_2E(\lambda'X_{13}U_3)$$

hence

$$E(U_1^2)\lambda'X_{12} = U_1U_2\frac{E(\lambda'X_{13}U_3)}{E(U_1^2)}$$

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

which contradicts our assumption that for any vector $\lambda \in \mathbb{R}^L$, $\mathbb{P}(\lambda' X_{12} = U_1 U_2) < 1$. Therefore K has all its eigenvalues < 1. Note that

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

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

Proof of corollary 2

Proof. 1. The function $\pi \to |\lambda_1(M(\pi)^2) - \lambda_2(M(\pi)^2)|$ is continuous on the compact $B(\pi_0, \frac{C}{\sqrt{N}})$. Let π_N be a minimizer on $B(\pi_0, \frac{C}{\sqrt{N}})$. We show in the proof of proposition 9 that $\lambda_1(M(\pi_N)^2) = O_p(N^2)$ and $\lambda_2(M(\pi_N)^2) = O_p(N)$. So $|\lambda_1(M(\pi_N)^2) - \lambda_2(M(\pi_N)^2)| = O_p(N^2)$. So with probability approaching 1, the largest eigenvalue of $M(\pi_N)$ in absolute value is simple on all of $B(\pi_0, \frac{C}{\sqrt{N}})$. The compactness of $B(\pi_0, \frac{C}{\sqrt{N}})$along with theorem 1 in Magnus (1985) allows to conclude that $\pi \to \nu(\pi)$ is infinitely continuously differentiable on $B(\pi_0, \frac{C}{\sqrt{N}})$.

2. Following 1), assume f_N is continuously differentiable on $B(\pi_0, \frac{2C}{\sqrt{N}})$. Let $\pi_{\max}$ be a maximizer of $||f_N'(\pi)||$ on $B(\pi_0, \frac{C}{\sqrt{N}})$. By equation (1.13): $\sqrt{N}(f_N(\pi_{\max}) - \pi_0) = K\sqrt{N}(\pi_{\max} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right)$
   also
   $$\sqrt{N}(f_N(\pi_{\max} + \frac{1}{\sqrt{N}}) - \pi_0) = K\sqrt{N}(\pi_{\max} + \frac{1}{\sqrt{N}} - \pi_0) + O_p\left(\frac{1}{\sqrt{N}}\right)$$
   taking the difference of the two last equations:
   $$\sqrt{N}\left(f_N\left(\pi_{\max} + \frac{1}{\sqrt{N}}\right) - f_N(\pi_{\max})\right) = K + O_p\left(\frac{1}{\sqrt{N}}\right)$$
   on the other side, by a Taylor expansion:[13]
   $$\sqrt{N}\left(f_N\left(\pi_{\max} + \frac{1}{\sqrt{N}}\right) - f_N(\pi_{\max})\right) = f_N'(\pi_{\max}) + o_p(1)$$

[13]Building on the results in Magnus (1985), and after some tedious computations, we can show that $\sup_{\pi \in B(\pi_0, \frac{C}{\sqrt{N}})} ||\frac{\partial \nu(\pi)}{\partial \pi_l}|| = O_p(1)$ and $\sup_{\pi \in B(\pi_0, \frac{C}{\sqrt{N}})} ||\frac{\partial^2 \nu(\pi)}{\partial \pi_l \partial \pi_q}|| = O_p(1)$, implying $\sup_{\pi \in B(\pi_0, \frac{C}{\sqrt{N}})} ||f_N''(\pi)|| = O_p(1)$.

hence

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

with a probability approaching 1:

$$||f'(\pi_{\max})|| = \sup_{\pi \in B(\pi_0, \frac{C}{\sqrt{N}})} ||f'(\pi)|| \leq \kappa$$

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 $1 - \epsilon$, $\hat{\pi}_1, \hat{\pi}_0 \in B(\pi_0, \frac{M}{2\sqrt{N}})$ so that $||\hat{\pi}_1 - \hat{\pi}_0|| \leq \frac{M}{\sqrt{N}}$ (let this be event E_N) . Assume f_N is continuously differentiable on $B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$ (denote this event F_N) and that $\sup_{\pi \in B\left(\pi_0, \frac{M}{\sqrt{N}}(1+\frac{1}{1-\kappa})\right)} ||f'(\pi)|| \leq \kappa$ (let this be event G_N). Then for any $\pi, \pi' \in B(\pi_0, \frac{M}{\sqrt{N}}(1 + \frac{1}{1-\kappa}))$, we have:

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

By induction on m, assume $\hat{\pi}_0, ..., \hat{\pi}_m \in B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$

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

and

$$\begin{aligned} ||\hat{\pi}_{m+1} - \pi_0|| &\leq ||\hat{\pi}_{m+1} - \hat{\pi}_m|| + ||\hat{\pi}_m - \hat{\pi}_{m-1}|| + ... + ||\hat{\pi}_1 - \hat{\pi}_0|| + ||\hat{\pi}_0 - \pi_0|| \\ &\leq \sum_{i=0}^{m} \kappa^i ||\hat{\pi}_1 - \hat{\pi}_0|| + ||\hat{\pi}_0 - \pi_0|| \\ &\leq \frac{M}{\sqrt{N}}(1 + \sum_{i=0}^{m} \kappa^i) \\ &\leq \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right) \end{aligned}$$

so $\hat{\pi}_{m+1} \in B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$.

So even though f_N is not necessarily contracting on $B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$, because it may not preserve $B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1 + \frac{1}{1-\kappa}\right)\right)$, we can follow the proof of the Banach fixed point theorem for the specific sequence $\hat{\pi}$ (or in fact any sequence initiated in a way that

the first two first elements are in $B\left(\pi_0, \frac{M}{2\sqrt{N}}\right)$ and not just $B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$).
First we show that the sequence $\hat{\pi}_m$ is a Cauchy sequence, let $p, q \in \mathbb{N}$, without loss of generality take $p > q$

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

so the sequence is a Cauchy sequence. Therefore, it has a limit in $B\left(\pi_0, \frac{M}{\sqrt{N}}\left(1+\frac{1}{1-\kappa}\right)\right)$ 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 $\hat{\pi}^*$ is a solution to the minimization problem 1.6 and is $\sqrt{N}$-consistent. Equation (1.13) yields

$$(I-K)\sqrt{N}(\hat{\pi}^* - \pi_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

and finally

$$(\hat{\pi}^* - \pi_0) = O_p\left(\frac{1}{\sqrt{N}}\right)$$

□

Proof of proposition 4

Proof. As for the proof of propositions 1 (section 1.8), assume that $\delta = 1$. The result for an unknown $\delta \in \{-1, 1\}$ immediately follows as described in the the proof section 1.8 . 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 (1.39):

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= \sum_{i\neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}U_iU_k \\
&+ \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}U_iU_k - \sum_{i\neq j, k\neq i,j} \nu_i(\tilde{\pi})\nu_j(\tilde{\pi})X_{jk,\eta}(U_iU_k + V_{ik}) \\
&= \sum_{i\neq j} X_{ij}(U_iU_j + V_{ij}) - \sum_{i\neq j, k\neq i,j} \frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk}(U_iU_k + V_{ik})
\end{aligned}$$

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

where the second equality results from equation (1.39) and the third from equation (1.37). Note that

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

   to see that, observe that from equation (1.35):

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

   combining both identities:

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

$$+\sum_{l=1}^{L}(\tilde{\pi}_l-\pi_{0,l})X_\eta X_l U+E(U_1^2)X_\eta U-\frac{\sum_k(\tilde{\pi}_k-\pi_{0,k})U'X_kU}{U'U}X_\eta U\Big)'$$
$$\times\Big(VM(\tilde{\pi})r(\tilde{\pi})-V^2U+O_p(1)VU+\sum_{l=1}^{L}(\tilde{\pi}_l-\pi_{0,l})VX_lU+E(U_1^2)VU$$
$$-\frac{\sum_k(\tilde{\pi}_k-\pi_{0,k})U'X_kU}{U'U}VU\Big)$$
$$=U'X_\eta VU+\frac{1}{\lambda_1(\tilde{\pi})}U'X_\eta V^2U+O_p(N\sqrt{N})$$

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:

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

Let's determine the asymptotic distribution of $U'X_\eta r(\tilde{\pi})$. From the equation (1.35)

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

Also

$$\begin{aligned}
\lambda_1(M(\tilde{\pi}))U'X_\eta M(\tilde{\pi})r(\tilde{\pi}) &= U'X_\eta M(\tilde{\pi})^2 r - U'X_\eta M(\tilde{\pi})VU + (\lambda_1(M(\tilde{\pi})) - U'U)U'X_\eta M(\tilde{\pi})U \\
&\quad + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta M(\tilde{\pi})X_l U + E(U_1^2)U'X_\eta M(\tilde{\pi})U \\
&= -U'X_\eta UU'VU - U'XV^2U - \sum_{l=1}^{L}(\pi_{0,l} - _l)U'X_\eta X_l VU \\
&\quad + (\lambda_1(M(\tilde{\pi})) - U'U)U'X_\eta UU'U \\
&\quad + (\lambda_1(M(\tilde{\pi})) - U'U)U'X_\eta VU \\
&\quad + (\lambda_1(M(\tilde{\pi})) - U'U)\sum_{l=1}^{L}(\pi_0 - \tilde{\pi})U'X_\eta X_l U \\
&\quad + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta UU'X_l U + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta VX_l U \\
&\quad + \sum_{k=1}^{L}(\pi_{0,k} - \tilde{\pi}_k)\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta X_k X_l U + E(U_1^2)U'X_\eta UU'U \\
&\quad + E(U_1^2)U'X_\eta VU + E(U_1^2)\sum_{l=1}^{L}(\pi_{0,l} - \tilde{\pi}_l)U'X_\eta X_l U + O_p(N^2\sqrt{N}) \\
&= -U'X_\eta UU'VU - U'X_\eta V^2U + (\lambda_1(M(\tilde{\pi})) - U'U)U'X_\eta UU'U \\
&\quad + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta UU'X_l U + E(U_1^2)U'X_\eta UU'U + O_p(N^2\sqrt{N})
\end{aligned}$$

By equation (1.34), when $\tilde{\pi} - \pi_0 = O_p\left(\frac{1}{N}\right)$ as we are assuming here, we get:

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

therefore:

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

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

$$\begin{aligned}
\lambda_1(M(\tilde{\pi}))U'X_\eta r(\tilde{\pi}) &= -\frac{1}{\lambda_1(\tilde{\pi})}U'X_\eta V^2U + \frac{U'V^2UU'X_\eta U}{\lambda_1(\tilde{\pi})U'U} - U'X_\eta VU \\
&\quad + (\lambda_1(M(\tilde{\pi})) - U'U)U'X_\eta U + \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta X_l U + E(U_1^2)U'X_\eta U
\end{aligned}$$

$$
\begin{aligned}
&+ O_p(N\sqrt{N}) \\
&= \frac{1}{U'U} U'X_\eta UU'VU - \frac{1}{\lambda_1(\tilde{\pi})} U'X_\eta V^2 U + \frac{U'V^2UU'X_\eta U}{\lambda_1(\tilde{\pi})U'U} - U'X_\eta VU \\
&+ \frac{U'V^2UU'X_\eta U}{(U'U)^2} - \frac{\sum_k (\tilde{\pi}_k - \pi_{0,k}) U'X_k UU'X_\eta U}{U'U} \\
&+ \sum_{l=1}^{L} (\tilde{\pi}_l - \pi_{0,l}) U'X_\eta X_l U + O_p(N\sqrt{N})
\end{aligned}
$$

so

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

plugging:

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

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

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

subsequently

$$
\sum_{i \neq j} X_{ij,\eta}(U_iU_j + V_{ij}) - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\pi})\nu_j(\tilde{\pi}) X_{jk,\eta}(U_iU_k + V_{ik})
$$

$$
\begin{aligned}
&= N\left(\frac{2}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2)\right) + \sum_{i\neq j} X_{ij,\eta}V_{ij} - \sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} \\
&+ \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU - \frac{1}{U'U}U'X_\eta VU \\
&- \frac{\sum_k(\tilde{\pi}_k - \pi_{0,k})U'X_kUU'X_\eta U}{(U'U)^2} + \frac{1}{U'U}\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})U'X_\eta X_l U + O_p(\sqrt{N}) \\
&= \sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{U'X_\eta X_l U}{U'U} - \frac{U'X_lU}{U'U}\frac{U'X_\eta U}{U'U}\right) \\
&+ N\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) - E(U_1X_{12}U_2)\right) \\
&+ \sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU + O_p(\sqrt{N}) \\
&= N^2\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{1}{E(U_1^2)}E(U_1U_3X_{12,\eta}X_{23,l}) - \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,l})E(U_1U_2X_{12,\eta})\right) \\
&+ N\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) - E(U_1X_{12}U_2)\right) \\
&+ \sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\sum_{i\neq j,k\neq i,j}\frac{U_i}{||U||}\frac{U_j}{||U||}X_{jk,\eta}V_{ik} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U'VU + O_p(\sqrt{N}) \\
&= N^2\sum_{l=1}^{L}(\tilde{\pi}_l - \pi_{0,l})\left(\frac{1}{E(U_1^2)}E(U_1U_3X_{12,\eta}X_{23,l}) - \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,l})E(U_1U_2X_{12,\eta})\right) \\
&\quad + R_{N,\eta} + O_p(\sqrt{N})
\end{aligned}
$$

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

$$
\begin{aligned}
R_{N,\eta} :=& N\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2}E(U_1U_2X_{12})\right) \\
&+ \sum_{i\neq j} X_{ij,\eta}V_{ij} - 2\frac{1}{NE(U_1^2)}\sum_{i\neq j,k\neq i,j}U_iU_jX_{jk,\eta}V_{ik} \\
&\quad + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})\sum_{i\neq j}U_iU_jV_{ij} + O_p(\sqrt{N}) \\
&= N\left(\frac{3}{E(U_1^2)}E(U_1^3X_{12,\eta}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2}E(U_1U_2X_{12})\right) \\
&+ \sum_{i\neq j} V_{ij}\left(X_{ij,\eta} - 2U_i\frac{1}{NE(U_1^2)}\sum_{k\neq i,j}U_kX_{jk,\eta} + \frac{1}{E(U_1^2)^2}E(U_1U_2X_{12,\eta})U_iU_j\right)
\end{aligned}
$$

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

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_k X_{jk,\eta} + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \\
&\quad + X_{ij,\eta} - 2U_j \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{ik,\eta} + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \Bigg)^2 \\
&= \sigma_V^2 \sum_{i<j} \Bigg(X_{ij,\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_1U_2X_{12,\eta}) U_i U_j \Bigg)^2 \\
&+ \sigma_V^2 \sum_{i<j} \Bigg(X_{ij,\eta} - 2U_j \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{ik,\eta} + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \Bigg)^2 \\
&+ 2\sigma_V^2 \sum_{i<j} \Bigg(X_{ij,\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_1U_2X_{12,\eta}) U_i U_j \Bigg) \\
&\qquad \times \Bigg(X_{ij,\eta} - 2U_j \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{ik,\eta} + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \Bigg) \\
&= \sigma_V^2 \sum_{i\neq j} \Bigg(X_{ij,\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_1U_2X_{12,\eta}) U_i U_j \Bigg)^2 \\
&+ 2\sigma_V^2 \sum_{i<j} \Bigg(X_{ij,\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_1U_2X_{12,\eta}) U_i U_j \Bigg) \\
&\qquad \times \Bigg(X_{ij,\eta} - 2U_j \frac{1}{NE(U_1^2)} \sum_{k\neq i,j} U_k X_{ik,\eta} + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \Bigg) \\
&= \sigma_V^2 \sum_{i\neq j} X_{ij,\eta}^2 + 4U_i^2 \frac{1}{N^2E(U_1^2)^2} \Bigg(\sum_{k\neq i,j} U_k X_{jk,\eta} \Bigg)^2 + \frac{1}{E(U_1^2)^4} E(U_1U_2X_{12,\eta})^2 U_i^2 U_j^2 \\
&- \sigma_V^2 \frac{4}{NE(U_1^2)} \sum_{i\neq j, k\neq i,j} X_{ij,\eta} U_i U_k X_{jk,\eta} + 2\frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) \sum_{i\neq j} X_{ij,\eta} U_i U_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^2 U_j U_k X_{jk,\eta} \\
&+ \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_1U_2X_{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_iU_j \frac{1}{N^2E(U_1^2)^2} \sum_{k,l\neq i,j} U_k U_l X_{ik,\eta} X_{jl,\eta}
\end{aligned}
$$

$$
\begin{aligned}
&- 2U_i \frac{1}{NE(U_1^2)} \sum_{k \neq i,j} U_k X_{jk,\eta} \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j \Bigg) \\
&+ \sigma_V^2 \sum_{i \neq j} \Bigg(+ \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) U_i U_j X_{ij,\eta} - 2 \frac{1}{NE(U_1^2)^3} E(U_1U_2X_{12,\eta}) \sum_{k \neq i,j} U_i U_j^2 U_k X_{ik,\eta} \\
&+ \frac{1}{E(U_1)^4} E(U_1U_2X_{12,\eta})^2 U_i^2 U_j^2 \Bigg) \\
&= N^2 \sigma_V^2 \Bigg(E(X_{12,\eta}^2) + 4 \frac{1}{E(U_1^2)} E(U_1U_3X_{12,\eta}X_{23,\eta}) + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&- \frac{4}{E(U_1^2)} E(U_1U_3X_{12,\eta}X_{23,\eta}) + \frac{2}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 - \frac{4}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&+ E(X_{12,\eta}^2) - 2 \frac{1}{E(U_1^2)} E(U_1U_3X_{12,\eta}X_{23,\eta}) + \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&- \frac{2}{E(U_1^2)} E(U_1U_3X_{12,\eta}X_{23,\eta}) + \frac{4}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&- \frac{2}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&+ \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 - 2 \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 \\
&+ \frac{1}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 + o(1) \Bigg) \text{ almost surely.} \\
&= N^2 \sigma_V^2 \Bigg(2E(X_{12,\eta}^2) + \frac{2}{E(U_1^2)^2} E(U_1U_2X_{12,\eta})^2 - \frac{4}{E(U_1^2)} E(U_1U_3X_{12,\eta}X_{23,\eta}) + o(1) \Bigg) \\
&\quad \text{(almost surely)}
\end{aligned}
$$

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

$$
\frac{1}{N} R_{N,\eta} \to_d \left(\frac{3}{E(U_1^2)} E(U_1^3 X_{12,\eta} U_2) - \frac{E(U_1^4)}{E(U_1^2)^2} E(U_1U_2X_{12,\eta}) \right) + \mathcal{N}\left(0, 2\sigma_V^2 \eta \Sigma \eta'\right)
$$

for

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

Finally:

$$
\begin{aligned}
N(\hat{\pi} - \pi_0) = &\left(E(X_{12}'X_{12}) - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}'X_{32}) \right)^{-1} \\
&\times \frac{1}{N} \left(\sum_{i \neq j} X_{ij}'(U_iU_j + V_{ij}) - \sum_{i \neq j, k \neq i,j} \nu_i(\tilde{\pi}) \nu_j(\tilde{\pi}) X_{jk}'(U_iU_k + V_{ik}) \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{E(U_1^2)} \left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}'X_{32}) \right)^{-1} \\
&\times \left(E(U_1U_3X_{12}X_{23}') - \frac{1}{E(U_1^2)} E(U_1U_2X_{12})E(U_1U_2X_{12}') \right) N(\tilde{\pi} - \pi_0) \\
&+ R_N + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= KN(\tilde{\pi} - \pi_0) + R_N + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

with

$$
\begin{aligned}
K :&= \frac{1}{E(U_1^2)} \left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}X_{23}') \right)^{-1} \\
&\times \left(E(U_1U_3X_{12}X_{23}') - \frac{1}{E(U_1^2)} E(U_1U_2X_{12})E(U_1U_2X_{12}') \right) \\
R_N &\to_d \left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}X_{32}') \right)^{-1} \\
&\times \left(\frac{3}{E(U_1^2)} E(U_1^3X_{12,\eta}U_2) - \frac{E(U_1^4)}{E(U_1^2)^2} E(U_1U_2X_{12}) \right) \\
&+ \left(E(X_{12}X_{12}') - \frac{1}{E(U_1^2)} E(U_1U_3X_{12}X_{23}') \right)^{-1} \mathcal{N}\left(0, 2\sigma_V^2\Sigma\right)
\end{aligned}
$$

□

Proof of proposition 5

Proof. Write:

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

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

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

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

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

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

Proof of lemma 3

Proof. First, note that, for $r \geq 2$

$$\left|||v(\tilde{\mu})||_r - ||U||_r\right| \leq ||v(\tilde{\mu}) - U||_r \leq ||v(\tilde{\mu}) - U||_2 = O_p(1)$$

Implying the first result: $\sum_i |\nu(\tilde{\pi})|^r = \frac{\sum_i |U_i|^r}{\max_i |\lambda_i(\tilde{\pi})|^{\frac{r}{2}}} + O_p\left(\frac{1}{N^{1/r}}\right)$. When $\delta = 1$, by proposition 9, with probability approaching 1 $\hat{\delta} = \frac{\lambda_1(\tilde{\pi})}{|\lambda_1(\tilde{\pi})|}$ and $\frac{\lambda_1(\tilde{\pi})}{|\lambda_1(\tilde{\pi})|} = 1$. When $\delta = -1$,

$$\begin{aligned}\hat{\delta} := &\frac{-\lambda_N(-M(\tilde{\pi}))\mathbb{1}\{-\lambda_N(-M(\tilde{\pi})) > \lambda_1(-M(\tilde{\pi}))\}}{\max_i |\lambda_i(\tilde{\pi})|} \\ &- \frac{\lambda_1(-M(\tilde{\pi}))\mathbb{1}\{-\lambda_N(-M(\tilde{\pi})) < \lambda_1(-M(\tilde{\pi}))\}}{\max_i |\lambda_i(\tilde{\pi})|}\end{aligned}$$

so with probability approaching 1:

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

by the same reasoning as for the case $\delta = 1$,

$$\begin{aligned}&\frac{\lambda_N(-M(\tilde{\pi}))\mathbb{1}\{|\lambda_N(-M(\tilde{\pi}))| > \lambda_1(-M(\tilde{\pi}))\}}{\max_i |\lambda_i(\tilde{\pi})|} \\ &+ \frac{\lambda_1(-M(\tilde{\pi}))\mathbb{1}\{|\lambda_N(-M(\tilde{\pi}))| < \lambda_1(-M(\tilde{\pi}))\}}{\max_i |\lambda_i(\tilde{\pi})|} = 1\end{aligned}$$

with probability approaching 1. so $\hat{\delta} = -1$ with probability approaching 1.

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

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

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

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

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

$$= O_p(1)$$

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

$$\begin{aligned}
&\left|\left|\frac{1}{N^2}\sum_{i\neq j,}\hat{U}_i\hat{U}_j X_{ij,\eta} - \frac{1}{N^2}\sum_{i\neq j}U_iU_jX_{ij,\eta}\right|\right| \\
&= \left|\left|\frac{1}{N^2}\left(\sum_{i\neq j}(\hat{U}_i - U_i)\hat{U}_j X_{ij,\eta} + \sum_{i\neq j}U_i(\hat{U}_j - U_j)X_{ij,\eta}\right)\right|\right| \\
&\leq \frac{1}{N^2}\left(||\hat{U} - U||_2||X||_2||U||_2 + ||\hat{U} - U||_2||X||_2||\hat{U}||_2\right) \\
&= O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

The two remaining results are proved similarly. □

Proof of equation (1.22):

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:

$$\begin{aligned}
e_{1,m}'e_{2,m} &= \left(\frac{1}{4}b_{m,N}^2c_{m,N}^2\sigma_A^2 - \frac{1}{4b_{m,N}^2}\right)N + b_{m,N}^2||A||^2 + b_{m,N}^2c_{m,N}\sigma_A\iota' A \\
&= \frac{1}{4b_{m,N}^2}\left(b_{m,N}^4\left(4||A||^2 + Nc_{m,N}^2\sigma_A^2 + 4c_{m,N}\sigma_A\iota' A\right) - N\right) \\
&= 0
\end{aligned}$$

□

Proof of proposition 6

In line with the proofs leading up to theorems 1 and 2, we beging by studying the behavior of the $M(\hat{\pi}_m)$'s largest eigenvalue ($\hat{\pi}_m$ defined in equation 1.21). First, decompose:

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

and likewise

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

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

Lemma 4.

- $\lambda_1(\hat{\pi}_m) - \lambda_{1,m} = -\frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_l e_{1,m} + O_p(1)$
- $\lambda_N(\hat{\pi}_m) - \lambda_{2,m} = -\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e'_{2,m}X_l e_{2,m} + O_p(1)$
- $||v_{11}(\hat{\pi}_m) - e_{1,m}|| = O_p(1)$, $v_{12}(\hat{\pi}_m) = O_p(1)$ *and* $r_1(\hat{\pi}_m) = O_p(1)$
- $||v_{21}(\hat{\pi}_m)|| = O_p(1)$, $||v_{22}(\hat{\pi}_m) - e_2|| = O_p(1)$ *and* $r_2(\hat{\pi}_m) = O_p(1)$
- $\lambda_1(\hat{\pi}_m)(v_{11}(\hat{\pi}_m) - e_{1,m}) = -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\pi}_m) + (\lambda_{1,m} - \lambda_1(\hat{\pi}_m) + O_p(1))e_{1,m}$ $+ \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m} + O_p(1)\right)e_{2,m} + Vv_{11}(\tilde{\pi}) - \sigma_A c_{m,N}v_{11}(\hat{\pi}_m)$
- $\lambda_N(\hat{\pi}_m)(v_{22}(\hat{\pi}_m) - e_{2,m}) = -\sum_l Z_{m,N,l}X_l v_{22}(\hat{\pi}_m) + (\lambda_{2,m} - \lambda_N(\hat{\pi}_m) + O_p(1))e_{2,m}$ $+ \left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_l e_{1,m} + O_p(1)\right)e_{1,m} + Vv_{22}(\tilde{\pi}) - \sigma_A c_{m,N}v_{22}(\hat{\pi}_m)$

Proof. We show the results successively:

- $\boldsymbol{||v_{11}(\hat{\pi}_m) - e_{1,m}|| = O_p(1)}$, $\boldsymbol{v_{12}(\hat{\pi}_m) = O_p(1)}$ **and** $\boldsymbol{r_1(\hat{\pi}_m) = O_p(1)}$

On one side:

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

on another side:

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

so

$$\begin{aligned} &M(\hat{\pi}_m)r_1(\hat{\pi}_m) + \lambda_1(\hat{\pi}_m)v_{11}(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m)v_{12}(\hat{\pi}_m) \\ &\qquad = -\sum_l Z_{m,N,l}X_l e_{1,m} + Ve_1 + \lambda_{1,m}e_{1,m} - \sigma_A c_{m,N}e_{1,m} \end{aligned}$$

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

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

similarly

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

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

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

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

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

and

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

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

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

and

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

so

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

also

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

which implies that

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

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

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

and by the Pythagorean theorem:

$$\begin{aligned}||M(\hat{\pi}_m)r(\hat{\pi}_m) - \lambda_{1,m} r_1(\hat{\pi}_m)||^2 &\leq ||M(\hat{\pi}_m)e_1 - \lambda_{1,m}e_1||^2 \\ &= ||-\sum_l Z_{m,N,l}X_l e_1 + Ve_1 - \sigma_A c_{m,N}e_1||^2 \\ &= O_p(N^2)\end{aligned}$$

in conclusion:

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

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

$$\begin{aligned}&M(\hat{\pi}_m)r_1(\hat{\pi}_m) + \lambda_1(\hat{\pi}_m)v_{11}(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m)v_{12}(\hat{\pi}_m) \\ &\quad = -\sum_l Z_{m,N,l}X_l e_{1,m} + Ve_{1,m} + \lambda_{1,m}e_{1,m} - \sigma_A c_{m,N}e_{1,m}\end{aligned}$$

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

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

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

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

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

so

$$\lambda_1(\hat{\pi}_m)(v_{11}(\hat{\pi}_m) - e_{1,m}) = -\sum_l Z_{m,N,l}X_l v_{11}(\hat{\pi}_m) + Vv_{11}(\hat{\pi}_m) - \sigma_A c_{m,N}v_{11}(\hat{\pi}_m)$$

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

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

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

so

$$
\begin{aligned}
\lambda_1(\hat{\pi}_m) e'_{2,m} v_{11}(\hat{\pi}_m) = &-\sum_l Z_{m,N,l} e_2 X_l v_{11}(\hat{\pi}_m) + e_2 V v_{11}(\hat{\pi}_m) + \lambda_{2,m}(e'_{2,m} v_{11}(\hat{\pi}_m)) \\
&- \sigma_A c_{m,N}(e'_{2,m} v_{11}(\hat{\pi}_m))
\end{aligned}
$$

implying

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

or

$$
\left(\lambda_{1,m} - \lambda_{2,m} + O_p(\sqrt{N})\right) e'_2 v_{11}(\tilde{\pi}) = -\sum_l Z_{m,N,l} e_2 X_l e_1 + O_p(N)
$$

with

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

and

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

therefore:

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

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

  Write:

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

so

$$\begin{aligned} \lambda_2(\hat{\pi}_m)(v_{22}(\hat{\pi}_m) - e_{2,m}) &= -\sum_l Z_{m,N,l} X_l v_{22}(\hat{\pi}_m) + V v_{22}(\hat{\pi}_m) - \sigma_A c_{m,N} v_{22}(\hat{\pi}_m) \\ &\quad + \left(-||v_{22}(\hat{\pi}_m)||^2 - \lambda_2(\hat{\pi}_m)\right) e_{2,m} + (e'_{2,m} v_{11}(\hat{\pi}_m)) e_{1,m} \\ &= -\sum_l Z_{m,N,l} X_l v_{22}(\hat{\pi}_m) + V v_{22}(\hat{\pi}_m) - \sigma_A c_{m,N} v_{22}(\hat{\pi}_m) \\ &\quad + \left(\lambda_{2,m} - \lambda_2(\hat{\pi}_m) + O_p(1)\right) e_{2,m} - (e'_{2,m} v_{11}(\hat{\pi}_m)) e_{1,m} \end{aligned}$$

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

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

so

$$\lambda_2(\hat{\pi}_m) e'_{1,m} v_{22}(\hat{\pi}_m) = -\sum_l Z_{m,N,l} e_{1,m} X_l v_{22}(\hat{\pi}_m) + e_{1,m} V v_{22}(\hat{\pi}_m) + \lambda_{1,m}(e'_{2,m} v_{11}(\hat{\pi}_m)) - \sigma_A c_{m,N}(e'_{1,m} v_{22}(\hat{\pi}_m))$$

implying

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

or

$$\left(\lambda_{2,m} - \lambda_{1,m} + O_p(\sqrt{N})\right) e'_1 v_{22}(\tilde{\pi}) = -\sum_l Z_{m,N,l} e_{1,m} X_l e_{2,m} + O_p(N)$$

therefore:

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

□

Note that:

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

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

$$\hat{\pi}_{m+1} - \pi_0 = \left(\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\hat{\pi}_m)\nu_j(\hat{\pi}_m) X_{jk}' X_{ik} \right)^{-1} \times \left(\sum_{i \neq j} X_{ij}' (A_i + A_j + V_{ij}) - \sum_{i \neq j, k \neq i,j} \nu_i(\hat{\pi}_m)\nu_j(\hat{\pi}_m) X_{jk}' (A_i + A_k + V_{ik}) \right)$$

First, note:

$$\begin{aligned}
&\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{i \neq j, k \neq i,j} \nu_i(\hat{\pi}_m)\nu_j(\hat{\pi}_m) X_{jk}' X_{ik} \\
&\quad = \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) \geq 0) \left(\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{i \neq j, k \neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} \frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk}' X_{ik} \right) \\
&\quad + \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) < 0) \left(\sum_{i \neq j} X_{ij}' X_{ij} - \sum_{i \neq j, k \neq i,j} \frac{v_{22,i}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||} \frac{v_{22,j}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||} X_{jk}' X_{ik} \right)
\end{aligned}$$

We treat each of the two terms separately:

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

likewise

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

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

For the term $\left(\sum_{i\neq j} X'_{ij}(A_i + A_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\hat{\pi}_m)\nu_j(\hat{\pi}_m)X'_{jk}(A_i + A_k + V_{ik})\right)$, 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 $X_{ij,\eta} := \eta' X_{ij}$.

$$\begin{aligned}
&\sum_{i\neq j} X'_{ij}(A_i + A_j + V_{ij}) - \sum_{i\neq j,k\neq i,j} \nu_i(\hat{\pi}_m)\nu_j(\hat{\pi}_m)X'_{jk}(A_i + A_k + V_{ik}) \\
&= \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) \geq 0) \\
&\quad \times \left(\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} \frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) \right) \\
&+ \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) < 0) \\
&\quad \times \left(\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{22,i}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||} \frac{v_{22,j}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) \right)
\end{aligned}$$

We have

$$\begin{aligned}
&\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} \frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) \\
&= \sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) \\
&\quad - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} \frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk,\eta} \left(M(\hat{\pi}_m)_{ik} - \sum_l (\pi_{0,l} - \hat{\pi}_{m,l}) X_{ik,l} \right) \\
&= \sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \frac{1}{||v_{11}(\hat{\pi}_m)||^2} v_{11}(\hat{\pi}_m)' M(\hat{\pi}_m) X_\eta v_{11}(\hat{\pi}_m) \\
&\quad - \frac{1}{||v_{11}(\hat{\pi}_m)||^2} \sum_l (\pi_{0,l} - \hat{\pi}_{m,l}) v_{11}(\tilde{\pi})' X_l X_\eta v_{11}(\hat{\pi}_m) \\
&\quad + \sum_{ik} \frac{v_{11,i}^2(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||^2} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right) + O_p(N)
\end{aligned}$$

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:

$$\mathbb{P}\left(\frac{1}{N\sqrt{N}} \max_k \sum_i \left(\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| - E\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right|\right) \geq x\right)$$
$$\leq \sum_k \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right| - E\left|X_{ik,\eta}(A_i + A_k + V_{ik})\right|\right) \geq x\right)$$
$$= N \times \mathbb{P}\left(\frac{1}{N\sqrt{N}} \sum_i \left(\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| - E\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right|\right) \geq x\right)$$
$$\leq \frac{1}{N^2} \frac{Var\left(\sum_i \left(\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right| - E\left|X_{i1,\eta}(A_i + A_1 + V_{i1})\right|\right)\right)}{x^2}$$
$$\leq \frac{1}{x^2}\Bigg(Var\left(\left|X_{12,\eta}(A_2 + A_1 + V_{12})\right|\right)$$
$$+ Cov\left(\left|X_{12,\eta}(A_2 + A_1 + V_{12})\right|, \left|X_{13,\eta}(A_3 + A_1 + V_{13})\right|\right)\Bigg)$$

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

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

hence

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

Note that

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

so

$$\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j, k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||}\frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right)$$
$$= N^2\Bigg(- \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2b_{m,N}}\right)^2$$

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

Similarly

$$\sum_{i\neq j}X_{ij,\eta}\left(A_i+A_j+V_{ij}\right)-\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}\frac{v_{22,j}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}X_{jk,\eta}\left(A_i+A_k+V_{ik}\right)$$
$$=-\frac{1}{||v_{22}(\hat{\pi}_m)||^2}v_{22}(\hat{\pi}_m)'M(\hat{\pi}_m)X_\eta v_{22}(\hat{\pi}_m)$$
$$+\frac{1}{||v_{22}(\hat{\pi}_m)||^2}\sum_l(\pi_{0,l}-\hat{\pi}_{m,l})v_{22}(\hat{\pi}_m)'X_lX_\eta v_{22}(\hat{\pi}_m)+O_p(N\sqrt{N})$$
$$=-\frac{\lambda_2(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||^2}v_{22}(\hat{\pi}_m)'X_\eta v_{22}(\hat{\pi}_m)+\frac{1}{||v_{22}(\hat{\pi}_m)||^2}\sum_l(\pi_{0,l}-\hat{\pi}_{m,l})v_{22}(\hat{\pi}_m)'X_lX_\eta v_{22}(\hat{\pi}_m)$$
$$+O_p(N\sqrt{N})$$
$$=-\frac{\lambda_{2,m}}{||e_{2,m}||^2}e_{2,m}'X_\eta e_{2,m}+\frac{c_{m,N}}{||v_{22}(\hat{\pi}_m)||^2}\sigma_Av_{22}(\hat{\pi}_m)'\iota\iota'X_\eta v_{22}(\hat{\pi}_m)+O_p(N\sqrt{N})$$
$$=e_{2,m}'X_\eta e_{2,m}+\frac{c_{m,N}}{||e_{2,m}||^2}\sigma_Ae_{2,m}'\iota\iota'X_\eta e_{2,m}+O_p(N\sqrt{N})$$

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

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

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

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

Therefore

$$\hat{\pi}_{m+1} - \pi_0$$

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

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

$$= \Bigg[- \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) \geq 0)$$

$$\times \left(\frac{\sigma_A}{\sqrt{4 + c_{m,N}^2}} + \frac{1}{4\sigma_A \sqrt{4 + c_{m,N}^2}} \left(c_{m,N}\sigma_A + \sigma_A \sqrt{4 + c_{m,N}^2} \right)^2 \right)$$

$$+ \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) < 0)$$

$$\times \left(\frac{\sigma_A}{\sqrt{4 + c_{m,N}^2}} + \frac{1}{4\sigma_A \sqrt{4 + c_{m,N}^2}} \left(c_{m,N}\sigma_A - \sigma_A \sqrt{4 + c_{m,N}^2} \right)^2 \right) \Bigg] (1, 0, 0, ..., 0)'$$

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

$$= \frac{\sigma_A}{\sqrt{4 + c_{m,N}^2}} \Bigg[- \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) \geq 0) \left(1 + \frac{1}{4} \left(c_{m,N} + \sqrt{4 + c_{m,N}^2} \right)^2 \right)$$

$$+ \mathbb{1}(\lambda_1(\hat{\pi}_m) + \lambda_N(\hat{\pi}_m) < 0) \left(1 + \frac{1}{4} \left(c_{m,N} - \sqrt{4 + c_{m,N}^2} \right)^2 \right) \Bigg] (1, 0, 0, ..., 0)'$$

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

Note that for any m, by lemma 4:

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

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 $m > 1$:

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

$$= c_{m,N} \sigma_A + O_p \left(\frac{1}{\sqrt{N}} \right)$$

so, for all $m > 1$:

$$\hat{\pi}_{m+1} - \pi_0 = \frac{\sigma_A}{\sqrt{4+c_{m,N}^2}}\Bigg[- \mathbb{1}(c_{m,N} \geq 0)\left(1 + \frac{1}{4}\left(c_{m,N} + \sqrt{4+c_{m,N}^2}\right)^2\right)$$
$$+ \mathbb{1}(c_{m,N} < 0)\left(1 + \frac{1}{4}\left(c_{m,N} - \sqrt{4+c_{m,N}^2}\right)^2\right)\Bigg](1,0,0,...,0)' + O_p\left(\frac{1}{\sqrt{N}}\right)$$

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

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

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

and

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

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 $m_0 \in \mathbb{N}$:

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

and

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

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

$$
\begin{aligned}
&= -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}{2 b_{m,N}} \right) \iota' X_\eta e_{1,m} \\
&\quad + 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} \\
&\quad - \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} \\
&\quad - \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} \\
&\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)' \\
&\qquad \times \left(2 X_\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}{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_{1,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} \\
&\quad + \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}{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_{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} \\
&\quad - \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} \\
&\quad - \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} \\
&\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)' \\
&\qquad \times \left(2 X_\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}{2 b_{m,N}} \right)^2 \sum_{ij} X_{ij,\eta}
\end{aligned}
$$

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

where

$T_{1,m,N}$

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

$$+\left(\left(\frac{Nc_{m,N}}{||e_{1,m}||^2}\sigma_A - 2\right) b_{m,N}\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right) + 2\right)\sum_{ij} X_{ij}A_j$$
$$= O_p\left(N\sqrt{N}\right)$$

so:

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

Note that

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

Hence:

$$\frac{1}{N^2}\left(\sum_{i\neq j} X_{ij,\eta}\left(A_i + A_j + V_{ij}\right) - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||}\frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk,\eta}\left(A_i + A_k + V_{ik}\right)\right)$$

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

$$+ \left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N} E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}\right.\right.$$

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

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

Define

$$\sum_{i\neq j} X_{ij}' X_{ij} - \sum_{i\neq j,k\neq i,j} \frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||}\frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||} X_{jk}' X_{ik}$$

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

likewise

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

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

Therefore:

$$\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) > 0)\sqrt{N} diag(0, 1, ..., 1)(\hat{\pi}_{m+1} - \pi_0)$$

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

$$
+ \frac{1}{\sqrt{N}} M_{1,m,N}^{-1} T_{1,m,N}
- \left(\frac{c_{m,N}}{E(e_{1,m}^2)} \sigma_A - 1 \right) \left(\frac{1}{2} b_{m,N} c_{m,N} \sigma_A + \frac{1}{2b_{m,N}} \right)^2 \sqrt{N} M_{1,m,N}^{-1} R_{1,m,N} M_{1,m,N}^{-1} E(X_{12})
+ O_p\left(\frac{1}{\sqrt{N}}\right)
$$

with

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

so

$$
\begin{aligned}
&\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) > 0)\sqrt{N} diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0) \\
= \quad & \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} \frac{\left(\frac{1}{2} b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} E(X_{12}X_{23}') \sqrt{N} Z_{m,N} + \frac{1}{\sqrt{N}} M_m^{-1} T_{1,m,N} \\
& - \left(\frac{c_m}{E(e_{1,m}^2)} \sigma_A - 1 \right) \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_{m,N}} \right)^2 \sqrt{N} M_m^{-1} R_{1,m,N} M_m^{-1} E(X_{12}) + O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}
$$

with

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

so that

$$
\begin{aligned}
&\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) > 0)\sqrt{N} diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0) \\
= \quad & \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{\left(\frac{1}{2} b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} E(X_{12}X_{23}') \sqrt{N} Z_{m,N} \\
& + \frac{1}{\sqrt{N}} M_m^{-1} T_{1,m,N} \\
& - \left(\frac{c_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \sigma_A - 1 \right) \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_{m,N}} \right)^2 \sqrt{N} M_m^{-1} R_{1,m,N} M_m^{-1} E(X_{12})
\end{aligned}
$$

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

Generally, including the intercept:

$$\begin{aligned}
&\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) > 0)\sqrt{N}(\hat{\pi}_{m+1} - \pi_0) \\
&= -\sigma_A c_{m+1,N}(1,0,...,0)' + \\
&\left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N} E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}\right.\right. \\
&\left.\left. - \frac{2\sigma_A c_{m,N} E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N} E(e_{1,m})^2 E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)(1,0,...,0)' E(X_{12}') Z_{m,N} \\
&+ \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{\left(\frac{1}{2} b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} E(X_{12} X_{23}') \sqrt{N} Z_{m,N} \\
&+ \frac{1}{\sqrt{N}} M_m^{-1} T_{1,m,N} \\
&- \left(\frac{c_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \sigma_A - 1\right)\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_{m,N}}\right)^2 \sqrt{N} M_m^{-1} R_{1,m,N} M_m^{-1} E(X_{12}) \\
&+ O_p\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

We treat the term

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

in the same way:

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

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

$$
\begin{aligned}
&+ \frac{1}{||e_{2,m}||^4} \sum_l Z_{m,N,l} e_{2,m}' X_l e_{2,m} e_{2,m}' X_\eta e_{2,m} - \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e_{2,m}' X_l X_\eta e_{2,m} \\
&+ \frac{1}{||e_{2,m}||^2} \Bigg(- \sum_l Z_{m,N,l} X_l e_{2,m} + \frac{1}{||e_{2,m}||^2} \sum_l Z_{m,N,l} e_{2,m}' X_\eta e_{2,m} e_{2,m} \\
&+ \left(\frac{1}{||e_{1,m}||^2 + ||e_{2,m}||^2} \sum_l Z_{m,N,l} e_{1,m} X_l e_{2,m} \right) e_{1,m} \Bigg)' \\
&\times \left(-2 X_\eta e_{2,m} - \frac{c_{m,N}}{||e_{2,m}||^2} \sigma_A \left(\iota\iota' X_\eta e_{2,m} + X_\eta \iota\iota' e_{2,m} \right) \right) + O_p(N)
\end{aligned}
$$

where

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

so:

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

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

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:

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

Define

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

where:

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

Therefore:

$$\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\sqrt{N} diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0)$$
$$= \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0) diag(0,1,...,1)$$
$$\times \left(\frac{E(e_{2,m})^2}{E(e_{2,m}^2)} \frac{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A + \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2}{\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 + b_{m,N}^2\sigma_A^2} M_{2,m,N}^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N} \right.$$
$$+ \frac{1}{N\sqrt{N}} M_{2,m,N}^{-1} T_{2,m,N}$$
$$- \left(\frac{c_{m,N}}{E(e_{2,m}^2)}\sigma_A + 1\right)\left(\frac{1}{2}b_{m,N}c_{m,N}\sigma_A - \frac{1}{2b_{m,N}}\right)^2 \sqrt{N} M_{2,m,N}^{-1} R_{2,m,N} M_{2,m,N}^{-1} E(X_{12})$$
$$\left. + O_p\left(\frac{1}{\sqrt{N}}\right)\right)$$

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

$$\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\sqrt{N} diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0)$$

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

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

$$\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\sqrt{N} diag(0, 1, ..., 1)(\hat{\pi}_{m+1} - \pi_0)$$
$$= \frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} \frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2} M_m^{-1} E(X_{12}X_{23}')\sqrt{N} Z_{m,N}$$
$$+ \frac{1}{N\sqrt{N}} M_m^{-1} T_{2,m}$$
$$- \left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right)$$
$$\times \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2} \sqrt{N} M_m^{-1} R_{2,m,N}(1, 0, ..., 0)'$$
$$+ O_p\left(\frac{1}{\sqrt{N}}\right)$$

Again, including the intercept:

$$\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\sqrt{N}(\hat{\pi}_{m+1} - \pi_0) = -c_{m+1,N}\sigma_A(1, 0, 0, ..., 0)'$$

$$+\frac{E(e_{1,m}^2)}{E(e_{1,m}^2)-E(e_{1,m})^2}\left(\frac{E(e_{2,m})^4}{E(e_{2,m}^2)^2}+\frac{1}{E(e_{2,m}^2)}\left(\frac{\sigma_A c_{m,N}E(e_{2,m})^2}{E(e_{2,m}^2)}-2\frac{E(e_{2,m})^4}{E(e_{2,m}^2)}\right.\right.$$

$$\left.\left.-2\frac{E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)+E(e_{1,m}^2)}-\frac{2\sigma_A c_{m,N}E(e_{2,m})^4}{E(e_{2,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{2,m})^2E(e_{1,m})^2}{E(e_{2,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\right)\right)$$

$$\times(1,0,0,...0)'E(X_{12}')Z_{m,N}$$

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

$$+\frac{1}{N\sqrt{N}}M_m^{-1}T_{2,m}$$

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

$$\times\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{2,m,N}(1,0,...,0)'$$

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

In conclusion:

$$\sqrt{N}diag(0,1,...,1)(\hat{\pi}_{m+1}-\pi_0)=diag(0,1,...,1)\times\Bigg(\tag{1.42}$$

$$\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N} \tag{1.43}$$

$$+\left(\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)+2\right) \tag{1.44}$$

$$\times\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j \tag{1.45}$$

$$+\big(\mathbb{1}(\lambda_1(\hat{\pi})+\lambda_N(\hat{\pi})\geq 0)-\mathbb{1}(\lambda_1(\hat{\pi})+\lambda_N(\hat{\pi})<0)\big) \tag{1.46}$$

$$\times\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2 \tag{1.47}$$

$$\times \sqrt{N}\left(\frac{1}{N^2}\sum_{ij} X_{ij} - E(X_{12})\right) \tag{1.48}$$

$$- \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \geq 0)\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right) \tag{1.49}$$

$$\times \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{1,m,N}(1,0,...,0)' \tag{1.50}$$

$$- \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\left(-\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A + 1\right) \tag{1.51}$$

$$\times \frac{\left(\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{2,m,N}(1,0,...,0)'\Bigg) \tag{1.52}$$

$$+ O_p\left(\frac{1}{\sqrt{N}}\right) \tag{1.53}$$

With the intercept:

$$\sqrt{N}(\hat{\pi}_{m+1} - \pi_0) = -\sigma_A c_{m+1,N}(1,0,...,0)'$$
$$+ \frac{E(e_{1,m}^2)}{E(e_{1,m}^2) - E(e_{1,m})^2}\left(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{1}{E(e_{1,m}^2)}\left(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}\right.\right.$$
$$\left.\left. + 2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)} - \frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2} - \frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2) + E(e_{2,m}^2))}\right)\right)$$
$$\times (1,0,...,0)'E(X_{12}')Z_{m,N}$$
$$\frac{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}$$
$$+ \left(\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right)b_m\left(\frac{1}{2}b_m c_m\sigma_A + \frac{1}{2b_m}\right) + 2\right)$$
$$\times \frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij} X_{ij}A_j$$
$$+ \big(\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \geq 0) - \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\big)$$

$$\times\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2$$

$$\times\sqrt{N}\left(\frac{1}{N^2}\sum_{ij}X_{ij}-E(X_{12})\right)$$

$$-\mathbb{1}(\lambda_1(\hat{\pi})+\lambda_N(\hat{\pi})\geq 0)\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)$$

$$\times\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{1,m,N}(1,0,...,0)'$$

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

$$\times\frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{2,m,N}(1,0,...,0)'\Bigg)$$

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

write

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

$$-\left(\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{11,i}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||}\frac{v_{11,j}(\hat{\pi}_m)}{||v_{11}(\hat{\pi}_m)||}X'_{jk}-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12})\right)$$

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

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

$$=\frac{1}{N^2||v_{11}(\hat{\pi}_m)||^2}\Bigg(\sum_{i,k\neq j}v_{11,i}(\hat{\pi}_m)v_{11,j}(\hat{\pi}_m)X'_{jk,\eta}-\sum_{j\neq k}v_{11,j}(\hat{\pi}_m)^2X_{jk,\eta}$$

$$-\sum_{j\neq k}v_{11,j}(\hat{\pi}_m)v_{11,k}(\hat{\pi}_m)X_{jk,\eta}\Bigg)-\frac{E(e_{1,m})^2}{E(e_{1,m}^2)}E(X_{12,\eta})$$

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

$$+\frac{1}{N^2||e_{1,m}||^4}\Bigg(-\sum_l Z_{m,N,l}X_le_{1,m}-\frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_le_{1,m}e_{1,m}e_{1,m}$$

$$+\left(\frac{1}{||e_{1,m}||^2+||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}X_le_{1,m}\right)e_{2,m}\Bigg)'\iota\iota' X_\eta e_{1,m}$$

$$+\frac{1}{N^2||e_{1,m}||^4}\left(e'_{1,m}\iota\right)\iota' X_\eta\Bigg(-\sum_l Z_{m,N,l}X_le_{1,m}-\frac{1}{||e_{1,m}||^2}\sum_l Z_{m,N,l}e'_{1,m}X_le_{1,m}e_{1,m}$$

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

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

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

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

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

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

hence:

$$diag(0,1,1,...,1)M_m^{-1}R_{1,m,N}(1,0,...,0)'$$

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

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

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

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

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

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

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

$$- diag(0,1,1,...,1) M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \Big(NE(e_{1,m})^2 \sum_{ij} X_{ij,\eta} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij,\eta}$$

$$+ NE(e_{1,m}) b_m \sum_{ij} A_i X_{ij,\eta} + b_m^2 \sum_i A_i \sum_{ij} A_i X_{ij,\eta} \Big)$$

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

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

$$- diag(0,1,1,...,1) M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \Big(NE(e_{1,m})^2 \sum_{ij} X_{ij,\eta} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij,\eta}$$

$$+ NE(e_{1,m}) b_m \sum_{ij} A_i X_{ij,\eta} \Big)$$

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

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

$$- diag(0,1,1,...,1) M_m^{-1} \frac{E(e_{1,m})}{N^2 E(e_{1,m}^2)} \Big(E(e_{1,m}) \sum_{ij} \left(X_{ij,\eta} - E(X_{ij,\eta}) \right) + b_m \sum_{ij} A_i X_{ij,\eta} \Big)$$

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

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

$$- diag(0,1,1,...,1) M_m^{-1} \frac{E(e_{1,m}) b_m}{N^2 E(e_{1,m}^2)} \sum_{ij} A_i X_{ij,\eta}$$

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

Including the intercept:

$M_m^{-1} R_{1,m,N} (1,0,...,0)'$

$$
\begin{aligned}
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - 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})^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) \\
&\qquad\qquad \times (1, 0, ..., 0)' E(X_{12})' Z_{m,N} \\
&\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) \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \Bigg(N E(e_{1,m})^2 \sum_{ij} X_{ij} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij} \\
&\quad + 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} \Bigg) + \frac{E(e_{1,m})^2}{E(e_{1,m}^2)} M_m^{-1} E(X_{12}) \\
&\quad + \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} \\
&\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) \\
&= \frac{1}{N^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) \\
&\quad - M_m^{-1} \frac{1}{N^2 ||e_{1,m}||^2} \Bigg(N E(e_{1,m})^2 \sum_{ij} X_{ij} + E(e_{1,m}) b_m \sum_i A_i \sum_{ij} X_{ij} \\
&\quad + N E(e_{1,m}) b_m \sum_{ij} A_i X_{ij} \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&+ \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) \\
&- M_m^{-1}\frac{E(e_{1,m})b_m}{N^2 E(e_{1,m}^2)}\sum_{ij} A_i X_{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^2 E(e_{1,m}^2)}\sum_{ij} A_i X_{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}
\end{aligned}
$$

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

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

$$-M_m^{-1}\frac{E(e_{1,m})b_m}{N^2E(e_{1,m}^2)}\sum_{ij}A_iX_{ij,\eta}$$

$$+\frac{1}{E(e_{1,m}^2)b_m\sigma_A^2}\left(b_m\left(\sigma_A^2-\frac{\sum A_i^2}{N}\right)-3E(e_{1,m})\frac{\sum_i A_i}{N}\right)(1,0,...,0)'$$

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

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

and similarly for $R_{2,m,N}$, write

$$R_{2,m,N}(1,0,...,0)'=\frac{1}{N^2}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$-\left(\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}\frac{v_{22,j}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}X_{jk}'-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12})\right)$$

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

$$\frac{1}{N^2}\sum_{i\neq j,k\neq i,j}\frac{v_{22,i}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}\frac{v_{22,j}(\hat{\pi}_m)}{||v_{22}(\hat{\pi}_m)||}X_{jk,\eta}'-\frac{E(e_{2,m})^2}{E(e_{2,m}^2)}E(X_{12,\eta})$$

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

$$+\frac{1}{N^2||e_{2,m}||^2\lambda_N(\hat{\pi}_m)}\Bigg(-\sum_l Z_{m,N,l}X_lv_{22}(\hat{\pi}_m)+(\lambda_{2,m}-\lambda_N(\hat{\pi}_m))e_{2,m}$$

$$+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_le_{2,m}\right)e_{1,m}\Bigg)'\iota\iota'X_\eta e_{2,m}$$

$$+\frac{1}{N^2||e_{2,m}||^2\lambda_N(\hat{\pi}_m)}\left(e_{2,m}'\iota\right)\iota'X_\eta\Bigg(-\sum_l Z_{m,N,l}X_lv_{22}(\hat{\pi}_m)+(\lambda_{2,m}-\lambda_N(\hat{\pi}_m))e_{2,m}$$

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

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

$$-\frac{1}{N^2||e_{2,m}||^4}\Bigg(-\sum_l Z_{m,N,l}X_le_{2,m}-\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_le_{2,m}e_{2,m}$$

$$+\left(\frac{1}{||e_{2,m}||^2+||e_{1,m}||^2}\sum_l Z_{m,N,l}e_{1,m}X_le_{2,m}\right)e_{1,m}\Bigg)'\iota\iota'X_\eta e_{2,m}$$

$$-\frac{1}{N^2||e_{2,m}||^4}\left(e_{2,m}'\iota\right)\iota'X_\eta\Bigg(-\sum_l Z_{m,N,l}X_le_{2,m}-\frac{1}{||e_{2,m}||^2}\sum_l Z_{m,N,l}e_{2,m}'X_le_{2,m}e_{2,m}$$

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

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

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

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

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

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

hence:

$$diag(0,1,1,...,1)M_m^{-1}R_{2,m,N}(1,0,...,0)'$$

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

$$-diag(0,1,1,...,1)M_m^{-1}\frac{E(e_{2,m})b_m}{N^2E(e_{2,m}^2)}\sum_{ij}A_iX_{ij,\eta}$$

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

$$= \frac{1}{N^2} diag(0,1,1,...,1) \frac{b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right)$$
$$+ diag(0,1,1,...,1) M_m^{-1} \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{1}{N^2} \sum_{ij} A_i X_{ij,\eta}$$
$$- \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2\right)^2} diag(0,1,1,...,1) M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right)$$

Including the intercept:

$$M_m^{-1} R_{2,m,N} (1,0,...,0)' =$$
$$\frac{1}{N^2} \left(1 - \frac{E(e_{2,m})^2}{E(e_{2,m}^2)}\right) M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right) - M_m^{-1} \frac{E(e_{2,m}) b_m}{N^2 E(e_{2,m}^2)} \sum_{ij} A_i X_{ij,\eta}$$
$$+ \frac{E(e_{2,m})}{E(e_{2,m}^2)^2} \left(-E(e_{2,m}) - 2\frac{E(e_{2,m})^3}{E(e_{2,m}^2)} + 2\frac{E(e_{1,m})^2 E(e_{2,m})}{E(e_{2,m}^2) + E(e_{1,m}^2)} \right) M_m^{-1} E(X_{12}) E(X_{12})' Z_{m,N}$$
$$+ \left(\frac{E(e_{2,m}^2) - ||e_{2,m}||^2/N}{E(e_{2,m}^2)^2} - \frac{E(e_{2,m}) b_m}{E(e_{2,m}^2)^2} \frac{\sum_i A_i}{N} \right) M_m^{-1} E(X_{12})$$
$$- \frac{E(e_{2,m})^2}{E(e_{2,m}^2)^2} M_m^{-1} E(X_{12} X_{23}') Z_{m,N} + O_p\left(\frac{1}{N}\right)$$
$$= \frac{1}{N^2} \frac{b_m^2 \sigma_A^2}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} M_m^{-1} \left(\sum_{ij} X_{ij} - E(X_{ij}) \right)$$
$$+ M_m^{-1} \frac{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right) b_m}{\left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2 \sigma_A^2} \frac{1}{N^2} \sum_{ij} A_i X_{ij,\eta}$$
$$+ \frac{E(e_{2,m})^2}{E(e_{2,m}^2) b_m^2 \sigma_A^2} \left(-1 - 2\frac{E(e_{2,m})^2}{E(e_{2,m}^2)} + 2\frac{E(e_{1,m})^2}{E(e_{2,m}^2) + E(e_{1,m}^2)} \right) (1,0,...,0)' E(X_{12})' Z_{m,N}$$
$$+ \frac{1}{E(e_{2,m}^2) b_m \sigma_A^2} \left(b_m \left(\sigma_A^2 - \frac{\sum A_i^2}{N} \right) - 3E(e_{2,m}) \frac{\sum_i A_i}{N} \right) (1,0,...,0)'$$

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

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

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

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

$$+ 2\frac{\left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 + \left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2 + 2b_m^2 \sigma_A^2}\Bigg)(1, 0, ..., 0)' E(X_{12})' Z_{m,N}$$

$$+ \frac{1}{E(e_{2,m}^2) b_m \sigma_A^2} \left(b_m \left(\sigma_A^2 - \frac{\sum A_i^2}{N}\right) + 3\left(\frac{1}{2}b_m c_m \sigma_A + \frac{1}{2b_m}\right) \frac{\sum_i A_i}{N}\right)(1, 0, ..., 0)'$$

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

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

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

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

$$+ 2b_m^2 \sigma_A^2 \left(\frac{1}{2}b_m c_m \sigma_A - \frac{1}{2b_m}\right)^2\Bigg)(1, 0, ..., 0)' E(X_{12})' Z_{m,N}$$

$$+\frac{1}{E(e_{2,m}^2)b_m\sigma_A^2}\left(b_m\left(\sigma_A^2-\frac{\sum A_i^2}{N}\right)+3\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)\frac{\sum_i A_i}{N}\right)(1,0,...,0)'$$
$$-\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)^2}M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

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:

$$diag(0,1,1,...,1)M_m^{-1}R_{1,m,N}(1,0,...,0)'$$
$$=\frac{1}{N^2}diag(0,1,1,...,1)\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$-diag(0,1,1,...,1)M_m^{-1}\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)b_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{1}{N^2}\sum_{ij}A_iX_{ij,\eta}$$
$$+\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)^2}diag(0,1,1,...,1)M_m^{-1}E(X_{12}X_{23}')Z_{m,N}+O_p\left(\frac{1}{N}\right)$$

Including the intercept:

$$M_m^{-1}R_{1,m,N}(1,0,...,0)'=\frac{1}{N^2}\frac{b_m^2\sigma_A^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}M_m^{-1}\left(\sum_{ij}X_{ij}-E(X_{ij})\right)$$
$$-M_m^{-1}\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)b_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\frac{1}{N^2}\sum_{ij}A_iX_{ij,\eta}$$
$$+\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2\right)b_m^2\sigma_A^2}\left(1+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}+\right.$$

$$- 2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A - \frac{1}{2b_m}\right)^2\Bigg)(1,0,...,0)'E(X_{12})'Z_{m,N}$$

$$+ \frac{1}{E(e_{1,m}^2)b_m\sigma_A^2}\left(b_m\left(\sigma_A^2 - \frac{\sum A_i^2}{N}\right) - 3\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)\frac{\sum_i A_i}{N}\right)(1,0,...,0)'$$

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

plugging in equation (1.42):

$$\sqrt{N}diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0) = diag(0,1,...,1) \times \Bigg[$$

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

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

$$\times \frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

$$+ \left(\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \geq 0) - \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 0)\right)$$

$$\times \left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2\sqrt{N}\left(\frac{1}{N^2}\sum_{ij}X_{ij} - E(X_{12})\right)$$

$$- \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \geq 0)\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)$$

$$\times \frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{1,m,N}(1,0,...,0)'$$

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

$$\times \frac{\left(\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2\right)\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^2}\sqrt{N}M_m^{-1}R_{2,m,N}(1,0,...,0)'\Bigg]$$

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

$$= diag(0,1,...,1) \times \Bigg($$

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

$$+ \Bigg[\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 2\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right) + 2+$$

$$+ \left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^2 + b_m^2\sigma_A^2}\sigma_A - 1\right)\frac{b_m\left(\frac{1}{2}b_mc_m\sigma_A + \frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2}\Bigg]\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

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

Including the intercept

$$\sqrt{N}(\hat{\pi}_{m+1} - \pi_0) = -\sigma_A c_{m+1,N}(1,0,...,0)'$$

$$+ \frac{E(e_{1,m}^2)}{E(e_{1,m}^2) - E(e_{1,m})^2}\Bigg(\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}$$

$$- \frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)} + 2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2) + E(e_{2,m}^2)}$$

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

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

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

$$
\begin{aligned}
&\quad \times \frac{1}{N\sqrt{N}} M_m^{-1} \sum_{ij} X_{ij} A_j \\
&+ \left(\mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \geq 0) - \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 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) \\
&\quad \times \left(\frac{1}{2} b_m c_m \sigma_A + \frac{1}{2b_m}\right)^2 \sqrt{N} \left(\frac{1}{N^2} \sum_{ij} X_{ij} - E(X_{12}) \right) \\
&- \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) \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) \\
&\quad \times \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_{1,m,N} (1, 0, ..., 0)' \\
&\quad - \mathbb{1}(\lambda_1(\hat{\pi}) + \lambda_N(\hat{\pi}) < 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) \\
&\quad \times \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_{2,m,N} (1, 0, ..., 0)' \Bigg) \\
&\quad + O_p\left(\frac{1}{\sqrt{N}}\right) \\
&= -\sigma_A c_{m+1,N} (1, 0, ..., 0)' \\
&\quad + \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} \\
&\quad - \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)} \\
&\quad - \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} \\
&\quad + \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}
\end{aligned}
$$

$$+\Bigg[\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)+2+$$

$$+\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{b_m\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2}\Bigg]\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

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

$$\times\left(1+2\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_m c_m\sigma_A-\frac{1}{2b_m}\right)^2\right)(1,0,...,0)'$$

$$\times\sqrt{N}E(X_{12})'Z_{m,N}$$

$$+3\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^3\sigma_A^4}\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'$$

$$+\left(\mathbb{1}(\lambda_1(\hat{\pi})+\lambda_N(\hat{\pi})\geq 0)-\mathbb{1}(\lambda_1(\hat{\pi})+\lambda_N(\hat{\pi})<0)\right)\left(\frac{\sum A_i^2}{N}-\sigma_A^2\right)$$

$$\times\left(\frac{c_m}{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_m c_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^4}\sqrt{N}(1,0,0,...,0)'$$

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

Let's simplify the coefficient of the term $M_m^{-1}E(X_{12})E(X_{12})'Z_{m,N}$:

$$\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_A c_{m,N}E(e_{1,m})^2}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^4}{E(e_{1,m}^2)}+2\frac{E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)+E(e_{2,m}^2)}$$

$$-\frac{2\sigma_A c_{m,N}E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_A c_{m,N}E(e_{1,m})^2E(e_{2,m})^2}{E(e_{1,m}^2)(E(e_{1,m}^2)+E(e_{2,m}^2))}\Bigg)$$

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

$$\left(1+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2\right)$$

$$=\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_Ac_{m,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)}+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)}+2b_m^6\sigma_A^6$$

$$-\frac{2\sigma_Ac_{m,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_Ac_{m,N}b_m^6\sigma_A^6}{E(e_{1,m}^2)}\Bigg)$$

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

$$\left(1+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2\right)$$

$$=\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_Ac_{m,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)}+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)}+2b_m^6\sigma_A^6$$

$$-\frac{2\sigma_Ac_{m,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_Ac_{m,N}b_m^6\sigma_A^6}{E(e_{1,m}^2)}\Bigg)$$

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

$$\left(1+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2\right)$$

To simplify notation, for every c, $c_{1,N}$ and σ_A denote

$$A(\sigma_A,c,c_{1,N}):=\frac{E(e_{1,m})^4}{E(e_{1,m}^2)^2}-\frac{1}{E(e_{1,m}^2)}\Bigg(\frac{\sigma_Ac_mc_{1,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{E(e_{1,m}^2)}+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)}$$

$$+2b_m^6\sigma_A^6-\frac{2\sigma_Ac_{m,N}\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^4}{E(e_{1,m}^2)^2}-\frac{2\sigma_Ac_{m,N}b_m^6\sigma_A^6}{E(e_{1,m}^2)}\Bigg)$$

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

$$\left(1+2\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}-2b_m^2\sigma_A^2\left(\frac{1}{2}b_mc_m\sigma_A-\frac{1}{2b_m}\right)^2\right)$$

$$B(\sigma_A,c):=\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-2\right)b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)+2+$$

$$+\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{b_m\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^2\sigma_A^2}$$

$$C(\sigma_A,c):=3\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^3}{b_m^3\sigma_A^4}$$

$$D(\sigma_A,c,c_{1,N}):=c_{1,N}\left(\frac{c_m}{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2+b_m^2\sigma_A^2}\sigma_A-1\right)\frac{\left(\frac{1}{2}b_mc_m\sigma_A+\frac{1}{2b_m}\right)^2}{b_m^2\sigma_A^4}$$

(remember $c_{1,N}:=\mathbb{1}(\lambda_1(\hat\pi)+\lambda_N(\hat\pi)\geq 0)-\mathbb{1}(\lambda_1(\hat\pi)+\lambda_N(\hat\pi)<0)$) so that:

$$\begin{aligned}
&\sqrt{N}(\hat\pi_{m+1}-\pi_0)\\
&=-\sigma_Ac_{m+1,N}(1,0,...,0)'+A(\sigma_A,c_m)M(c_m)^{-1}E(X_{12})E(X_{12}')Z_{m,N}\\
&\quad+M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}+B(\sigma_A,c_m)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j\\
&\quad+C(\sigma_A,c_m)\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'+D(\sigma_A,c_m)\sqrt{N}\left(\frac{\sum A_i^2}{N}-\sigma_A^2\right)(1,0,0,...,0)'\\
&\quad+O_p\left(\frac{1}{\sqrt{N}}\right)\\
&=-\sigma_Ac_{m+1,N}(1,0,...,0)'+M(c_m)^{-1}\left(E(X_{12}X_{23}')+A(\sigma_A,c_m)E(X_{12})E(X_{12}')\right)Z_{m,N}\\
&\quad+B(\sigma_A,c_m)\frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j+C(\sigma_A,c_m)\sqrt{N}\frac{\sum_i A_i}{N}(1,0,...,0)'
\end{aligned}$$

$$+ D(\sigma_A, c_m)\sqrt{N}\left(\frac{\sum A_i^2}{N} - \sigma_A^2\right)(1,0,0,...,0)' + O_p\left(\frac{1}{\sqrt{N}}\right)$$

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

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

therefore:

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

and

$$\sqrt{N}diag(0,1,...,1)(\hat{\pi}_{m+1} - \pi_0) = diag(0,1,...,1) \times \Bigg(M_m^{-1}E(X_{12}X_{23}')\sqrt{N}Z_{m,N}$$

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

$$\times \frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

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

$$= diag(0,1,...,1) \times \Bigg(M_m^{-1}E(X_{12}X_{23}')\sqrt{N}diag(0,1,...,1)(\hat{\pi}_m - \pi_0)$$

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

$$\times \frac{1}{N\sqrt{N}}M_m^{-1}\sum_{ij}X_{ij}A_j$$

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

Lemma for the proof of Theorem 3

Lemma 5.

$$\hat{K} \to_p K_0$$

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 1.8. □

Chapter 2

A general estimation procedure for exchangeable random graph models

2.1 Introduction

Consider a the following general model:

$$Y_{ij} := h(X_i, X_j, U_i, U_j, V_{ij}; \beta) \tag{2.1}$$

for all $i, j \leq N$ and for some measurable and known measurable function h, *i.i.d.* variables X_i, U_i and V_{ij}, which are also mutually independent, and for some parameter β.[1] We are interested in estimating β using the observations $(Y_i j)_{i,j \leq N}$ and $(X_i)_{i \leq N}$.

Given that, in general, we know much more about *i.i.d.* models than about models with dyadic dependence such as the one in the equation (2.1), it would interesting to extract an *i.i.d.* sub-sample from a full sample (Y_{ij}) and (X_i). Observe that the set of edges $\{Y_{1,2}, Y_{3,4}, ..., Y_{N-1,N}\}$ (assuming N is even) are *i.i.d.*. Denoting $Y_{(i)} := Y_{2i-1,2i}$ and $X_{(i)} := (X_{2i-1}, X_{2i})$ for all $i = 1..N/2$, the observations $(Y_{(i)}, X_{(i)})_{i \leq N/2}$ become *i.i.d.* and follow:

$$Y_{(i)} = h(X_{(i)}, \epsilon_{(i)}, \beta) \tag{2.2}$$

with $\epsilon_{(i)}$ for all $i = 1..N/2$. Assuming the parameter β is identified under the model (2.2), it is also identified under (2.1). Moreover, any estimator for β with certain desirable properties in (2.2) would have those same properties under (2.1). In fact, there are many ways to extract *i.i.d.* sub-samples like the one used in (2.2): for any permutation $\sigma \in \mathbb{S}_N$, the observations $\{Y_{\sigma(2i-1),\sigma(2i)}, i = 1..N/2\}$ are *i.i.d.*.

This approach is too naive, it disregards most of the data. A more sensible estimator would be one that averages all - or a large number of - the estimators obtained through the *i.i.d.* sub-samples. This paper studies these *averaged* estimator for parametric binomial

[1] This model is in fact very general: any X-exchangeable random array would have a representation of the form (2.1). See Crane and Towsner (2018) for details.

models (e.g. logit models). I show that if the set of permutations used to extract the *i.i.d.* samples is "diverse" enough, that is, if the sub-samples do not intersect too much (in a sense that I precise in the proposition 14), then the "average MLE" has the same asymptotic distribution as the composit maximum likelihood estimator (c.f. section 4.2. in Graham (2020) for details on the composite maximum likelihood). In the next section, I formally describe the procedure, the diversity and condition and show the asymptotic distribution of the "averaged" estimators. The third section discusses an interesting application: the procedure can also be useful when the network is not observable in its entirety. The last section concludes. All the proofs are relegated to the end of the paper.

2.2 The model and the main results

Consider the model:

$$Y_{ij} = \mathbb{1}(X_{ij}\beta_0 + U_i + U_j + V_{ij} \geq 0) \tag{2.3}$$

where: $X_{ij} = g(X_i, X_j)$ with (X_i) are i.i.d. random variable, U_i and V_{ij} are i.i.d random variables with mean 0 such that $\epsilon_{ij} = U_i + U_j + V_{ij}$ is distributed following CDF Φ and PDF ϕ. β_0 is the parameter of interest, β_0 is known to be in a set $K \subset \mathbb{R}^k$.

Assume we have an even number of observations $i, j = 1..N$, I am interested the following estimator: first, for every permutation $\sigma \in \mathbb{S}_N$ consider the i.i.d. observations $\left(Y_{\sigma(2i-1),\sigma(2i)}, X_{\sigma(2i-1),\sigma(2i)}\right)_{i=1}^{\frac{N}{2}}$, to simplify, denote: $Y_{\sigma,i} := Y_{\sigma(2i-1),\sigma(2i)}$ and similarly for X. Define $\hat{\beta}_\sigma$ the maximum likelihood estimator of β_0 computed using the i.i.d. sample $(Y_\sigma, X_\sigma) = (Y_{\sigma,i}, X_{\sigma,i})$:

$$\hat{\beta}_\sigma := arg\max_\beta \sum_{i=1}^{\frac{N}{2}} Y_{\sigma,i}\log(\Phi(X_{\sigma,i}\beta)) + (1 - Y_{\sigma,i})\log(1 - \Phi(X_{\sigma,i}\beta))$$

For every σ, denote: $\mathcal{L}_\sigma(X, Y; \beta) := \frac{2}{N}\sum_{i=1}^{\frac{N}{2}} Y_{\sigma,i}\log(\Phi(X_{\sigma,i}\beta)) + (1-Y_{\sigma,i})\log(1-\Phi(X_{\sigma,i}\beta))$.
Fix some set $S \subset \mathbb{S}_N$, define:

$$\hat{\beta}_S := \frac{1}{|S|}\sum_{\sigma \in S}\hat{\beta}_\sigma$$

the objective is to determine the asymptotic distribution of $\hat{\beta}_S$.
Note that for any $\sigma \in \mathbb{S}_N$, whenever $\hat{\beta}_\sigma$ is an interior point of the parameter space K:

$$0 = \frac{\partial\mathcal{L}_\sigma(X, Y; \hat{\beta}_\sigma)}{\partial\beta} = \frac{\partial\mathcal{L}_\sigma(X, Y; \beta_0)}{\partial\beta} + \frac{\partial^2\mathcal{L}_\sigma(X, Y; \bar{\beta}_\sigma)}{\partial\beta^2}(\hat{\beta}_\sigma - \beta_0)$$

for some $\bar{\beta}_\sigma \in [\beta_0, \hat{\beta}_\sigma]$.[2] Therefore:

[2]Throughout, as I will state in each proposition, I assume the parameter space to be convex.

$$\hat{\beta}_S - \beta_0 = -\frac{1}{|S|}\sum_{\sigma\in S}\left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$
$$= -\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} \tag{2.4}$$
$$+\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$

where $\Sigma(\beta_0) := E\left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta^2}\right)$.

Before discussing the asymptotic behavior of $\hat{\beta}_S$, a few technical comments are in order. First, these Taylor expansions are only valid if *all* the $\hat{\beta}_\sigma$'s are interior points. How can we be sure they are? Second, the equation (2.4) requires that $\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}$ is invertible for any $\bar{\beta}_\sigma$ and for any σ.

The two following propositions and their corollary address these two concerns. I show that the $\hat{\beta}_\sigma$'s are not only all interior points with high probability (when the true parameter is itself an interior point), but that they are uniformly consistent as long as S does not grow too fast in N. Moreover, I show that $\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}$ converge to their common expectation uniformly both in σ and in β.

Further, these uniform convergence results will allow me to neglect the second term of the equation (2.4): $\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$, relative to its first term. That is, the asymptotic distribution of $\hat{\beta}_S$ will be that of the first term of the equation (2.4): $-\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$.

Proposition 11. *Assume*

- *K is compact and convex;*
- *X has a compact support and*
- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta\in K$, that is:*

$$\inf_{\beta\in K}\lambda_{\min}(\Sigma(\beta)) > 0$$

then, for any $\sigma\in\mathbb{S}_N$, for any $\epsilon\in\mathbb{R}_+^*$:

$$\mathbb{P}\left(\sup_{\beta\in K}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| > \epsilon\right) \leq A\exp(-BN)$$

for some constants A and positive B that depend only on K, ϵ, $||.||$ the norm chosen on the matrix space and Σ.[3]

The second proposition shows that the $\hat{\beta}'_\sigma$s are close to the true parameter with a probability that grows exponentially to 1 with N:

Proposition 12. *Under the the assumptions of proposition 11, for all $\epsilon > 0$ there exist scalars A and $B > 0$ that do not depend on N such that:*

$$\mathbb{P}\left(||\hat{\beta}_{id} - \beta_0|| > \epsilon\right) \leq A\exp(-BN)$$

where id denotes the identity permutation, i.e. $id \in \mathbb{S}_N$ with $id(i) = i$ for all $i \in N$.

That each estimator $\hat{\beta}_\sigma$ is close to the true parameter with a probability that increases this fast (exponentially) has very strong implications: if the set S is small enough (with a cardinality that grows polynomial in N), then the $\hat{\beta}_\sigma$'s are uniformly consistent *almost surely*. The following corollary shows uniform consistency in probability for any set S that grows sub-exponentially but not necessarily polinomialy!) because convergence in probability is enough for our purposes. The claim on the uniform almost sure convergence follows by Borel-Cantelli.

Corollary 7. *In addition to the assumptions of proposition 1, assume that S grows sub-exponentially, that is: $|S| = o(\exp(AN))$ for all $A \in \mathbb{R}$. Then*

$$\sup_{\sigma \in S} |\hat{\beta}_\sigma - \beta_0| \to_p 0$$

If in addition β_0 is an interior point in K, then with probability approaching 1, β_σ is in the interior of K for all $\sigma \in S$.

Now that we dealt with the technical concerns regarding the validity of the Taylor expansion in (2.4), the two following propositions look at the asymptotic distribution of each of the terms in the final formula (2.4).

Proposition 13. *Fix some (sequence) $S \subset \mathbb{S}_N$. Define*

$$C_{S,ij} := |\{\sigma \in S : \exists k : \{\sigma(2k-1), \sigma(2k)\} = \{i, j\}\}|$$

[3] In fact, what I show is that

$$\mathbb{P}\Bigg(\left[\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} \text{ is not invertible for some } \beta\right] \\ \text{OR } \left[\text{it is invertible AND } \sup_{\beta \in K} \left|\left|\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}\right|\right| > \epsilon\right]\Bigg) \\ \leq A\exp(-BN)$$

I omit this detail in the statement of the proposition to simplify the exposition.

the number of times the pair $\{i,j\}$ appears in the subset of edges in S. In addition to the assumptions of proposition 11, assume that $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$, then

$$\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial \mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$
$$\rightarrow_d \mathcal{N}\left(0, 4\times\Sigma(\beta_0)^{-1}Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)}-(1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1,U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

I will delay the discussion over the new condition in this theorem: $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ until we state the main result in this paper in proposition 14. Putting all the previous propositions together, we are now able to determine the asymptotic distribution of $\hat{\beta}_S$:

Proposition 14. *Assume that:*

- *K is compact and convex;*
- *X has a compact support and*
- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta\in K$, that is:*

$$\inf_{\beta\in K}\lambda_{\min}(\Sigma(\beta))>0$$

- *S grows sub-exponentially, that is: $|S| = o(\exp(AN))$ for all $A\in\mathbb{R}$*
- $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$

Then:

$$\sqrt{N}(\hat{\beta}_S-\beta_0)\rightarrow_d$$
$$\mathcal{N}\left(0, 4\times\Sigma(\beta_0)^{-1}Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)}-(1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1,U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

Remarks regarding the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$
First, notice that:

$$\frac{\sum_{i<j} C_{S,ij}^2}{N|S|^2} = \frac{\sum_{ij}\sum_{\sigma,\pi\in S}\mathbb{1}(i,j\in\sigma\cap\pi)}{N|S|^2} = \frac{\sum_{\sigma,\pi\in S}|\sigma\cap\pi|}{N|S|^2}$$

where I notationaly identify permutations with perfect matchings (sets of edges), so that $\sigma\cap\pi := \{\{i,j\} : \exists k,k' \text{ s.t. } \{i,j\} = \{\sigma(2k-1),\sigma(2k)\} = \{\pi(2k'-1),\pi(2k')\}\}$. The condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ is then equivalent to $\sum_{\sigma,\pi\in S}|\sigma\cap\pi| = o(N|S|^2)$. This

alternative formulation clarifies the need for the condition: it is a diversification requirement on the set S. S is not allowed to include permutations (perfect matchings) that share too many edges. Specifically, the average overlap between all the perfect matchings in S shouldn't grow faster than N.

This condition restricts the choice of the set of permutations $|S|$, for instance, $|S|$ can't be bounded (as a function of N), since for all i, j:

$$C_{S,ij}^2 \geq C_{S,ij}$$

therefore:

$$\sum_{i<j} C_{S,ij}^2 \geq \sum_{i<j} C_{S,ij} = \frac{N-1}{2}|S|$$

if $|S|$ does not go to infinity as $N \to \infty$, then the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ can't be satisfied.

On the other side, any S such that $C_{S,ij} \in \{0,1\}$ for all i, j, i.e. where each pair appears at most once, and such that $|S| \to +\infty$ as $N \to \infty$, satisfies the condition. That is because in that case: $C_{S,ij}^2 = C_{S,ij}$ for all i, j, therefore $\sum_{i<j} C_{S,ij}^2 = \sum_{i<j} C_{S,ij} = \frac{N-1}{2}|S| = o(N|S|^2)$. Such an S is always guaranteed to exist. Fix some N (even) and consider the set of permutations where I first include the identity permutation, then I "rotate" the second elements in each pair (rotate the even indices). In other words, consider the following set of permutations:

$$\begin{aligned} S := \{&(1,2,3,4,...,N-3,N-2,N-1,N); \\ &(1,4,3,6,...,N-3,N,N-1,2); \\ &(1,6,3,8,...,N-3,2,N-1,4); \\ &. \\ &. \\ &. \\ &(1,N,3,2,...,N-3,N-4,N-1,N-2)\} \end{aligned}$$

where $\sigma = (i_1, ..., i_N)$ denotes the permutation $\sigma(k) = i_k$. Notice that the odd indices (1,3, ...) do not change from one permutation to the other, whereas the even indices are rotated. In this example, $|S| = \frac{N}{2}$ and $C_{S,ij} \in \{0,1\}$ for all i, j.

For computational reasons, one would want to choose a set S that is as small as possible. Any subset of S defined above would work provided that its size explodes with N.

A weaker sufficient condition for $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ would be that each edge is allowed to be repeated in S at most $c_N = o(|S|)$. In which case, for any pair $\{i, j\}$:

$$C_{S,ij}^2 \leq c_N C_{S,ij}$$

so:

$$\sum_{i<j} C_{S,ij}^2 \leq c_N \sum_{i<j} C_{S,ij} = c_N \frac{N-1}{2}|S| = o(N|S|^2)$$

as desired.

Importantly, S can be random as long as it is independent from all other variables (X, U and V). In fact, picking a random S can relieve from the burden of verifying the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ as discussed in the following corollary.

Corollary 8. *Assume that:*

- *K is compact convex;*
- *X has a compact support and*
- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta \in K$, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

- *For all N, the perfect matchings in S are drawn uniformly with replacement from the set of perfect matchings and $|S|$ is a deterministic function of N with: $|S| = O(\log(N))$.*

Then:

$$\sqrt{N}(\hat{\beta}_S - \beta_0) \to_d$$
$$\mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

An interesting application is one where the econometrician doesn't observe the complete N-node network, but can only observe a subgraph containing only a subset of $\mathbb{S}_N$. I discuss this application in the next section.

2.3 The average estimator for networks with missing data

Assume that $Y_N = (Y_{ij})_{i,j \leq N}$ is generate following equation (2.3) as before. However, assume that now the econometrician cannot observe the entire network, instead, the econometrician observes a subgraph of Y only. Specifically, assume there is a random graph $G_N = (G_{ij})_{i,j \leq N}$ with $G_{ij} \in \{0, 1\}$ for all i, j such that

1. for all $i, j \leq N$, Y_{ij} is observed if and only if $G_{ij} = 1$ and
2. G_N and Y_N are independent.

If G_N has a set of perfect matchings S_N that meets the conditions of proposition 14 above, that is such that S_N grows sub-exponentially and $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$, then S_N can be used to construct the estimator $\hat{\beta}_{S_N}$ and the proposition 14 can be readily applied. However,

checking that such a set S_N exists is hard. To the best of my knowledge, using the fastest algorithms available, the enumeration of all the perfect matchings in G_N can performed at a time complexity $O(N) \times$ the number of perfect matchings in G_N (Uno (1997)). As in the corollary 8, we can get around this difficulty by sampling from S_N, as long as we know that $\sum_{i<j} C_{S,ij}^2 = o_p(N|S|^2)$ (there is no need to assume that $|S_N|$ grows sub exponentially).

Proposition 15. *Assume that:*

- *K is compact convex.*
- *X has a compact support.*
- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta \in K$, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

- *Assume the set S_N of all perfect matchings in G_N is such that $\sum_{i<j} C_{S_N,ij}^2 = o_p(N|S_N|^2)$.*

For all N, construct $\tilde{S}_N$, a tuple of perfect matchings uniformly drawn (with replacement) from S_N with a deterministic cardinality and $|\tilde{S}_N| \to +\infty$. Then:

$$\sqrt{N}(\hat{\beta}_{\tilde{S}} - \beta_0) \to_d$$
$$\mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1 - Y_{12})\frac{\phi(X_{12}\beta_0)}{1 - \Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

Thanks to the proposition 15, we no longer need to check that $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ conditional on G_N as suggested by the proposition 14. It is instead enough that the underlying model that generates G_N be such that $\sum_{i<j} C_{S,ij}^2 = o_p(N|S|^2)$ in probability. But how large is the class of models for G_N satisfying $\sum_{i<j} C_{S,ij}^2 = o_p(N|S|^2)$? The next proposition provides a partial answer for bipartite graphs.[4]

For the purposes of proposition 16 only, consider the following model instead of model (2.3):

$$Y_{ij} = \mathbb{1}(X_{ij}\beta_0 + U_i + W_j + V_{ij} \geq 0) \tag{2.5}$$

for $i \in N$ and $j \in M$, N and M being two sets of nodes. We assume $|N| = |M|$, a necessary condition for perfect matchings to exist. We will overload the notation N: when there is no ambiguity, it will also refer to the cardinality $|N|$.

Proposition 16. *For any $\epsilon > 0$, if G_N is drawn uniformly from the set of bipartite graphs with at least $N^{\frac{3}{2}+\epsilon}$ edges, then $\sum_{i<j} C_{S,ij}^2 = o_p(N|S|^2)$.*

[4]The model in equation (2.3) that is one for a non bi-partite graph Y_N. However, the propositions 11 to 15 would hold, under the same proofs, for the bipartite graph model (2.5).

2.4 Concluding remarks

This paper offers a systematic procedure to translate what we know about *i.i.d.* models to exchangeable array models. It could be particularly useful for models where no other estimators have been analysed. I am particularly thinking about semi-parametric models where the composite likelihood estimator is not available. The proofs for other models would be basically the same as the ones in this paper at the cost of adding some smoothness assumptions (that are satisfied by the binomial parametric model studied here and that I did not need to emphasize).

The estimator, however, is likely to be (very) inefficient. It does not exploit the dependence structure that dyadic models exhibit. That is clear in the parametric model studied in this paper: the average MLE cannot outperform the composite MLE, which also suffers from the same flaw. However, one question that I am leaving under the shadow is: what happens if we take exponentially many *i.i.d.* samples? If we do, the average MLE -were it to be well defined - would be computationally infeasible, but what would its theoretical properties be? clearly, from the proofs in this paper (again, if the MLE's are all defined and are all interior points!), the asymptotic distribution of the average MLE would be nothing like the composite maximum likelihood anymore.

Perhaps related to the question of inefficiency, the use of this procedure for data sets with missing observations could be interesting. First, it intuitively illustrates how inefficient the estimators obtained are: in the last proposition, I show that in general, for balanced bi-partite models, around $1/\sqrt{N}$th of the total number of observations (edges) is in general enough to perform like the estimator returned by the procedure if every edge were used (or like the composite MLE)! Second, the result in the last section sheds only a very dim light over the question of what observations are allowed to be missing in general graphs: non bi-partite or unbalanced bi-partite. The same proof strategy does not seem to work for other settings and I am curious to know what other models for the observable graph (G_N in the last section) would guarantee that the diversity condition on the set of all perfect matching be satisfied with high probability. Of course, allowing the observables's graph G_N to be correlated with the actual graph Y_N is yet another interesting and probably much more challenging question. I have not put enough thought towards an answer to the last question yet.

Finally, this paper did not discuss how the standard errors could be estimated. That was not my focus so far. However, given that each "averaged estimator" is computed based on a set of *i.i.d.* sub-samples that are themselves drawn *i.i.d.* uniformly from the set of available *i.i.d.* sub-samples, we end-up with a huge number of "averaged" estimators. The idea of computing the standard errors by computing multiple "averaged" estimators each on a different *i.i.d.* sub-sample, then computing an "empirical standard error" based on all these "averaged" estimators, is an appealing place to start thinking about standard error estimation.

2.5 Proofs

Proof of proposition 11

Proof. (Proposition 11) The proof follows in 3 steps:

Step 1: Prove that for any continuous function W from $support(X) \times support(Y) \times K$ into $\mathbb{R}$, with mean: $\mu(\beta) = E(W(X_{\sigma,i}, Y_{\sigma,i}, \beta))$, there are constants A' and B' such that for all $\epsilon > 0$:

$$\mathbb{P}\left(\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| > 4\epsilon\right) \leq A' \exp(-B'N)$$

with $\bar{W}(\beta) := \frac{2}{N}\sum_{i=1}^{N/2} W(X_{\sigma,i}, Y_{\sigma,i}; \beta)$

Fix $\epsilon > 0$. For any $\beta \in K$, define:
$\lambda_\delta(\beta) := E\left(\sup_{\beta':||\beta'-\beta|| \leq \delta} |W(X_{\sigma,i}, Y_{\sigma,i}, \beta) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')|\right)$ and $\delta > 0$ such that $\lambda_\delta(\beta) < \epsilon$ for all $\beta \in K$. Such δ exists because by theorem 9.1 in Keener (2010):

$$\sup_{\beta \in K} \lambda_\delta(\beta) \rightarrow_{\delta \rightarrow 0} 0$$

Since K is compact, let $(\beta_i)_{i=1..m}$ be a finite set of elements in K such that the open balls O_i centered at β_i with radius δ cover K. Following the proof of theorem 9.2 in Keener (2010), note that:

$$\begin{aligned}\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| &= \max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \mu(\beta)| \\ &\leq \max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \bar{W}(\beta_i)| + |\bar{W}(\beta_i) - \mu(\beta_i)| + |\mu(\beta_i) - \mu(\beta)|\end{aligned}$$

Note that for all i and for all $\beta \in O_i$:

$$|\mu(\beta_i) - \mu(\beta)| \leq \lambda_\delta(beta_i) \leq \epsilon$$

second, observe:

$$\bar{M}_{\delta,N}(\beta) := \frac{2}{N}\sum_{i=1}^{\frac{N}{2}} \sup_{\beta':||\beta'-\beta|| \leq \delta} |W(X_{\sigma,i}, Y_{\sigma,i}, \beta) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')|$$

and note that:

$$\begin{aligned}\max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \bar{W}(\beta_i)| &\leq \max_{i=1..m} \bar{M}_{\delta,N}(\beta_i) \\ &\leq \max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \max_{i=1..m} \lambda_\delta(\beta_i) \\ &\leq \max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \epsilon\end{aligned}$$

Therefore:

$$\sup_{\beta\in K}|\bar{W}(\beta)-\mu(\beta)| \leq \max_{i=1..m}|\bar{M}_{\delta,N}(\beta_i)-\lambda_\delta(\beta_i)| + \max_{i=1..m}|\bar{W}(\beta_i)-\mu(\beta_i)| + 2\epsilon$$

Hence:

$$\begin{aligned}
&\mathbb{P}(\sup_{\beta\in K}|\bar{W}(\beta)-\mu(\beta)| \geq 4\epsilon)\\
&\leq \mathbb{P}(\max_{i=1..m}|\bar{M}_{\delta,N}(\beta_i)-\lambda_\delta(\beta_i)| + \max_{i=1..m}|\bar{W}(\beta_i)-\mu(\beta_i)| + 2\epsilon \geq 4\epsilon)\\
&\leq \mathbb{P}(\max_{i=1..m}|\bar{M}_{\delta,N}(\beta_i)-\lambda_\delta(\beta_i)| + \max_{i=1..m}|\bar{W}(\beta_i)-\mu(\beta_i)| \geq 2\epsilon)\\
&\leq \mathbb{P}(\max_{i=1..m}|\bar{M}_{\delta,N}(\beta_i)-\lambda_\delta(\beta_i)|| \geq \epsilon) + \mathbb{P}(\max_{i=1..m}|\bar{W}(\beta_i)-\mu(\beta_i)| \geq \epsilon)\\
&\leq m \times \big(\mathbb{P}(|\bar{M}_{\delta,N}(\beta_1)-\lambda_\delta(\beta_1)|| \geq \epsilon) + \mathbb{P}(|\bar{W}(\beta_1)-\mu(\beta_1)| \geq \epsilon)\big)
\end{aligned}$$

By the compactness of $support(X) \times support(Y) \times K$ and the continuity of W,

$$(W(X_{\sigma,i}, Y_{\sigma,i}, \beta_1) - \mu(\beta_1))_i$$

and

$$\sup_{\beta':||\beta'-\beta_1||\leq\delta}|W(X_{\sigma,i}, Y_{\sigma,i}, \beta_1) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')| - \lambda_\delta(\beta_1))_i$$

are i.i.d. and bounded, Hoeffding's inequality allows to conclude.

Step 2: Show that for any $\sigma \in \mathbb{S}_N$, for any $\epsilon \in \mathbb{R}_+^*$ there exist constants some constants A'' and B'' that depend only on K and ϵ

$$\mathbb{P}\left(\sup_{\beta\in K}||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}|| > 4\epsilon\right) \leq A''\exp(-B''N)$$

To see that, it is enough to apply the result in step 1 element wise on the matrix $\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}$ then use a union bound to obtain the desired result for the max norm $||.||_\infty$.

Step 3: Show the final result.

Note that for any β, and any given σ (using a sub-multiplicative matrix norm this time):

$$\begin{aligned}
||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| &\leq ||\Sigma(\beta)^{-1}|| \times ||\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| \times ||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}||\\
&\leq \frac{||\Sigma(\beta)^{-1}||}{\lambda_{\min}(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2})} \times ||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}||
\end{aligned}$$

where $\lambda_{\min}(.)$ returns the smallest eigen value. Take some $x \in \mathbb{R}^k$ such that $||x|| = 1$ and $x'\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}x = \lambda_{\min}(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2})$, then :

$$\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq x'\Sigma(\beta)x - x'\left(\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}\right)x$$
$$= x'\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}x$$

implying:[5]

$$\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq \lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2})$$

under the event: $\sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| < \inf_\beta \lambda_{\min}(\Sigma(\beta))$ so:

$$\begin{aligned}
&||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| \\
&\leq \frac{||\Sigma(\beta)^{-1}||}{\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||} \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \\
&\leq \frac{\sup_\beta\{||\Sigma(\beta)^{-1}||\}}{\inf_\beta\{\lambda_{\min}(\Sigma(\beta))\} - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||} \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||
\end{aligned}$$

Therefore for any $\epsilon > 0$, there exists a function that only depends on epsilon $\gamma(\epsilon) > 0$ such that, under the event under the event $E_N := \sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| < \inf_\beta \lambda_{\min}(\Sigma(\beta))$ we have

$$\sup_\beta ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| \geq 4\epsilon \Rightarrow \sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \geq \gamma(\epsilon)$$

[5] Note here that we could similarly show:

$$\lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq \lambda_{\min}(\Sigma(\beta))$$

implying that:

$$||\lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}) - \lambda_{\min}(\Sigma(\beta))|| \leq ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||$$

and leading to the result alluded to in a footnote to the proposition's statement:

$$\mathbb{P}\left(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} \text{ is not invertible for some } \beta\right) \leq A'' \exp(-B''N)$$

for some generic A'', $B'' > 0$ that are independent of N.

then:

$$\mathbb{P}\left(\sup_{\beta}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| \geq 4\epsilon\right)$$
$$= \mathbb{P}\left(\sup_{\beta}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| \geq 4\epsilon; E_N\right)$$
$$+ \mathbb{P}\left(\sup_{\beta}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}|| \geq 4\epsilon; not(E_N)\right)$$
$$\leq \mathbb{P}\left(\sup_{\beta}||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}|| \geq \gamma(\epsilon); E_N\right)$$
$$+ \mathbb{P}(not(E_N))$$
$$\leq \mathbb{P}\left(\sup_{\beta}||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}|| \geq \gamma(\epsilon)\right)$$
$$+ \mathbb{P}\left(\sup_{\beta}||\Sigma(\beta) - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}|| \geq \inf_{\beta}\lambda_{\min}(\Sigma(\beta))\right)$$

which allows to conclude by step 2. □

Proof of proposition 12

Proof. Note that:

$$\hat{\beta}_{id} - \beta_0 = \left(\frac{\partial^2\mathcal{L}_{id}(X,Y;\bar{\beta}_{id})}{\partial\beta^2}\right)^{-1}\frac{\partial\mathcal{L}_{id}(X,Y;\beta_0)}{\partial\beta}$$
$$= -\Sigma(\bar{\beta}_{id})^{-1}\frac{\partial\mathcal{L}_{id}(X,Y;\beta_0)}{\partial\beta} + \left(\Sigma(\bar{\beta}_{id})^{-1} - \left(\frac{\partial^2\mathcal{L}_{id}(X,Y;\bar{\beta}_{id})}{\partial\beta^2}\right)^{-1}\right)\frac{\partial\mathcal{L}_{id}(X,Y;\beta_0)}{\partial\beta}$$

Thanks to the compactness of the support of X and of K, $\frac{\partial\mathcal{L}_{id}(X,Y;\beta_0)}{\partial\beta}$ is bounded by some constant M and $\beta \to \Sigma(\beta)^{-1}$ is bounded by some constant L.

Hence:

$$||\hat{\beta}_{id} - \beta_0|| \leq L||\frac{\partial\mathcal{L}_{id}(X,Y;\beta_0)}{\partial\beta}|| + M\sup_{\beta\in K}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}||$$

Applying the Hoeffding bound to the first term and proposition 11 to the second, we obtain the desired result. □

Proof of corollary 7

Proof. Fix $\epsilon > 0$.

$$\mathbb{P}(\sup_{\sigma \in S} |\hat{\beta}_\sigma - \beta_0| > \epsilon) \leq |S|\mathbb{P}(|\hat{\beta}_{id} - \beta_0| > \epsilon)$$

and the proposition (12) completes the proof. □

Proof of proposition 13

Proof. Fix some $S \subset \mathbb{S}_N$ and some $\lambda \in \mathbb{R}^k$. I want to determine the asymptotic distribution of:

$$\begin{aligned}
&\lambda' \frac{1}{|S|} \Sigma(\beta_0)^{-1} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} \\
&= \lambda' \frac{1}{|S| N/2} \sum_{\sigma \in S} \sum_{i=1}^{\frac{N}{2}} \left(Y_{\sigma,i} \frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)} - (1 - Y_{\sigma,i}) \frac{\phi(X_{\sigma,i}\beta_0)}{1 - \Phi(X_{\sigma,i}\beta_0)} \right) \Sigma(\beta_0)^{-1} X_{\sigma,i} \\
&=: \frac{1}{|S| N/2} \sum_{\sigma \in S} \sum_{i=1}^{\frac{N}{2}} f(X_{\sigma,i}, Y_{\sigma,i})
\end{aligned}$$

where $f(X_{\sigma,i}, Y_{\sigma,i}) := \left(Y_{\sigma,i} \frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)} - (1 - Y_{\sigma,i}) \frac{\phi(X_{\sigma,i}\beta_0)}{1 - \Phi(X_{\sigma,i}\beta_0)} \right) \lambda' \Sigma(\beta_0)^{-1} X_{\sigma,i}$. Although f depends on λ and β, they are omitted to simplify the notation. We can rearrange:

$$\lambda' \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} = \frac{1}{|S| N/2} \sum_{i<j} C_{S,ij} f(X_{i,j}, Y_{i,j})$$

with $C_{S,ij} := |\{\sigma \in S : \exists k : \{\sigma(2k-1), \sigma(2k)\} = \{i, j\}\}|$, the number of times the pair $\{i, j\}$ appears in the subset of observations generated by S. Observe that for all i, j, by definition: $C_{S,ij} = C_{S,ji}$ and $C_{S,ii} = 0$. Also note that for all i:

$$\sum_{j=1}^{N} C_{S,ij} = |S|$$

since every pair appears exactly once per permutation $\sigma \in S$. Denote:

$$q(X_i, X_j, U_i, U_j) := E(f(X_{i,j}, Y_{i,j}) | X_i, X_j, U_i, U_j)$$

$$h(X_i, U_i) := E(q(X_i, X_j, U_i, U_j) | X_i, U_i)$$

and

$$\tilde{q}(X_i, X_j, U_i, U_j) := q(X_i, X_j, U_i, U_j) - h(X_i, U_i) - h(X_j, U_j)$$

where it is assumed that $C_{S,ii} = 0$ for all i and $C_{S,ij} = C_{S,ji}$ for all i and j. Observe, following O'Neil and Redner (1993), that

$$\begin{aligned}\sum_{i<j} C_{S,ij} q(X_i, X_j, U_i, U_j) &= \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + \sum_i \left(\sum_{j=1}^{N} C_{S,ij}\right) h(X_i, U_i) \\ &= \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + |S| \sum_i h(X_i, U_i)\end{aligned}$$

So:

$$\begin{aligned}\lambda' \Sigma(\beta_0)^{-1} \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \frac{1}{|S|N/2} \sum_{i<j} C_{S,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j)\right) \\ &\quad + \frac{1}{|S|N/2} \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + \frac{2}{N} \sum_i h(X_i, U_i)\end{aligned} \tag{2.6}$$

We have:

$$\begin{aligned}Var\left(\sum_{i<j} C_{S,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j)\right)\right) &= Var\left(f(X_{1,2j}, Y_{1,2}) - q(X_1, X_2, U_1, U_2)\right) \\ &\quad \times \sum_{i<j} C_{S,ij}^2 \\ Var\left(\sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j)\right) &= Var\left(\tilde{q}(X_1, X_2, U_1, U_2)\right) \sum_{i<j} C_{S,ij}^2\end{aligned} \tag{2.7}$$

Assuming :

$$\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$$

then:

$$\begin{aligned}\sqrt{N} \lambda' \Sigma(\beta_0)^{-1} \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \frac{2}{\sqrt{N}} \sum_i h(X_i, U_i) + o_p(1) \\ \to_d \mathcal{N}(0, 4Var(h(X_1, U_1)))\end{aligned} \tag{2.8}$$

but

$$\begin{aligned}&Var(h(X_1, U_1)) \\ &= \lambda' \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12} \frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0} - (1 - Y_{12}) \frac{\phi(X_{12}\beta_0)}{1 - \Phi(X_{12}\beta_0} |X_1, U_1\right)\right) \Sigma(\beta_0)^{-1} \lambda\end{aligned} \tag{2.9}$$

therefore, using the wold device:

$$\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$
$$\to_d \mathcal{N}\left(0, 4\times\Sigma(\beta_0)^{-1}Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1,U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

□

Proof of proposition 14

Proof. (Proposition 14)

Remember:

$$\begin{aligned}&\hat{\beta}_S - \beta_0\\ &= -\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\\ &\quad + \frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\end{aligned}$$

First, I show that:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} = o_p(1)$$

Note that because K is compact and $\beta \to \Sigma(\beta)^{-1}$ is conitnuously differentiable, then $\beta \to \Sigma(\beta)^{-1}$ is Lipschitz on K, let η be the Lipschitz constant.

$$\begin{aligned}&\left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|\\ &\leq \left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\bar{\beta}_\sigma)^{-1} - \Sigma(\beta_0)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|\\ &+ \left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\bar{\beta}_\sigma)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|\\ &\leq \left(\eta\sup_{\sigma\in S}||\hat{\beta}_\sigma - \beta_0|| + \sup_{\beta\in K}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}||\right)\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|\end{aligned}$$

First, note:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| = O_p(1)$$

because:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|$$
$$\leq \sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}\right|\right|$$

and

$$Var\left(\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}\right|\right|\right)$$
$$=\frac{N}{|S|^2}[|S|Var\left(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}\right|\right|\right)$$
$$+|S|(|S|-1)Cov(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}\right|\right|,$$
$$\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma',i}\frac{\phi(X_{\sigma',i}\beta_0)}{\Phi(X_{\sigma',i}\beta_0)}-(1-Y_{\sigma',i})\frac{\phi(X_{\sigma',i}\beta_0)}{1-\Phi(X_{\sigma',i}\beta_0)}\right)X_{\sigma',i}\right|\right|)]$$
$$\leq NVar\left(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}\left|\left|\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}\right|\right|\right)$$
$$=4Var\left(\left|\left|\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)}-(1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}\right)X_{12}\right|\right|\right)$$

By proposition 11:

$$\sup_{\beta\in K}||\Sigma(\beta)^{-1}-\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}||=o_p(1)$$

and

$$\sup_{\sigma\in S}||\hat{\beta}_\sigma-\beta_0||=o_p(1)$$

since:

$$\mathbb{P}(\sup_{\sigma\in S}||\hat{\beta}_\sigma-\beta_0||>\epsilon)\leq|S|\mathbb{P}(||\hat{\beta}_{id}-\beta_0||>\epsilon)\to 0$$

where the limit is obtained by proposition 12 and by the assumption that $|S| = o(\exp(AN))$ for all $A \in \mathbb{R}$.

Finally:

$$\sqrt{N}(\hat{\beta}_S - \beta_0) = -\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma \in S}\frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} + o_p(1) \tag{2.10}$$

and proposition 13 allows to conclude.

□

Proof of corollary 8

Proof. (Corollary 8.) Assume that for all N, S (in fact, S_N) is constructed by drawing permutations (or perfect matchings) with replacement from the set of perfect matchings. Denote $C_{ij,N}$ the number of perfect matchings in which the pair i, j appears in the set S_N. Let c_N be a deterministic sequence such that $c_N \to +\infty$ and $c_N = o(|S|)$.
Define the events $E_N := \{C_{ij,N} > c_N \text{ for some pair } i, j\}$. Then:

$$\begin{aligned}
\mathbb{P}(E_N) &\leq \sum_{ij} \mathbb{P}(C_{ij,N} > c_N) \\
&= \frac{N(N-1)}{2}\mathbb{P}(C_{12,N} > c_N) \\
&= \frac{N(N-1)}{2}\sum_{k=c_N+1}^{|S|}\mathbb{P}(C_{12,N} = k)
\end{aligned}$$

Let σ be the random variable corresponding to a single uniform draw from the set of all perfect matchings (i.e. permutations in $\mathbb{S}_N$). For any fixed pair i, j we have:[6]

$$\mathbb{P}(i, j \in \sigma) = \frac{\frac{(N-2)!}{(N/2-1)!2^{N/2-1}}}{\frac{N!}{(N/2)!2^{N/2}}} = \frac{2N/2}{N(N-1)} = \frac{1}{N-1}$$

so

$$\mathbb{P}(C_{12,N} = k) = \binom{|S|}{k}\left(\frac{1}{N-1}\right)^k\left(\frac{N-2}{N-1}\right)^{|S|-k} \leq \binom{|S|}{k}\left(\frac{1}{N-1}\right)^{c_N+1}$$

[6]I abuse notation here: $i, j \in \sigma$ means that the pair i, j forms an edge in the perfect matching σ, or in the language of permutations that there exists some k such that $\{\sigma(2k-1), \sigma(2k)\} = \{i, j\}$.

and

$$\begin{aligned}\mathbb{P}(E_N) &\leq \frac{N(N-1)}{2}\left(\frac{1}{N-1}\right)^{c_N+1}\sum_{k=c_N+1}^{|S|}\binom{|S|}{k} \\ &\leq \frac{N(N-1)}{2}\left(\frac{1}{N-1}\right)^{c_N+1} \times 2^{|S|} \\ &= O\left(\left(\frac{1}{N}\right)^{C_N-2}\right) \\ &= o(\frac{1}{N^2})\end{aligned}$$

Therefore

$$\sum_N \mathbb{P}(E_N) < +\infty$$

By the Borel–Cantelli lemma:

$$\mathbb{P}(\limsup E_N) = \mathbb{P}(\cap_{N\geq 1} \cup_{k=N}^{\infty} E_k) = 0$$

or equivalently:

$$\mathbb{P}(\exists N_0\ \forall N > N_0\ \forall i,j \leq N :\ C_{ij,N} \leq c_N) = 1$$

as shown earlier, if for all pairs i, j $C_{ij,N} \leq c_N = o(|S|)$, then:

$$\sum_{i<j} C_{ij,N}^2 \leq c_N \sum_{i<j} C_{ij,N} = c_N \times \frac{N-1}{2}|S| = o(N|S|^2)$$

hence with probability one, the condition: $\sum_{i<j} C_{ij,N}^2 = o(N|S|^2)$ is satisfied.[7] The rest of the proof for proposition 14 follows. □

Proof of proposition 15

Proof. Of proposition 15 Given the conditions of the proposition, denote $\tilde{\mathbb{P}}$ the probability conditional on S_N. By defintion:

$$\tilde{C}_{\tilde{S}_N,ij} := \sum_{\sigma\in\tilde{S}_N} \mathbb{1}(ij \in \sigma)$$

the terms in this sum are *i.i.d.* conditional on S_N, because the perfect matchings in $\tilde{S}_N$ are *i.i.d.*, therefore:

[7] Here I showed that $\frac{C_{ij,N}^2}{N|S|^2} \to 0$ almost surely. In fact, it was enough to show convergence in probability since that is enough to obtain equation 2.10 and conclude.

$$\tilde{\mathbb{E}}(\tilde{C}_{\tilde{S}_N,ij}) = |\tilde{S}_N|\frac{C_{S_N,ij}}{|S_N|}$$

and

$$\tilde{\mathbb{V}}(\tilde{C}_{\tilde{S}_N,ij}) = |\tilde{S}_N|\frac{C_{S_N,ij}}{|S_N|}\left(1-\frac{C_{S_N,ij}}{|S_N|}\right)$$

then

$$\tilde{\mathbb{E}}(\tilde{C}^2_{\tilde{S}_N,ij}) = |\tilde{S}_N|\frac{C_{S_N,ij}}{|S_N|} + (|\tilde{S}_N|^2 - |\tilde{S}_N|)\frac{C^2_{S_N,ij}}{|S_N|^2}$$

and:

$$\begin{aligned}\tilde{\mathbb{E}}\left(\frac{\sum_{i<j}\tilde{C}^2_{\tilde{S}_N,ij}}{N|\tilde{S}_N|^2}\right) &= \frac{\sum_{i<j}C_{S_N,ij}}{N|\tilde{S}_N|\times|S_N|} + \left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\\ &= \frac{1}{|\tilde{S}_N|} + \left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\end{aligned}$$

Remember equation (2.6):

$$\begin{aligned}\lambda'\Sigma(\beta_0)^{-1}\frac{1}{|\tilde{S}_N|}\sum_{\sigma\in\tilde{S}_N}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} &= \frac{1}{|\tilde{S}_N|N/2}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j}) - q(X_i,X_j,U_i,U_j)\right)\\ &+ \frac{1}{|\tilde{S}_N|N/2}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\tilde{q}(X_i,X_j,U_i,U_j) + \frac{2}{N}\sum_i h(X_i,U_i)\end{aligned}$$

equation (2.7) becomes:

$$\begin{aligned}&Var\left(\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j}) - q(X_i,X_j,U_i,U_j)\right)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)\\ &\quad= Var\left(f(X_{1,2j},Y_{1,2}) - q(X_1,X_2,U_1,U_2)\right)\sum_{i<j}\tilde{C}^2_{\tilde{S}_N,ij}\\ &Var\left(\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\tilde{q}(X_i,X_j,U_i,U_j)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)\\ &\quad= Var\left(\tilde{q}(X_1,X_2,U_1,U_2)\right)\sum_{i<j}\tilde{C}^2_{\tilde{S}_N,ij}\end{aligned}$$

then

$$Var[\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j))|(S_N)_{N\geq 0}]$$
$$= Var\left(f(X_{1,2j},Y_{1,2})-q(X_1,X_2,U_1,U_2)\right)\tilde{\mathbb{E}}\left(\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}^2_{\tilde{S}_N,ij}\right)$$
$$+\tilde{\mathbb{V}}\left(E\left(\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j)\right)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)\right)$$
$$= Var\left(f(X_{1,2j},Y_{1,2})-q(X_1,X_2,U_1,U_2)\right)\left(\frac{1}{|\tilde{S}_N|}+\left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\right)$$

where the second equality results from the observation that for all i,j

$$E\left(\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j)\right)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)=0$$

likewise:

$$Var[\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\tilde{q}(X_i,X_j,U_i,U_j)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}]$$
$$= Var\left(\tilde{q}(X_1,X_2,U_1,U_2)\right)\left(\frac{1}{|\tilde{S}_N|}+\left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\right)$$

so equation (2.8) still holds, conditionally on $(S_N)_{N\geq 0}$ this time:

$$\sqrt{N}\lambda'\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}=\frac{2}{\sqrt{N}}\sum_i h(X_i,U_i)+o_p(1)$$
$$\to_d \mathcal{N}(0,4Var(h(X_1,U_1)))$$

by dominated convergence, equation (2.8) also holds unconditionally. The rest of the argument for propositions 13 and 14 follows. □

Proof of proposition 16

Proof. Proposition 16.
The proof is for some fixed $\epsilon>0$.

First, without loss of generality, I assume that the sequence of graphs G_N are independent. Otherwise, I would work on another sequence (a coupling) $(G'_N)_N$ such that $G_N =_d G'_N$ and the G'_N are independent. In that case, $\frac{\sum_{i<j}C^2_{S',ij}}{N|S'|^2} =_d \frac{\sum_{i<j}C^2_{S,ij}}{N|S|^2}$ (the first ratio is computed on

G'_N and the second on G_N), therefore proving the proposition for G'_N implies that it also holds for G_N.

Let $e(G_N)$ denote the number of edges of the graph G_N. I will show the result conditionally on the sequence $(e(G_N)_N)_{N\geq 0}$, then proposition 16 will follow by dominated convergence. For the rest of the proof, the "ambient" probability is that conditional on $(e(G_N)_N)_{N\geq 0}$: I will omit to condition by $(e(G_N)_N)_{N\geq 0}$ in my notation. Further, I will use the notation e_N for $e(G_N)$.

Note that, conditional on $e(G_N)$, G_N is uniformly drawn from the set of graphs with exactly $e(G_N)$ edges.

First, I show that as $N \to +\infty$:

$$\begin{aligned} E(S_N) &\sim N! \left(\frac{e_N}{N^2}\right)^N \exp\left(-\frac{1}{2}\left(\frac{N^2}{e_N}-1\right)\right) \\ E(S_N^2) &\sim E(S_N)^2 \\ E(C_{S_N,ij}^2) &\sim (N-1)!^2 \exp\left(1-\frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1} \end{aligned} \tag{2.11}$$

The first two statements result immediately from the theorems 1 and 2 in O'Neil (1970) (cf. the section 8.1 in Lovász and Plummer (2009) for details about the link between perfect matchings in a bipartite graph and the permanent of its bi-adjacency matrix). The proof for $E(C_{S_N,ij}^2)$ follows similar steps as those of the proofs for theorems 1 and 2 in O'Neil (1970). As in O'Neil (1970), denote, for any permutation $\sigma \in \mathbb{S}_N$,

$$x_\sigma := 1\left\{(i,\sigma(i)) \text{ is an edge in } G_N \text{ for all } i \in N\right\}$$

and for any integers M and $k \leq M$, define:

$$B_k^M := \{(\sigma,\pi) \in \mathbb{S}_M^2 : |\{i \in M : \pi(i) = \sigma(i)\}| = k\}$$

that is, if we identify every permutation in $\mathbb{S}_M$ to a perfect matching between two sets of carnality M each, then B_k^M is the set of perfect matching pairs that have exactly k edges in common.

By definition:

$$C_{S_N,ij} = \sum_{\sigma\in\mathbb{S}_N:\sigma(i)=j} x_\sigma$$

and:

$$C_{S_N,ij}^2 = \sum_{\sigma,\pi\in\mathbb{S}_N:\sigma(i)=\pi(i)=j} x_\sigma x_\pi$$

so

$$E(C_{S_N,ij}^2) = \sum_{k=1}^{N} |B_{k-1}^{N-1}| \mathbb{P}(x_\sigma x_\pi = 1 | (\pi,\sigma) \in B_k^N)$$

the equation (1.9) in O'Neil (1970) yields:

$$|B_k^N| = \frac{N!^2}{k!}e^{-1}\left(1 + O(\frac{1}{(N-k+1)!})\right)$$

and for $k \leq k_1 := [N^{5/8}]$, equation (1.14) in O'Neil (1970):

$$\begin{aligned}&\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B_k^N)\\ &= \left(\frac{e_N}{N^2}\right)^{2N-k} \exp\left(-2\left(1-\frac{k}{N}\right)^2\left(\frac{N^2}{e_N}-1\right)\right)\left(1+O(N^{-1/4-\epsilon})+O(N^{-2\epsilon})\right)\end{aligned}$$

hence

$$\begin{aligned}&E(C^2_{S_N,ij})\\ &= (N-1)!^2\exp\left(1-\frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N}\sum_{k=1}^{k_1}\frac{1}{(k-1)!}\left[\frac{N^2}{e_N}\exp\left(4(N/e_N-1/N)\right)\right]^k\\ &\times\left(1+O(N^{-1/4-\epsilon})+O(N^{-2\epsilon})\right)+\sum_{k=k_1+1}^{N}|B_{k-1}^{N-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B_k^N)\\ &= (N-1)!^2\exp\left(1-\frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N}\times\left[\frac{N^2}{e_N}\exp\left(4(N/e_N-1/N)\right)\right]\\ &\times\sum_{k=0}^{\infty}\frac{1}{(k-1)!}\left[\frac{N^2}{e_N}\exp\left(4(N/e_N-1/N)\right)\right]^k\times(1+o(1))\times\left(1+O(N^{-1/4-\epsilon})+O(N^{-2\epsilon})\right)\\ &+\sum_{k=k_1+1}^{N}|B_{k-1}^{N-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B_k^N)\\ &= (N-1)!^2\exp\left(1-\frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N}\times\left[\frac{N^2}{e_N}\exp\left(4(N/e_N-1/N)\right)\right]\\ &\times\exp\left[\frac{N^2}{e_N}\exp\left(4(N/e_N-1/N)\right)\right]\times(1+o(1))\\ &+\sum_{k=k_1+1}^{N}|B_{k-1}^{N-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B_k^N)\\ &= (N-1)!^2\exp\left(1-\frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1}\times(1+o(1))\\ &+\sum_{k=k_1+1}^{N}|B_{k-1}^{N-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B_k^N)\end{aligned}$$

noting that $|B_k^N| \leq \frac{N!^2}{k!}$ (from equation 1.8 in O'Neil (1970)), we have:

$$\sum_{k=k_1+1}^{N} |B_{k-1}^{N-1}| \mathbb{P}(x_\sigma x_\pi = 1 | (\pi, \sigma) \in B_k^N) \leq \sum_{k=k_1+1}^{N} \frac{(N-1)!^2}{(k-1)!} \left(\frac{e_N}{N^2}\right)^{2N-k}$$
$$= (N-1)!^2 \left(\frac{e_N}{N^2}\right)^{2N-1} \times O(N^{-(1/8)N^{5/8}})$$

finally:

$$E(C_{S_N,ij}^2) \sim (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1}$$

Given the asymptotic results in (2.11), I can now show that:

$$\frac{|S_N|}{E(|S_N|)} \to_p 1 \tag{2.12}$$

indeed, for any $\epsilon > 0$:

$$\mathbb{P}\left(|\frac{|S_N|}{E(|S_N|)} - 1| > \epsilon\right) = \mathbb{P}\left(\frac{||S_N| - E(|S_N|)|}{E(|S_N|)} > \epsilon\right)$$
$$\leq \frac{Var(|S_N|)}{\epsilon^2 E(|S_N|)^2}$$
$$= \frac{E(|S_N|^2) - E(|S_N|)^2}{\epsilon^2 E(|S_N|)^2}$$
$$\to 0$$

where the inequality is Markov's and where the limit is obtained thanks to the equation (2.11).

Observe that:

$$\frac{E(\sum_{ij} C_{S_N,ij}^2)}{NE(|S_N|^2)} \sim \frac{N^2 \times (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1}}{N \times N!^2 \left(\frac{e_N}{N^2}\right)^{2N} \exp\left(1 - \frac{N^2}{e_N}\right)}$$
$$= \frac{N}{e_N}$$
$$\to 0$$

therefore:

$$\frac{\sum_{ij} C_{S_N,ij}^2}{NE(|S_N|^2)} \to_p 0$$

then the equation (2.12) gives:

$$\frac{\sum_{ij} C_{S_N,ij}^2}{N|S_N|^2} \to_p 0$$

as desired. □

Chapter 3

Asymptotic efficiency of the maximum likelihood estimator for exponential families

Introduction

For parametric models on i.i.d. data, the asymptotic properties of the maximum likelihood estimator are know at least since the work of Le Cam (1953) LeCam (1953) (Stigler (2007)Stigle (2007) for a historical note). Under some regularity conditions, the maximum likelihood estimator is known to be consistent, asymptotically normal and efficient within a class of regular estimators (see for instance chapters 7 and 8 in van der Vaart (1998) van der Vaart (1998)). These results then allowed for the computation of semi-parametric efficiency bounds for models on i.i.d. data (Chamberlain (1987) Chamberlain (1987b) and Newey (1990) Newey (1990)). Apart from extensions to some time series models (Chamberlain (1987) Chamberlain (1987b) and Newey (1990) Newey (1990)), very little is known about efficiency (both parametric and semi-parametric) or the asymptotic properties of the maximum likelihood estimator when the i.i.d. assumption is relaxed. The difficulty when trying to relax independence is that the space of possibilities in terms of the dependence patterns the data could follow is huge, so huge that it is hard to imagine universal and unifying results such as those we know for independent data.

In this work, I study a specific class of models, the exponential family models, with a focus on exponential random graph models. These models are appealing in at least two regards. First, they impose enough structure for us to be able to establish fairly general results and, at the same time, they are general enough to include various models of interest. In fact, the most general version of the exponential random graph models (corresponding to a data generating process of the form in equation (3.4) discussed below) includes the multinomial models which are known to approximate arbitrarily well any well behaved model (Chamberlain (1987) Chamberlain (1987b)).

Second, exponential random graphs are widely used to model social networks. They are particularly suitable to reproduce the distribution of a given statistic (or set of statistics) that is observed in the network being modeled. For instance, assume the analyst is interested in modeling a network that is known to have a number of triangles following, say, a Poisson distribution with parameter λ (to be estimated). Further, the analyst assumes that the observed network is drawn uniformly conditionally on the number of triangles, that is, all the networks with the same number of triangles are equally likely.[1] Then this data generating process corresponds to an exponential random graph model with a sufficient statistic equal to the number of triangles and with parameter λ. To look at a real world example, in many contexts, social networks follow have degree distribution that follows a power law (see for instance Stephen and Toubia (2009) Stephen and Toubia (2009) for social commerce networks), a random graph model that could reproduce this feature is an exponential random graph model where the degree distribution is drawn first following a power law with a certain parameter, then the network is generated following some given distribution conditional on the degree sequence generated in the previous stage.[2,3]

Finally, for some choices of the sufficient statistic, exponential random graph models can be micro-founded. Mele (2017) Mele (2017) presents a model of network formation where N agents, picked sequentially randomly, meet at each period and decide whether they want to send a link to each other or not. Agent i's utility from a graph with adjacency matrix g and playing with $N-1$ other agents is given by:

$$U_i(g, X, \theta) = \sum_{j=1}^{N} g_{ij} u(X_i, X_j, \theta_u) + \sum_{j=1}^{N} g_{ij} g_{ji} m(X_i, X_j, \theta_m) + \sum_{j=1}^{N} g_{ij} \sum_{k=1, \neq i,j}^{N} g_{jk} v(X_i, X_k, \theta_v) + \sum_{j=1}^{N} g_{ij} \sum_{k=1, \neq i,j}^{N} g_{ki} w(X_k, X_j, \theta_w) \tag{3.1}$$

where $\theta = (\theta_u, \theta_m, \theta_v, \theta_w)$; $(X_i)_i$ are covariates. To the expression of the utility in 3.1, a different extreme value type 1 error term is added in every period where the agent gets to decide. The shock terms are assumed to be i.i.d across agents and across time.

The terms in 3.1 can be seen as the utilities from direct links, mutual links, indirect links and "popularity", respectively.

Mele (2017) makes 3 additional assumptions: i) when making the decision, agents do not take the other agents' future actions into account, but only the state of the graph at the moment they are making their decision to pair with a given player or not; ii) the meeting technology is such that the probability of two agents meeting in a given period does not

[1] In fact, the conditional distribution could be any distribution that does not depend on the parameter λ and the unconditional distribution would still be exponential.

[2] To be precise: the degrees of the nodes in a given network can't be indenpendent, since the degree sequence would need to be "graphic". One still has to determine the joint distribution of the degree sequence that yields power law marginals.

[3] For a more detailed introduction to exponential random graph models: Snijders (2002) Snijders (2002).

depend on whether they are currently linked or not; and iii) the indirect and mutual links terms in the utility expression are covariate-symmetric, that is, for all i, j:

$$m(X_i, X_j, \theta_m) = m(X_j, X_i, \theta_m)$$

$$w(X_i, X_j, \theta_v) = v(X_i, X_j, \theta_v)$$

Given these assumptions, Mele (2017) (theorem 1) shows that the network formation game converges to a unique stationary distribution over graphs

$$\pi(g, X, \theta) := \frac{\exp[Q(g, X, \theta)]}{\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X, \theta)]}$$

$\mathcal{G}$ being the set of all directed graphs of N agents and Q is the game's potential function, defined by:

$$Q(g, X, \theta) \\ := \sum_{i=1}^{N} \sum_{j=1}^{N} g_{ij} u(X_i, X_j, \theta_u) + \sum_{i=1}^{N} \sum_{j>i}^{N} g_{ij} g_{ji} m(X_i, X_j, \theta_m) + \sum_{j=1}^{N} g_{ij} \sum_{k=1, \neq i,j}^{N} g_{jk} v(X_i, X_k, \theta_v)$$

In particular, if the utility components u, v m are linear in θ, the unique stationary distribution becomes exponential:

$$\pi(g, X, \theta) := \frac{\exp[\theta' t(g, X)]}{\sum_{\omega \in \mathcal{G}} \exp[\theta' t(\omega, X)]} \tag{3.2}$$

For another illustration micro-founding ERGMs with a different set of sufficient statistics, see Chandrasekhar and Jackson (2012)Chandrasekhar and Jackson (2012).

This prospectus is organized as follows: the model is presented and discussed in section 1. In section 2, I show that, under some sufficient conditions, the MLE is asymptotically normal then I discuss few applications in section 3. Section 4 establishes the parametric efficiency result, which holds under the same conditions as in section 3. Finally, in section 5, I discuss a possible extension to an example of a more general exponential random graph model then the one studied in sections 2 to 4.

3.1 The model

Assume that for all N, $Y_N = (Y_{ij})_{i<j \leq N}$ is drawn from an exponential family with the same k-dimensional parameter η. That is, assume that for all N, the density of $(Y_{ij})_{\{i,j\} \leq N}$ can be expressed:

$$f_N(Y_N; \eta) = h_N(Y_N) e^{T_N(Y_N)'\eta - B_N(\eta)} \tag{3.3}$$

for some functions $h_N : \mathbb{Y}_N \to \mathbb{R}_+^*$, $h_N > 0$ a.s., $T_N : \mathbb{Y}_N :\to \mathbb{R}^k$ and $B_N : \mathbb{R}^k \to \mathbb{R}$; $\mathbb{Y}_n$ being the support of Y_N.

The model in 3.3 is restrictive in two main respects. First, I am assuming that the parameter η remains unchanged as the network size increases. A more general specification would be:

$$f_N(Y_N;\eta) = h_N(Y_N)e^{T_N(Y_N)'A_N(\eta)-B_N(\eta)} \tag{3.4}$$

for some function $A_n : \mathbb{R}^k \to \mathbb{R}^k$. Indeed, the specification in (3.4) accounts for some very important cases that cannot be modeled through (3.3). For instance, if $(Y_{ij})_{ij}$ follows a multinomial exchangeable distribution, then we know that the density of $(Y_{ij})_{ij}$ can be written in form (3.4) and not (3.3). The multinomial distribution is by itself enough to motivate the interest in (3.4): multinomial distributions can approximate arbitrarily well other distributions that are not within the exponential family, opening the possibility to extend some asymptotic or efficiency results on the exponential families to more general families. However, the results shown in this note heavily rely on the assumption $A_N(\eta) = \eta$ for all η. The proofs presented here don't go through for the more general models (3.4) and would most likely require that one makes heavy assumptions on A_N and uses different technology from what I use here; this is work in progress.

A second - perhaps more worrying - concern with the model in (3.3) is its potential inconsistency: by assuming that Y_{N+1} and Y_N are both distributed following (3.3), the distribution of Y_N ought to be obtained by marginalizing (or projecting) that of Y_{N+1}. Shalizi and Rinaldo (2013) Shalizi and Rinaldo (2013) characterize the constraints imposed on the model (3.3) in order for it to be consistent. These constraints (on the sufficient statistic $T_N(Y_N)$) turn out to be very restrictive and exclude some of the specifications of interest in econometrics (for instance, Shalizi and Rinaldo (2013) Shalizi and Rinaldo (2013) show that, if $T_N(Y_N)$ is, say, the number of tryads or stars in the graph Y_N, then consistency is violated, to mention just two important examples). Shalizi and Rinaldo (2013) Shalizi and Rinaldo (2013) conclude that "*Since these models are not projective, however, it is impossible to improve parameter estimates by getting more data, since parameters for smaller sub-graphs just cannot be extrapolated to larger graphs (or vice versa).*"[4]

Although Shalizi and Rinaldo's point in the preceding quote does make a lot of sense, one can still use exclusively the assumption that Y_N is drawn from the model (3.3), without assuming consistency, and the results would still be mathematically valid. The proofs of the results in this note do not involve marginalization. In addition, Shalizi and Rinaldo's argument does not account for instances where the distribution in (3.3) results from some underlying micro-founded model. Mele (2017) Mele (2017)'s model (discussed in more detail above) is a good illustration. In Mele's model, the observed network is assumed to be drawn from the stationary (asymptotic) distribution of some sequential network formation game (asymptotic in the sense that the game's stages repeat "infinitely" many times, but keeping the number of agents fixed). There is no reason why the (stationary) graph distribution for the game with N players should be consistent with the graph distribution obtained in a game with $N + 1$ players. Mele's model does not impose any consistency restrictions on the choice

[4]"Projective"= "consistent"; as I loosely defined it earlier.

of sufficient statistic $T_N(Y_N)$. Other examples can be found in Chandrasekhar and Jackson (2012) Chandrasekhar and Jackson (2012).

One last remark to connect these two concerns: Shalizi and Rinaldo's conclusion holds true for models in the form (3.3), but it is silent about more general models in the form (3.4).

3.2 The asymptotic distribution of the maximum likelihood estimator

Under the specification in (3.3), the log-likelihood function can be written:

$$l_n(Y_N;\eta) = \log(h_N(Y_n)) + T_N(Y_N)'\eta - B_N(\eta)$$

Assuming $B_N(.)$ is twice continuously differentiable - which would be the case for instance if $T_N(Y_N)$ has its first 2 moments, then $\hat{\eta}$, the MLE estimator of η, is characterized by:

$$\frac{\partial B_N(\hat{\eta})}{\partial \eta} = T_N(Y_N) \tag{3.5}$$

and the fisher information matrix is given by:

$$I_N(\eta) := E\left[\left(T_N(Y_N) - \frac{\partial B_N(\eta)}{\partial \eta}\right)'\left(T_N(Y_N) - \frac{\partial B_N(\eta)}{\partial \eta}\right)\right] = \left[\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right]$$

Proposition 1. *Denote $k := dim(\eta)$. Assume that:*

1. *$I_N(\eta) = K_N\Gamma_N(\eta)K_N'$, Where $\Gamma_N(.)$ is bounded function such that: $0 < \Gamma_N(\eta) < b$ for some $b > 0$ and $\inf_n \inf_{\eta\in K}\Gamma_N(\eta) > 0$ for any compact K and K_N is an invertible matrix such that: $K_N^{-1} = O(1)$.*

2. *For any compact set K and for any $j \leq k$:*

$$\sup_{\eta,\tilde{\eta}\in K} \frac{||\frac{\partial^3 B_N(\eta)}{\partial \eta_j \partial^2 \eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial \eta^2}||^{3/2}} = o(1)$$

3. *$\hat{\eta}$ is tight.*

then:

$$I_N(\eta)^{1/2}(\hat{\eta} - \eta) \to_d N(0, I_k) \tag{3.6}$$

where I_k is the identity matrix of dimension k.

The K_N in the first condition can be seen as a matrix of rates of convergence. For the i.i.d. case, K_N is simply $\sqrt{N} \times I_k$. For non i.i.d. settings, different components of the MLE estimator can converge at different rates, one example is the dyadic Gaussian linear regression discussed in the next section. The conditions imposed on $\Gamma_N(.)$ are simply to guarantee that K_N fits with this interpretation: for instance, if $\Gamma_N(.)$ converges to 0 as N grows, then the actual rate of convergence would be slower than K_N in the representation: $I_N(\eta) = K_N \Gamma_N(\eta) K_N'$, the same remark applies if $\Gamma_N(.)$ diverges to ∞, i.e. if some of its eigenvalues diverges to infinity. That is why I would need the eigenvalues of $\Gamma_N(.)$ to be bounded away from 0 and $+\infty$.

To gain some intuition around the role of the second condition, note that the moments of the score function map one to one to the differentials of B_N:

$$Var(s_N) = \frac{\partial^2 B_N(\eta)}{\partial \eta^2}$$

and:

$$\mathbb{E}(s_{N,i} s_{N,j} s_{N,k}) = \frac{\partial^3 B_N(\eta)}{\partial \eta_i \partial \eta_j \partial \eta_k}$$

Therefore, the second condition assumes that the third moment of the score grows at a higher speed than the second moment so that, asymptotically, only the first two moments matter, as if the case for the normal distribution. Showing the asymptotic linearity of the score is first step towards showing the final result and does in fact not require the last condition.

Finally, the third condition is important in performing a "delta-method like" manoeuvre, which constitutes the final step of the proof.

Before discussing some examples, note that the first two conditions are always satisfied for exponential i.i.d. models. This remark yields the following corollary:

Corollary. *For i.i.d. exponential models, if $\hat{\eta}$ is tight then:*

$$I_N(\eta)^{1/2}(\hat{\eta} - \eta) \to_d N(0, I_k) \tag{3.7}$$

in particular, $\hat{\eta}$ is consistent.

The proof of proposition 1 comes in two steps. First, I show that the score has an asymptotically normal distribution, precisely:

$$I_N(\eta)^{-1/2} s_N(\eta) \to_d N(0, I_k)$$

this is done by showing the convergence of the moment generating function of the left hand-side to that of the (multivariate) standard normal. Note that:

$$
\begin{aligned}
M_N(\phi;\eta) :&= E(e^{\phi' I_N(\eta)^{-1/2} s_N(\eta)}) \\
&= \int h_N(y) \exp\left(\phi' I_N(\eta)^{-1/2}(T_N(y) - \frac{\partial B_N(\eta)}{\partial \eta}) + T_N(y)'\eta - B_N(\eta)\right) dy \\
&= e^{-\frac{1}{2}\phi'\phi} \int h_N(y) \exp\left((\eta+\epsilon_N)' T_N(y) - B_N(\eta+\epsilon_N)\right) \exp(-\text{Remainder of order 3}) \\
&= e^{-\frac{1}{2}\phi'\phi} \exp(-\text{Remainder of order 3}) \int h_N(y) \exp\left((\eta+\epsilon_N)' T_N(y) - B_N(\eta+\epsilon_N)\right) \\
&\approx e^{-1/2\phi'\phi}
\end{aligned}
$$

where $\epsilon_N := \phi' I_N(\eta)^{-1/2}$. The rest of the argument consists of showing that the last approximation holds true asymptotically, that is, showing that the remainder vanishes to 0. This is one place where the second hypotheses of the proposition is required.

The second part of the argument starts by noting that:

$$
s_N(\eta) = \frac{\partial B_N(\hat{\eta})}{\partial \eta} - \frac{\partial B_N(\eta)}{\partial \eta}
$$

then using a ”delta method-like” argument to show:

$$
\left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{1/2} (\hat{\eta} - \eta) = \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta) + o_p(1)
$$

and conclude. Please refer to the appendix for a detailed formal proof.

3.3 Examples

The linear regression with Gaussian residuals

Let $(Y_{ij})_{i<j\leq N}$ be a random variable such that:

$$
Y_{ij} = X_{ij}\beta + A_i + A_j + V_{ij}
$$

with:

- β is a $K \times 1$-dimensional parameter;
- $(X_{ij})_{ij}$ is a random vector such that $X_{ij} = X_{ji}$ for all i, j;
- $A_i \sim_{i.i.d} N(0,1)$;
- $V_{ij} = V_{ji} \sim_{i.i.d} N(0,1)$;
- $(X_{ij})_{i<j\leq N}$, $(A_i)_{i\leq N}$ and $(V_{ij})_{i<j\leq N}$ are independent of each other.

Consider the function $n : (i, j) \to \frac{(j-2)(j-1)}{2} + i$ for all $i < j$, note that n is one to one from $\{(i, j) : i, j \in \mathbb{N}^+ \text{ and } i < j\}$ into $\mathbb{N}^+$. Consider the following stacked vectors/matrices:

- $Y_N = (Y_i)_{i \leq \frac{N(N-1)}{2}}$, where $Y_i = Y_{n^{-1}(i)}$ for all i;
- $X_N = (X_1', ..., X_{\frac{N(N-1)}{2}}')'$ is a $\frac{N(N-1)}{2} \times K$ matrix, where $X_i = X_{n^{-1}(i)}$.[5]

Note that:

$$Y_N - X_N\beta \sim N(0, \Omega_N)$$

where $\Omega_N := (\omega_{m,n})_{m,n \leq \frac{N(N-1)}{2}}$ and, using the same abuse of notation above:

$$\omega_{ij,kl} := \begin{cases} 3 & \text{if } \{i,j\} = \{k,l\} \\ 1 & \text{if } |\{i,j\} \cap \{k,l\}| = 1 \\ 0 & \text{otherwise} \end{cases}$$

It is easy to check that the density of Y_N (conditional on X) can be written as a member of the family (3.3) with:

$$B_N(\beta) = -\frac{1}{2}\beta' X_N' \Omega_N X_N \beta$$

Therefore:

$$I_N(\beta) = X_N' \Omega_N X_N$$

and:

$$\frac{\partial^3 B_N(\beta)}{\partial \beta^3} = 0 = o(I_N(\beta)^{3/2})$$

Hence, applying proposition 1:

$$(X_N' \Omega_N X_N)^{1/2}(\hat{\beta}_{ML} - \beta) \to N(0, I_k)$$

In fact, in this easy case, this result could have been obtained without using the proposition, simply by observing that:

$$\hat{\beta}_{ML} = (X_N' \Omega_N^{-1} X_N)^{-1}(X_N' \Omega_N^{-1} Y_N) \sim_{|X} N(\beta, (X_N' \Omega_N^{-1} X_N)^{-1})$$

To take a concrete example, consider the example where:

$$X_{ij} = (1, X_i X_j)$$

[5] This is to say that I am ordering the pairs i, j with $i < j$ in a "lexicographic-like" order: $(1, 2) \to 1$, $(1, 3) \to 2$, $(2, 3) \to 3$, $(1, 4) \to 4$, $(2, 4) \to 5$, $(3, 4) \to 6$... This is only one labeling among others and this choice plays no role in what follows.

for some i.i.d random variables $(X_i)_{1\leq i\leq N}$. That is, we are considering the regression model:

$$Y_{ij} = \eta_1 + \eta_2 X_i X_j + U_i + U_j + V_{ij}$$

where the U_i's and V_{ij}'s satisfy the same set of conditions as above.

After some tedious computations (detailed in the appendix), I find:

$$X_N' \Omega_N^{-1} X_N = \begin{pmatrix} N\left(\frac{1}{4} + o_p(1)\right) & N\left(\frac{E(X_1)^2}{4} + o_p(1)\right) \\ N\left(\frac{E(X_1)^2}{4} + o_p(1)\right) & N^2\left(Var(X_1)^2 + o_p(1)\right) \end{pmatrix}$$

which can be decomposed:

$$X_N' \Omega_N^{-1} X_N = K_N \times \begin{pmatrix} \frac{1}{4} + o_p(1) & \frac{1}{\sqrt{N}}\left(\frac{E(X_1)^2}{4} - \frac{1}{4} + o_p(1)\right) \\ \frac{1}{\sqrt{N}}\left(\frac{E(X_1)^2}{4} - \frac{1}{4} + o_p(1)\right) & N^2\left(Var(X_1)^2 + o_p(1)\right) \end{pmatrix} \times K_N'$$

where:

$$K_N := \begin{pmatrix} \sqrt{N} & 0 \\ \sqrt{N} & N \end{pmatrix}$$

Assuming $E(X_1) \neq 0$ and $Var(X_1) \neq 0$, K_N is a rate matrix satisfying the conditions of proposition 1. Other choices of the rate matrix are possible. Trivially, αK_N, for any $\alpha \neq 0$ is another rate matrix.

Here, since K_N is lower triangular, the diagonal terms have a simple interpretation. They represent uni-dimentional convergence rates of the individual components of $\hat{\eta}$. That is, in this example, the intercept is converging at a rate of $\sqrt{N}$, the usual convergence rate in *i.i.d.* models, whereas the slope (coefficient on $X_{ij} = X_i X_j$) converges at rate N.

Inference on the Erdos-Renyi graph

Let's now look at inference on few examples of Erdos Renyi graphs. These are random graphs such that a link between any two nodes forms with a given probability p_N, independently of all the other links in the graph. The probability p_N is allowed to depend on the number of nodes N. The independence between edges immensely simplifies the computations and allows for many tractable illustrations.

An Erdos-Renyi graph with probability of link formation p_N is drawn from a distribution with probability mass function:

$$f_N(y) = \exp\left(\sum_{ij} y_{ij}\log(p_N) + (1-y_{ij})\log(1-p_N)\right)$$
$$= \exp\left(\sum_{ij} y_{ij}\log(\frac{p_N}{1-p_N}) + \log(1-p_N)\right)$$

In the following three examples, I look at three specific cases: one where $p_N = p$, independent of N, which corresponds to an $i.i.d$ Bernouilli model. This example will simply confirm that the result of proposition 1 extends the results on maximum likelihood estimation in the $i.i.d.$ setting. The second example corresponds to the more popular Erdos-Renyi graph model with $p_N = \frac{p}{N}$.[6] The value of p is this second case plays a very important role in determining the shape of the graph as N grows to infinity: it is well known (see van der Hofstad (2016) Hofstad (2016), chapter 4) that there is a regime change at $p = 1$, if $p < 1$, then asymptotically, the graph is almost surely composed of small (size of $O(\log(N))$) disconnected subgraphs and if $p > 1$, then we get a giant component along with smaller components of size $O(\log(N))$. The question of testing whether p is larger or smaller than one is of great importance. The last example will be a less common choice of p_N with the single purpose of illustrating the use of proposition 1 to show the inconsistency of a maximum likelihood estimator.

Example 1:

Assume the link between nodes i and j forms following the probability $p = \frac{e^\eta}{1+e^\eta}$. The model is then of the form in equation (3.3):

$$f(y) = \exp(\sum_{ij} y_{ij}\eta - B_N(\eta))$$

with:

$$B_N(\eta) = \log(\sum_{\omega\in\mathcal{G}} \exp[\eta t(\omega)])$$
$$= \log\left(\sum_{k=0}^{N(N-1)/2} \binom{N(N-1)}{k} e^{\eta k}\right)$$
$$= \log\left((1+e^\eta)^{N(N-1)/2}\right)$$

[6] In fact, I will be looking at a very close model: $p_N = \frac{p}{N+p}$. The usual Erdos Renyi model with $p = \frac{p}{N}$ is not an Exponential random Graph model in the sense of equation (3.3).

$$= \frac{N(N-1)}{2} \log(1 + e^{\eta})$$

The information matrix is given by:

$$I_N(\theta) = \frac{\partial^2 B_N}{\partial \eta^2}(\theta) = \frac{N(N-1)}{2} \frac{e^{\eta}}{(1 + e^{\eta})^2}$$

It is easy to check that the condition of the proposition 1 is satisfied:

$$\frac{\partial^3 B_N}{\partial \eta^3}(\tilde{\eta}) = O_N(N^2) = o(I_N(\eta)^{3/2})$$

for any η and $\tilde{\eta}$.

The maximum likelihood estimator $\hat{\eta}_{ML}$ is given by:

$$\hat{\eta}_{ML} = \log\left(e^{\frac{2t(Y_N)}{N(N-1)}} - 1\right)$$

where Y_N here is the observed adjacency matrix.

By the proposition 1:

$$I_N(\eta)^{1/2}(\hat{\eta}_{ML} - \eta) \to_d N(0,1)$$

or:

$$N(\hat{\eta}_{ML} - \eta) \to_d N\left(0, \frac{(1 + e^{\eta})^2}{e^{\eta}}\right)$$

Example 2:

Consider a graph with N nodes with a probability of link formation equal to:

$$p_N := \frac{p}{N+p}$$

The probability mass function of the model can be expressed:

$$\begin{aligned} f_N(y) &= \exp\left(\sum_{ij} y_{ij} \log(\frac{p}{N}) - \log(\frac{N}{N+p})\right) \\ &= \exp\left(-\sum_{ij} y_{ij} \log(N)\right) \exp\left((\sum_{ij} y_{ij}) \log(p) - \frac{N(N-1)}{2} \log(\frac{N}{N+p})\right) \end{aligned}$$

So by re-parametrizing: $\eta := \log(p)$, we get a model of the form in equation (3.3) with:

$$T_N(y) = \sum_{ij} y_{ij}$$
$$B_N(\eta) = \frac{N(N-1)}{2}\log(\frac{N}{N+e^\eta}) = \frac{N(N-1)}{2}\log(N) - \frac{N(N-1)}{2}\log(N+e^\eta)$$

we can compute:

$$B'_N(\eta) = \frac{N(N-1)}{2}\frac{e^\eta}{N+e^\eta}$$
$$B''_N(\eta) = I_N(\eta) = \frac{N(N-1)}{2}\frac{Ne^\eta}{(N+e^\eta)^2}$$
$$B'''_N(\eta) = \frac{N(N-1)}{2}\frac{N^2e^\eta - Ne^{2\eta}}{(N+e^\eta)^3}$$

so:

$$I_N(\eta)^{-1} = O(\frac{1}{N}) = O(1)$$

also, for any compact K:

$$\sup_{\eta,\tilde{\eta}\in K}\frac{||\frac{\partial^3 B_N(\eta)}{\partial\eta_j\partial^2\eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial\eta^2}||^{3/2}} = O(\frac{1}{N^{3/2}}) = o(1)$$

finally, the MLE estimator can be expressed:

$$\hat{\eta}_{MLE} = \log\left(\frac{N\bar{y}}{1-\bar{y}}\right)$$

To see that $\hat{\eta}_{MLE}$ is tight, it is enough to show that $\frac{N\bar{y}}{1-\bar{y}}$ converges in probability to a positive number. Note that:

$$E(\bar{y}) = \frac{e^\eta}{(e^\eta+N)} \to 0$$

$$Var(\bar{y}) = \frac{2}{N(N-1)}\frac{Ne^\eta}{(e^\eta+N)^2} \to 0$$

so

$$\bar{y} \to_p 0$$

also:

$$E(N\bar{y}) \to e^\eta$$

$$Var(N\bar{y}) = \frac{2N^2}{N(N-1)}\frac{Ne^\eta}{(e^\eta+N)^2} \to 0$$

this allows to conclude that the MLE estimator is consistent:

$$\hat{\eta}_{MLE} \to_p \eta$$

it is therefore tight. Proposition 1 gives:

$$\sqrt{\frac{N(N-1)}{2}\frac{Ne^\eta}{(N+e^\eta)^2}}(\hat{\eta}_{MLE} - \eta) \to_d \mathcal{N}(0,1)$$

or:

$$\sqrt{N}(\hat{\eta}_{MLE} - \eta) \to_d \mathcal{N}(0, \frac{2}{e^\eta})$$

Note that this model could be extended to allow for the probability of link formation to depend on covariates that are specific to the pair (i, j). That is, a more general model would be one where a link between nodes i and j forms with probability:

$$\mathbb{P}(y_{ij} = 1|X) = \frac{\phi(X_{ij}\beta)}{N + \phi(X_{ij}\beta)}$$

for some known function ϕ, where I condition on $X = (X_{ij})_{ij}$. Links form independently conditionally on X.

For this model to be an Exponential Random Graph Model, one has to set: $\phi(x) = e^x$ for all x. The function B_N then becomes:

$$B_N(\eta, X_N) := B_N(\eta)\frac{N(N-1)}{2}\log(N) - \sum_{ij}\log(N + e^{X_{ij}\beta})$$

Example 3:

Assume that the random graph is drawn from a distribution with probability mass function:

$$f(y) = \exp\left(\frac{\sum_{ij} y_{ij}}{N}\eta - B_N(\eta)\right)$$

This corresponds to an Erdos-Renyi model with probability of link formation:

$$\mathbb{P}(y_{ij} = 1) = \frac{e^{\eta/N}}{1 + e^{\eta/N}}$$

$$\begin{aligned} B_N(\eta) &= \frac{N(N-1)}{2}\log(1+e^{\eta/N}) \\ B_N'(\eta) &= \frac{N-1}{2}\frac{e^{\eta/N}}{1+e^{\eta/N}} \\ B_N''(\eta) &= I_N(\eta) = \frac{N-1}{2N}\frac{e^{\eta/N}}{(1+e^{\eta/N})^2} = O(1) \\ B_N'''(\eta) &= \frac{N-1}{2N^2}\frac{e^{\eta/N}(1-e^{\eta/N})}{(1+e^{\eta/N})^3} \end{aligned}$$

Hence, for any compact K:

$$\sup_{\eta,\tilde{\eta}\in K} \frac{||\frac{\partial^3 B_N(\eta)}{\partial\eta_j\partial^2\eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial\eta^2}||^{3/2}} = o(1/N) = o(1)$$

The expression of the MLE estimator:

$$\hat{\eta}_{MLE} = N\log\left(\frac{\bar{y}}{1-\bar{y}}\right)$$

I need to show that $\hat{\eta}_{MLE}$ is tight, note:

$$\bar{y} \to_p \frac{1}{2}$$

and that:

$$\hat{\eta}_{MLE} = N\frac{\log(1+\frac{2\bar{y}-1}{1-\bar{y}})}{\frac{2\bar{y}-1}{1-\bar{y}}}\frac{2\bar{y}-1}{1-\bar{y}}$$

since:

$$\begin{aligned} 1-\bar{y} &\to_p \frac{1}{2} \\ \frac{\log(1+\frac{2\bar{y}-1}{1-\bar{y}})}{\frac{2\bar{y}-1}{1-\bar{y}}} &\to_p 1 \\ Var(N(2\bar{y}-1)) = 4N^2Var(\bar{y}) &= \frac{8N^2}{N(N-1)}\frac{e^{\eta/N}}{(1+e^{\eta/N})^2} \to 2 \end{aligned}$$

proposition 1 yields:

$$\frac{1}{2}(\hat{\eta}_{MLE}-\eta) \to_d \mathcal{N}(0,1)$$

Interestingly, this shows that MLE is not consistent. In this example, any consistent estimator is more efficient, in a sense that I will precise shortly, than the MLE. The result on efficiency in the next section will take this finding even further to show that there exists no consistent estimator for η in this model.

The fact that the probability of link formation in converging to $\frac{1}{2}$ as $N \to \infty$ does not explain this phenomenon. There is nothing particular about $\frac{1}{2}$, I could have the probability converge to any number $p := \frac{c}{1+c}$strictly between 0 and 1 (i.e. $c > 0$, known), by setting:

$$\mathbb{P}(y_{ij} = 1) = \frac{ce^{\eta/N}}{1 + ce^{\eta/N}}$$

which corresponds to the model:

$$f(y) = \exp\left(\frac{\sum_{ij} y_{ij}}{N} \log(c)\eta - B_N(\eta)\right)$$

assuming c is know.

Note that this amounts to a rescaling of the initial model, the same conclusions follow.

3.4 The parametric efficiency of MLE for model (3.3)

Let K_N be any given sequence of invertible square matrices of dimension equal to $dim(\eta)$.[7] I will adopt the same definition of a (K_N -) regular estimator as the one presented in Hajèk (1970) Hájek (1970) :

Definition. (K_N - regular estimator) An estimator η_N is said to be regular if there exists a distribution $L_\eta(.)$ such that for any $h \in \mathbb{R}^k$:

$$\mathbb{P}_{\eta + K_N^{-1}h}(K_N(\eta_N - \eta) - h \leq v) \to L_\eta(v)$$

Before asking whether the MLE is efficient, let's first show that it is regular. The following lemma provides a necessary and sufficient condition for the ML estimator of a model satisfying the condition of proposition 1 to be regular for some rate matrix K_N:

Lemma:

[7] So far, I have assumed the density function in equation (3.3) was defined for any η. If η were constrained to belong to some set E, then K_N should be such that for any h, the perturbed parameter $\eta + K_N^{-1}h$ be in E for N large enough. That would for instance be the case if E is open and the minimal eigen value of $K_N K_N'$ goes to $+\infty$.

Assume that the assumptions of proposition 1 hold, let K_N be a sequence of invertible symmetric matrices. Then $\hat{\eta}_{ML}$ is K_N-regular if and only if for all η there exists a symmetric definite positive $\Gamma(\eta)$ such that:

$$K_N^{-1}\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N^{-1} \to \Gamma(\eta)$$

Proof. Assume η_{ML} is K_N regular. We know, by proposition 1, that:

$$\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{1/2}(\eta_{ML}-\eta) \to_d \mathcal{N}(0, I_k)$$

Then, using the definition of regularity with $h = 0$:

$$K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{-1/2}\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{1/2}(\eta_{ML}-\eta) = K_N(\eta_{ML}-\eta) \to_d L_\eta$$

First, this implies that $K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{-1/2}$ is bounded (otherwise L_η would be degenerate, can be checked by assuming the maximum eigen value of $K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{-1/2}$ diverges then using the characterization of convergence in distribution in terms of pointwise convergence of characteristic functions). Then assume $K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)^{-1/2}$ has a subsequence that converges to some $\Sigma(\eta)$, then:

$$L_\eta \sim_d \mathcal{N}(0, \Sigma(\eta)\Sigma(\eta)')$$

Therefore $K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N$ converges to $\Sigma(\eta)\Sigma(\eta)'$ for any converging subsequence. Given that $K_N\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N$ is bounded, then it converges to $\Sigma(\eta)\Sigma(\eta)'$ along all subsequences.

Conversely, assume that for all η there exists a symmetric definite positive $\Gamma(\eta)$ such that:

$$K_N^{-1}\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N^{-1} \to \Gamma(\eta).$$

I want to show that:

$$\mathbb{E}_{\eta+K_N^{-1}h}(e^{\phi' K_N(\hat{\eta}_{ML}-\eta-K_N^{-1}h)}) \to_N e^{\frac{1}{2}\phi'\Gamma(\eta)^{-1}\phi}$$

For that, note that:

$$\begin{aligned}
&\mathbb{E}_{\eta+K_N^{-1}h}(e^{\phi' K_N(\hat{\eta}_{ML}-\eta-K_N^{-1}h})\\
&= e^{-\phi' h}\int e^{\phi' K_N(\hat{\eta}_{ML}-\eta)} h_N(y) e^{T_N(y)(\eta+K_N^{-1}h)-B_N(\eta+K_N^{-1}h)}dy\\
&= e^{-\phi' h}\int e^{\phi' K_N(\hat{\eta}_{ML}-\eta)} h_N(y) e^{T_N(y)\eta-B_N(\eta)} e^{T_N(y)K_N^{-1}h)-(B_N(\eta+K_N^{-1}h)-B_N(\eta))}dy\\
&= e^{-\phi' h} e^{-(B_N(\eta+K_N^{-1}h)-B_N(\eta)-\frac{\partial B_N}{\partial \eta}(\eta))}\\
&\quad\times\int e^{\phi' K_N(\hat{\eta}_{ML}-\eta)} h_N(y) e^{T_N(y)\eta-B_N(\eta)} e^{(T_N(y)-\frac{\partial B_N}{\partial \eta}(\eta))' K_N^{-1}h)}dy\\
&= e^{-\phi' h} e^{-(B_N(\eta+K_N^{-1}h)-B_N(\eta)-\frac{\partial B_N}{\partial \eta}(\eta)K_N^{-1}h)}\\
&\quad\times\int \exp\left(\phi' K_N(\hat{\eta}_{ML}-\eta)+(T_N(y)-\frac{\partial B_N}{\partial \eta}(\eta))' K_N^{-1}h\right) h_N(y) e^{T_N(y)\eta-B_N(\eta)}dy\\
&= e^{-\phi' h} e^{-(B_N(\eta+K_N^{-1}h)-B_N(\eta)-\frac{\partial B_N}{\partial \eta}(\eta)K_N^{-1}h)}\\
&\quad\times\int \exp\left(\phi' K_N(\hat{\eta}_{ML}-\eta)+h' K_N^{-1}s_N(\eta)\right) h_N(y) e^{T_N(y)\eta-B_N(\eta)}dy
\end{aligned}$$

Denote: $K_N^{-1}\left(\frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N^{-1} =: \Gamma_N(\eta)$; and note from the proof of proposition 1:

$$K_N(\hat{\eta}_{ML}-\eta) = \Gamma_N(\eta)^{-1}K_N^{-1}s_N(\eta)+o_p(1)$$

hence:

$$\begin{aligned}
\mathbb{E}_{\eta+K_N^{-1}h}(e^{\phi' K_N(\hat{\eta}_{ML}-\eta-K_N^{-1}h}) &= e^{-\phi' h} e^{-(B_N(\eta+K_N^{-1}h)-B_N(\eta)-\frac{\partial B_N}{\partial \eta}(\eta)K_N^{-1}h)}\\
&\quad\times \mathbb{E}_\eta\left(\exp((\phi'\Gamma_N(\eta)^{-1}+h')K_N^{-1}s_N(\eta)\right)
\end{aligned}$$

Note that (shown in the proof of proposition 1):

$$K_N^{-1}s_N(\eta)\to_d \mathcal{N}(0,\Gamma(\eta))$$

So:

$$\phi'\Gamma_N(\eta)^{-1}+h')K_N^{-1}s_N(\eta)\to_d \mathcal{N}(0,(\phi'\Gamma_N(\eta)^{-1}+h')\Gamma(\eta)(\phi'\Gamma_N(\eta)^{-1}+h')')$$

and:

$$\begin{aligned}
\mathbb{E}_\eta\left(\exp((\phi'\Gamma_N(\eta)^{-1}+h')K_N^{-1}s_N(\eta)\right) &\to_N \exp\left(\frac{1}{2}(\phi'\Gamma_N(\eta)^{-1}+h')\Gamma(\eta)(\phi'\Gamma_N(\eta)^{-1}+h')'\right)\\
&= \exp\left(\frac{1}{2}[\phi'\Gamma_N(\eta)\phi+2\phi' h+h'\Gamma_N(\eta)^{-1}h]\right)
\end{aligned}$$

on the other side:

$$B_N(\eta + K_N^{-1}h) - B_N(\eta) - \frac{\partial B_N}{\partial \eta}(\eta)K_N^{-1}h = \frac{1}{2}h'K_N^{-1}\frac{\partial^2 B_N}{\partial \eta^2}(\eta)K_N^{-1}h + \frac{1}{6}\sum_{j=1}^{k} h'K_N{-}1\frac{\partial^3 B_N}{\partial \eta^2 \partial \eta_j}(\bar{\eta})K_N^{-1}h(K_N^{-1}h)_j$$

for some $\bar{\eta}$ between η and $\eta + K_N^{-1}h$. By assumption, since $K_N^{-1} = O(1)$, there exists a compact K such that $\bar{\eta} \in K$ for all N. Therefore, by proposition 1's assumptions:

$$B_N(\eta + K_N^{-1}h) - B_N(\eta) - \frac{\partial B_N}{\partial \eta}(\eta)K_N^{-1}h = \frac{1}{2}h'\Gamma(\eta)h + o(1)$$

Finally:

$$\mathbb{E}_{\eta+K_N^{-1}h}(e^{\phi' K_N(\hat{\eta}_{ML}-\eta-K_N^{-1}h)}) \to_N e^{\frac{1}{2}\phi'\Gamma(\eta)^{-1}\phi}$$

which shows that the MLE is regular with:

$$L_\eta =_d \mathcal{N}(0, \Gamma(\eta)^{-1})$$

□

We can now state the second result:

Proposition 2.
Assume that the model (3.3) satisfies the condition of proposition 1, then the maximum likelihood estimator is efficient within the class of K_N regular estimators.

Proof. The likelihood ratio (following the notation in Hajèk (1970) Hájek (1970)):

$$\begin{aligned} r_N(h, Y_N) &= \exp\left(h'K_N^{-1}T_N(\eta) - (B_N(\eta + K_N^{-1}h) - B_N(\eta))\right) \\ &= \exp\left(h'K_N^{-1}T_N(\eta) - h'K_N^{-1}\frac{\partial B_N}{\partial \eta}(\eta) - \frac{1}{2}h'K_N^{-1}\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta})K_N^{-1}h\right) \end{aligned} \tag{3.8}$$

where $\bar{\eta}$ is some vector such that $\bar{\eta}_i$ is between η_i and $(\eta + K_N^{-1}h)_i$ for all indices i.
Hence:

$$\begin{aligned} r_N(h, Y_N) = \exp[h'K_N^{-1}(T_N(\eta) - \frac{\partial B_N}{\partial \eta}(\eta)) - \frac{1}{2}h'\Gamma(\eta)h - \frac{1}{2}h'\left(K_N^{-1}\frac{\partial^2 B_N}{\partial \eta^2}(\eta)K_N^{-1} - \Gamma(\eta)\right)h \\ - \frac{1}{2}h'K_N^{-1}\left(\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta}) - \frac{\partial^2 B_N}{\partial \eta^2}(\eta)\right)K_N^{-1}h] \end{aligned} \tag{3.9}$$

By the proposition's first assumption:

$$\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta}) - \frac{\partial^2 B_N}{\partial \eta^2}(\eta) \to 0$$

and by the second assumption:

- the maximum likelihood estimator is K_N- regular; and
- $K_N^{-1} \frac{\partial^2 B_N}{\partial \eta^2}(\eta) K_N^{-1} - \Gamma(\eta) \to 0$

Therefore by the Hajèk- Le Cam convolution theorem (Hajèk (1970) Hájek (1970)), β_{ML} is efficient (within the class of K_N-regular estimators).

□

The fact that this efficiency property, along with the result on asymptotic normality (proposition 1), allow for "constant rates of convergence" yields interesting conclusions when the information matrix is bounded. Take the example 3 of the Erdos-Renyi random graph model discussed in the previous section. I have shown that in that example the MLE is inconsistent. Note that any consistent estimator is 1-regular, therefore, MLE is efficient compared to 1-regular estimators, i.e. there can't be any consistent estimator for η in that example.

3.5 An example of model (3.4): a simple binomial

In this section, I look at a basic example of a model that falls under the specification in equation (3.4), that is in the form:

$$f_N(Y_N; \eta) = h_N(Y_N) e^{T_N(Y_N)' A_N(\eta) - B_N(\eta)}$$

with $A_N(.) \neq id(.)$. As I mentioned earlier, this more general exponential model allows for many specifications of interest including in particular the multinomial.

Assume:

$$Y_{ij} := 1\left(a + U_i + U_j + V_{ij} \geq 0\right) \tag{3.10}$$

for some $i.i.d$ standard normal U_i and V_{ij}. I am interested in estimating a.

To see that the model is of the form (3.4), note that the probability mass function (Y_{ij}) can be expressed:

$$f_N(y; a) = e^{T_N(y)' A_N(a) - B_N(a)}$$

with:

- $T_N(y)$ is a $2^{\frac{N(N-1)}{2}} \times 1$ vector that has zeroes everywhere except in the line corresponding to the value of y

- $A_N(a) = (\log(\mathbb{P}_a(Y = y_1)), \log(\mathbb{P}_a(Y = y_1)), ..., \log(\mathbb{P}_a(Y = y_{\frac{N(N-1)}{2}})))'$
- $B_N(a) = 0$ for all a.

The maximum likelihood estimator cannot be expressed in closed form and even it's numerical approxiamtion is time consuming. At this stage, I do not have formal results about MLE in this setting. In the rest of this section, I explore different estimators of a, I try compare the three through simulations. The first estiamtor is a simple method of moments estimator for which inference is simple. The second estimator is a fixed effects estimator and the last is the MLE. For these two last estimators, I am not able to determine the asymptotic distribution. For the MLE, I try to apply the same arguments I used for model (3.3) and show why they don't go through in this general setting.

The three estimators

The method of moments and composit likelihood

Note that for any i, j:

$$\begin{aligned} E(Y_{ij}) &= P(Z \leq a) \\ &= \Phi(\frac{a}{\sqrt{3}}) \end{aligned} \tag{3.11}$$

where Z is some random variable $Z \sim N(0, 3)$, and Φ is the CDF of the standard normal. The equation 3.11 suggests the following (method of moments) estimator:

$$\hat{a}_{MM} := \sqrt{3}\Phi^{-1}(\bar{Y})$$

$\bar{Y}$ being the sample mean of Y_{ij}.

Straightforward algebra shows that $\hat{a}_{MM}$ coincides with the composit likelihood estimator of a.I follow standard arguments to find the asymptotic distribution of $\hat{a}_{MM}$. First, note that:

$$\bar{Y} = \frac{2}{N(N-1)} \sum_{ij} (Y_{ij} - E(Y_{ij}|U_i, U_j)) + \frac{2}{N(N-1)} \sum_{ij} E(Y_{ij}|U_i, U_j)$$

and that:

$$\begin{aligned} Var\left(\frac{2\sqrt{N}}{N(N-1)} \sum_{ij} Y_{ij} - E(Y_{ij}|U_i, U_j\right) &= \frac{4}{N(N-1)^2} \sum_{ij} Var\left(Y_{ij} - E(Y_{ij}|U_i, U_j)\right) \\ &= \frac{4}{(N-1)} Var\left(Y_{12} - E(Y_{12}|U_i, U_j)\right) \\ &\to_N 0 \end{aligned}$$

Since $\frac{2}{N(N-1)} \sum_{ij} E(Y_{ij}|U_i, U_j)$ is a U-statistic, we get:

$$\sqrt{N}(\bar{Y} - E(Y)) \to_d N\left(0, 4Var[E(Y_{12}|U_1)]\right)$$

The delta method yields:

$$\sqrt{N}(\hat{a}_{MM} - a) \to N\left(0, 12\frac{Var[E(Y_{12}|U_1)]}{\phi(\frac{a}{\sqrt{3}})^2}\right)$$

ϕ: the PDF of the standard normal.

The fixed effect estimator

For every i:

$$E(Y_{ij}|U_i) = \Phi(\frac{a + U_i}{\sqrt{2}})$$

and note that:

$$\frac{1}{N}\sum_{i=1}^{N} a + U_i \to_{a.s.} a$$

this suggests the following estimator:

$$\hat{a}_{FE} := \frac{\sqrt{2}}{N}\sum_{i=1}^{N} \Phi^{-1}\left(\frac{1}{N-1}\sum_{j\neq i} Y_{ij}\right)$$

The maximum likelihood estimator.

Let's write the likelihood function of $Y = (Y_{ij})_{i<j\in\{1...N\}}$. For that, denote, for any $y \in \{0,1\}^{\frac{N(N-1)}{2}}$:

$$N(y) := 2y - 1$$

That is, $N(y)$ is an $\frac{N(N-1)}{2} \times 1$ vector with entry i, j equal to 1 when $y_{ij} = 1$ and -1 when $y_{ij} = 0$.

Also, $M(y) := (M(y)_{(ij),k})_{i<j\in\in\{1...N\};k\leq N}$ is the $\frac{N(N-1)}{2} \times N$ matrix, with :

$$M(y)_{(ij),k} = \begin{cases} 1 & \text{if } y_{ij} = 1 \text{ and } k \in \{i,j\} \\ -1 & \text{if } y_{ij} = 0 \text{ and } k \in \{i,j\} \\ 0 & \text{otherwise.} \end{cases}$$

write, $U = (U_1, ..., U_N)'$; and $V = (V_{12}, V_{13}, ..., V_{N-1,N})$. We can express the likelihood function, for all $y \in \{0,1\}^{\frac{N(N-1)}{2}}$:

$$f(y;a) := P_a(Y=y) = P_a(aN(y) + M(y)U + diag(N(y))V \geq 0) \\ = P_a(Z \geq 0) \tag{3.12}$$

for some random variable $Z \sim N\left(aN(y), I_{N(N-1)/2} + M(y)M(y)'\right) = N\left(aN(y), \Omega\right)$, $I_{N(N-1)/2}$ being the identity matrix of dimension $N(N-1)/2$, and $\Omega := I_{N(N-1)/2}+M(y)M(y)'$. The inequality in 3.12 is in the element-wise sense.

Therefore:

$$f(y;a) \propto e^{-\frac{1}{2}a^2 N(y)'\Omega(y)^{-1}N(y)} \int_{z\geq 0} e^{-\frac{1}{2}z'\Omega(y)^{-1}z} e^{aN(y)'\Omega(y)^{-1}z} dz$$

Hence, the loglikelihood:

$$\log(f(y;a)) = constant + \log\left(\int_{z\geq 0} e^{\frac{-1}{2}z'\Omega(y)^{-1}z} e^{aN(y)'\Omega(y)^{-1}z} dz\right) - \frac{1}{2}a^2 N(y)'\Omega(y)^{-1}N(y)$$

This function is strictly concave, the maximum likelihood estimator solves:

$$\frac{\int_{z\geq 0} N(y)'\Omega(y)^{-1} z e^{\frac{-1}{2}z'\Omega(y)^{-1}z} e^{aN(y)'\Omega(y)^{-1}z} dz}{\int_{z\geq 0} e^{\frac{-1}{2}z'\Omega(y)^{-1}z} e^{aN(y)'\Omega(y)^{-1}z} dz} = aN(y)'\Omega(y)^{-1}N(y)$$

which is equivalent to:

$$E\left(N(y)'\Omega(y)^{-1}Z | Z \geq 0\right) = aN(y)'\Omega(y)^{-1}N(y) \tag{3.13}$$

where, as before, Z is some random vector distributed following $N\left(aN(y), \Omega\right)$.

The MLE can't be obtained in closed form, but can be computed numerically using the equation (3.13). The question now is to determine the asymptotic distribution of $\hat{a}_{MLE}$ and compare it to the method of moments (or composit likelihood) estimator obtained earlier.

The likelihood function of Y can be written as an exponential distribution of the form the equation (3.4):

$$f_N(y;\eta) = e^{T_N(y)'A_N(\eta) - B_N(\eta)}$$

where $\eta = a$ (to link the notation of the example to that of the general framework), $T_N(y)$ is a $2^{\frac{N(N-1)}{2}} \times 1$ vector that has zeroes everywhere except in the line corresponding to the value of y, and $A_N(\eta) = (f(y_1, \eta), f(y_2, \eta), ..., f(y_{\frac{N(N-1)}{2}}, \eta))'$, where $\{y_1, ..., y_{\frac{N(N-1)}{2}}\} = \{0,1\}^{\frac{N(N-1)}{2}}$ is list of al possible values of y. Finally, $B_N(\eta) = 0$ for all η.

First, I want to show that:

$$\frac{s_N(\eta)}{\sqrt{-T_N(_N)A"_N(\eta) + B"_N(\eta)}} \rightarrow_d N(0,1)$$

for $s_N(\eta) = E\left(N(y)'\Omega(y)^{-1}Z|Z \geq 0\right)|_{y=Y_N} - aN(y)'\Omega(y)^{-1}N(y) = T_N(Y_N)A'_N(\eta) - B'_N(\eta)$, the score.

It is enough to show that the characteristic function of $\frac{s_N(\eta)}{\sqrt{-T_N(_N)A"_N(\eta)+B"_N(\eta)}}$ converges to $\phi \to e^{-\frac{1}{2}\phi^2}$, the characteristic function of the standard normal.

$$
\begin{aligned}
&E_\eta\left(e^{\phi\frac{s_N(\eta)}{\sqrt{-T_N(Y_N)A"_N(\eta)+B"_N(\eta)}}}\right) \\
&= \int h_N(y)\exp\left(\phi\frac{T_N(y)A'_N(\eta) - B'_N(\eta)}{\sqrt{-T_N(y)A"_N(\eta) + B"_N(\eta)}} + T_N(y)A_N(\eta) - B_N(\eta)\right)d\mu(y) \\
&= e^{-\frac{1}{2}\phi^2}\int h_N(y)\exp\left(\phi\frac{T_N(y)A'_N(\eta) - B'_N(\eta)}{\sqrt{-T_N(y)A"_N(\eta) + B"_N(\eta)}} + T_N(y)A_N(\eta) - B_N(\eta) + \frac{1}{2}\phi^2\right)d\mu(y) \\
&= e^{-\frac{1}{2}\phi^2}\int h_N(y)\exp[T_N(y)A_N(\eta + \phi/\sqrt{-T_N(y)A"_N(\eta) + B"_N(\eta)}) \\
&- B_N(\eta + \phi/\sqrt{-T_N(y)A"_N(\eta) + B"_N(\eta)})]\exp\left(-\frac{1}{2}\phi^2\frac{T_N(y)A^{(3)}(\bar\eta) - B_N^{(3)}(\bar\eta)}{-T_N(y)A"_N(\eta) + B"_N(\eta)}\right)d\mu(y)
\end{aligned}
\tag{3.14}
$$

the last equality uses a Taylor expansion, $\bar\eta$ is some scalar in $[\eta, \eta + \frac{\phi}{\sqrt{-T_N(y)A"_N(\eta)+B"_N(\eta)}}]$.

Therefore, we obtain the following proposition:

Proposition. Assume that for any η and $\bar\eta$:

$$\frac{[T_N(Y_N)A_N^{(3)}(\bar\eta - B_N^{(3)}(\bar\eta]^{\frac{1}{3}}}{[T_N(Y_N)A"_N(\eta) - B''_N(\eta)]^{\frac{1}{2}}} \to 0;\ \text{almost surely.}$$

then:

$$\frac{s_N(\eta)}{\sqrt{-T_N(_N)A"_N(\eta) + B"_N(\eta)}} \to_d N(0,1)$$

Note that the proof is not over, I haven't shown that the binomial model satisfies the (very strong) condition of the proposition. Moreover, the condition is sufficient but no necessary, I suspect it is too strong for our purposes. The fact that the argument used with model (3.3) requires such a strong condition to be satisfied suggests that those arguments might not yield any results in the more general model (3.4).

The simulations

Because the MLE cannot be computed in closed form, the runtime to numerically approximate it makes its computation more time consuming than the fixed effect and method of moment estimators. Therefore, I run two sets of simulations, the first is (figures 3.1, 3.2 and 3.3) is composed of 250 simulations of graphs of 25 nodes, the second (figures 3.4 and 3.5)is composed of 1000 simulations of graphs of 250 nodes . These simulations are performed using Gnez (2007)Genz (1992)'s sampler sampler that exploits the properties of Gaussian random vectors.

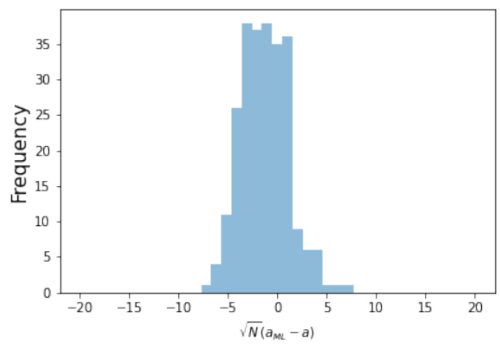

Figure 3.1: Maximum Likelihood, N=25 nodes, n=250 simulations

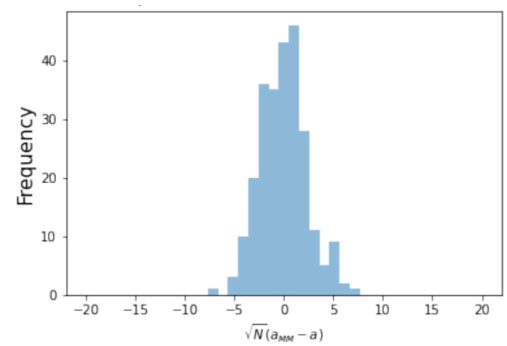

Figure 3.2: Method of Moments, N=25 nodes, n=250 simulations

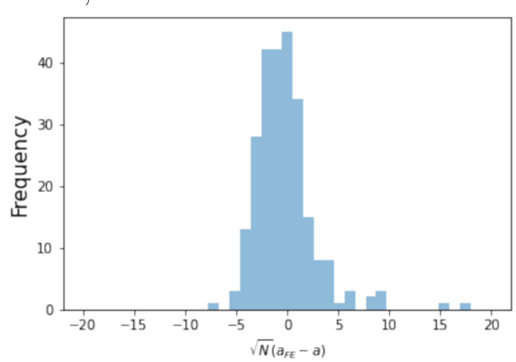

Figure 3.3: Fixed Effects, N=25 nodes, n=250 simulations

These simulations suggest that the fixed effects estimator is asymptotically normally distributed but that it is asymptotically biased. This relates to the well known incidental parameters problem that pertains to the fixed effects estimation in the panel data literature (see for instance Hahn and Newey (2004) Hahn and Newey (2004)). The simulations are less conclusive for the maximum likelihood estimator but suggest that it is also normally distributed. The simulations also confirm what we already knew about the method of moments estimator (it is unbiased and asymptotically normal).

Conclusion

The results in this prospectus provide partial answers to the question of parametric efficiency for models of exponential families. Many related questions are left open. First, the sufficient

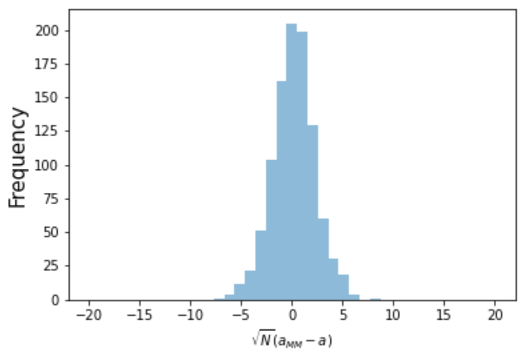

Figure 3.4: Method of Moments, N=250 nodes, n=1000 simulations

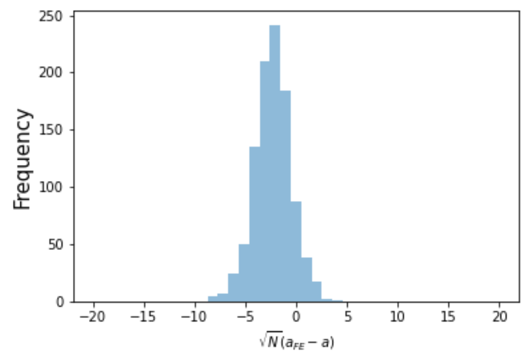

Figure 3.5: Fixed Effects, N=250 nodes, n=1000 simulations

conditions that I propose for asymptotic efficiency are often hard to check. When the normalizing constant $B_N(\eta)$ in model (3.13) cannot be obtained in closed form, as is usually the case, then it is not clear how one can verify the condition on its second and third differentials.

Second, this work does not discuss the implementation of the maximum likelihood estimator. When the normalizing constant $B_N(\eta)$ can't be computed in closed form, it has to be approximated numerically. MCMC methods are often used for this purpose but remain unsatisfactory (Snijders (2002) Snijders (2002)). Some more recent developments exploit large deviation principles of random graphs (Chatterjee and Diaconis (2013) Chatterjee and Diaconis (2013), Mele and Zhu (2020) Mele and Zhu (2019)) but the existing results impose rather binding constraints on the choice of the sufficient statistic.

Finally, extending these results to the more general exponential family model (equation (3.4)) seems to be hard, the arguments behind the proofs in this prospectus do not extend in an obvious way. Model (3.4) is much more general than the class of models studied here and could pave the way to extensions to non-exponential models.

3.6 Proofs

Proof of proposition 1

Define the score:

$$s_N(\eta) := \frac{\partial l_n(Y_n;\eta)}{\partial \eta} = T_N(Y_N) - \frac{\partial B_N(\eta)}{\partial \eta}$$

Let's first show that the score has an asymptotically normal distribution:

$$I_N(\eta)^{-1/2} s_N(\eta) \to_d N(0, I_k)$$

The moment generating function of $I_N(\eta)^{-1/2}s_N(\eta)$ evaluated at some given ϕ:

$$\begin{aligned}
&M_N(\phi;\eta)\\
:=\ &E(e^{\phi' I_N(\eta)^{-1/2}s_N(\eta)})\\
=\ &\int h_N(y)\exp\left(\phi' I_N(\eta)^{-1/2}(T_N(y)-\frac{\partial B_N(\eta)}{\partial\eta})+T_N(y)'\eta-B_N(\eta)\right)dy\\
=\ &e^{-\frac{1}{2}\phi'\phi}\int h_N(y)\exp\left((\eta+\phi' I_N(\eta)^{-1/2})'T_N(y)-B_N(\eta+\phi' I_N(\eta)^{-1/2})\right)\\
&\qquad\times\exp(-Rem_N(\phi' I_N(\eta)^{-1/2},\eta)
\end{aligned}$$

where the last equality results from a Taylor expansion and where $Rem_N(\phi' I_N(\eta)^{-1/2},\eta)$, the reminder term, can be expressed as:

$$Rem_N(\phi' I_N(\eta)^{-1/2},\eta) := \sum_{|\beta|=3} R_{\beta,N}(\phi' I_N(\eta)^{-1/2}+\eta)(\phi' I_N(\eta)^{-1/2})^{\beta}$$

where β here is a multi-index, and

$$\begin{aligned}
R_{\beta,N}(\phi' I_N(\eta)^{-1/2}+\eta) &= -\frac{3}{\beta!}\int_0^1(1-\tau)^2 D_\eta^\beta B_N|_{\eta+\tau\phi' I_N(\eta)^{-1/2}}d\tau\\
&= -\frac{3}{\beta!}\int_0^1(1-\tau)^2\frac{\partial^3 B_N}{\partial\eta^\beta}(\eta+\tau\phi' I_N(\eta)^{-1/2})d\tau
\end{aligned}$$

Therefore:

$$M_N(\phi,\eta)=\exp\left(-\frac{1}{2}\phi'\phi\right)\exp\left(-Rem_N(\phi' I_N(\eta)^{-1/2},\eta)\right)$$

Since $I_N(\eta)^{-1}$ is bounded, then there exists a compact K such that for all N: $\eta+\tau\phi' I_N(\eta)^{-1/2}\in K$. Also, given the condition that for any $j\leq k$:

$$\sup_{\eta,\tilde{\eta}\in K}\frac{||\frac{\partial^3 B_N}{\partial\eta_j\partial^2\eta}||}{||\frac{\partial^2 B_N}{\partial\eta^2}||^{3/2}}=o(1)$$

we get:

$$M_N(\phi,\eta)\to_n \exp(-\frac{1}{2}\phi\phi')$$

$\phi\to\exp(-\frac{1}{2}\phi\phi')$ being the moment generating function of the standard normal, we obtain the desired result.

Now let's show:

$$I_N(\eta)^{1/2}(\hat{\eta}-\eta) \to_d N(0, I_k)$$

Note that: $T_N(Y_N) = \frac{\partial B_N(\hat{\eta})}{\partial \eta}$ (remember, $\hat{\eta}$ is the MLE estimator of η, since $\frac{\partial^2 B_N(\eta)}{\partial \eta^2}$ is positive definit, it is unique whenever it exists).

Observe:

$$\begin{aligned}
\hat{\eta}-\eta &= (\frac{\partial B_N}{\partial \eta})^{-1}(\frac{\partial B_N(\hat{\eta})}{\partial \eta}) - (\frac{\partial B_N}{\partial \eta})^{-1}(\frac{\partial B_N(\eta)}{\partial \eta}) \\
&= \frac{\partial}{\partial \eta}(\frac{\partial B_N}{\partial \eta}^{-1})(\frac{\partial B_N(\eta)}{\partial \eta})\left(\frac{\partial B_N(\hat{\eta})}{\partial \eta} - \frac{\partial B_N(\eta)}{\partial \eta}\right) \\
&+ \frac{1}{2}\sum_{j=1}^{k} \frac{\partial^2}{\partial \eta_j \partial \eta}(\frac{\partial B_N^{-1}}{\partial \eta})(\frac{\partial B_N(\bar{\eta})}{\partial \eta})\left(\frac{\partial B_N(\hat{\eta})}{\partial \eta} - \frac{\partial B_N(\eta)}{\partial \eta}\right)\left(\frac{\partial B_N(\hat{\eta})}{\partial \eta} - \frac{\partial B_N(\eta)}{\partial \eta}\right)_j
\end{aligned}$$

for some $\bar{\eta}$ between η and $\hat{\eta}$, also denote $\bar{x} := \frac{\partial B_N(\bar{\eta})}{\partial \eta}$.

Then:

$$\begin{aligned}
&\hat{\eta}-\eta \\
&= \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1} s_N(\eta) \\
&+ \frac{1}{2}\sum_{j,l=1}^{k} s_{N,j}(\eta)\left[\frac{\partial}{\partial \eta_j}\left(\frac{\partial B_N}{\partial \eta}\right)^{-1}(\bar{x})\right]_l \left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} \frac{\partial}{\partial \eta_l}\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} s_N(\eta) \\
&= \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1} s_N(\eta) \\
&+ \frac{1}{2}\sum_{l=1}^{k}\left[\frac{\partial}{\partial \eta}\left(\frac{\partial B_N}{\partial \eta}\right)^{-1}(\bar{x})\right]_{l,\cdot} s_N(\eta)\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1}\left(\frac{\partial^3 B_N(\bar{\eta})}{\partial \eta^2 \partial \eta_j}\right)\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} s_N(\eta) \\
&= \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1} s_N(\eta) \\
&+ \frac{1}{2}\sum_{l=1}^{k}\left[\left(\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta})\right)^{-1}\right]_{l,\cdot} s_N(\eta)\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1}\left(\frac{\partial^3 B_N(\bar{\eta})}{\partial \eta^2 \partial \eta_j}\right)\left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} s_N(\eta)
\end{aligned}$$

Finally:

$$\left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{1/2} (\hat{\eta}-\eta)$$
$$= \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta)$$
$$+ \frac{1}{2}\sum_{l=1}^{k} \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{1/2} \left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} \left(\frac{\partial^3 B_N(\bar{\eta})}{\partial \eta^2 \partial \eta_j}\right) \left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1}$$
$$\times s_N(\eta) \left[\left(\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta})\right)^{-1}\right]_{l,.} s_N(\eta)$$

If there exists a compact K such that $\bar{\eta} \in K$ almost surely, then for all $l = 1..k$, if:

$$\Delta_{N,l} := \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{1/2} \left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1} \left(\frac{\partial^3 B_N(\bar{\eta})}{\partial \eta^2 \partial \eta_j}\right) \left(\frac{\partial^2 B_N(\bar{\eta})}{\partial \eta^2}\right)^{-1}$$
$$\times s_N(\eta) \left[\left(\frac{\partial^2 B_N}{\partial \eta^2}(\bar{\eta})\right)^{-1}\right]_{l,.} s_N(\eta)$$

then:

$$||\Delta_{N,l}|| \leq \sup_{\eta,\tilde{\eta}\in K} \frac{||\frac{\partial^3 B_N(\eta)}{\partial \eta_j \partial^2 \eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial \eta^2}||^{3/2}} \times || \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta)||^2 = o_p(1)$$

therefore, since $\hat{\eta}$ is tight, $\bar{\eta}$ is also tight, and fix any $\epsilon > 0$, I want to show that:

$$\mathbb{P}\left(||\Delta_{N,l}|| > \epsilon\right) \to_{N\to\infty} 0$$

by tightness, for any $\epsilon' > 0$, there exists an $M > 0$ such that $\mathbb{P}(||\bar{\eta} - \eta|| > M) < \epsilon'$ for all N; denote $K_M := \{x : ||x - \eta|| \leq M\}$ hence:

$$\mathbb{P}\left(||\Delta_{N,l}|| > \epsilon\right)$$
$$= \mathbb{P}\left(||\Delta_{N,l}|| > \epsilon, ||\bar{\eta}-\eta|| > M\right) + \mathbb{P}\left(||\Delta_{N,l}|| > \epsilon, ||\bar{\eta}-\eta|| \leq M\right)$$
$$\leq \mathbb{P}\left(||\bar{\eta}-\eta|| > M\right) + \mathbb{P}\left(\sup_{\eta,\tilde{\eta}\in K_M} \frac{||\frac{\partial^3 B_N(\eta)}{\partial \eta_j \partial^2 \eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial \eta^2}||^{3/2}} \times || \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta)||^2 > \epsilon\right)$$
$$\leq \epsilon' + \mathbb{P}\left(\sup_{\eta,\tilde{\eta}\in K_M} \frac{||\frac{\partial^3 B_N(\eta)}{\partial \eta_j \partial^2 \eta}||}{||\frac{\partial^2 B_N(\tilde{\eta})}{\partial \eta^2}||^{3/2}} \times || \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta)||^2 > \epsilon\right)$$

Taking limits on both sides, we get:

$$\lim_{N\to\infty} \mathbb{P}\left(||\Delta_{N,l}|| > \epsilon\right) \leq \epsilon'$$

this inequality holds for any ϵ', therefore:

$$\lim_{N\to\infty} \mathbb{P}\left(||\Delta_{N,l}|| > \epsilon\right) = 0$$

i.e. $\Delta_{N,l} = o_p(1)$, so:

$$\left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{1/2} (\hat{\eta} - \eta) = \left(\frac{\partial^2 B_N(\eta)}{\partial \eta^2}\right)^{-1/2} s_N(\eta) + o_p(1)$$

which allows to conclude.

Gaussian linear regression example: computational details

Remember: $\Omega_N := (\omega_{m,n})_{m,n \leq \frac{N(N-1)}{2}}$ and, using the same abuse of notation above:

$$\omega_{ij,kl} := \begin{cases} 3 & \text{if } \{i,j\} = \{k,l\} \\ 1 & \text{if } |\{i,j\} \cap \{k,l\}| = 1 \\ 0 & \text{otherwise} \end{cases}$$

and the (conditional) MLE estimator of β:

$$\hat{\beta} := (X_N' \Omega_N^{-1} X_N)^{-1} (X_N' \Omega_N^{-1} Y_N) \sim_{|X} N(\beta, (X_N' \Omega_N^{-1} X_N)^{-1})$$

Let's evaluate the information matrix $X_N' \Omega_N^{-1} X_N$. For that, I will first need to compute $\Omega_N^{-1} =: ({\omega^*}_{ij,kl})_{ij,kl}$.

Computing Ω_N^{-1}

I conjecture[8] that ${\omega^*}_{ij,kl}$ has the following shape:

$$\omega^*_{ij,kl} := \begin{cases} a_N & \text{if } \{i,j\} = \{k,l\} \\ b_N & \text{if } |\{i,j\} \cap \{k,l\}| = 1 \\ c_N & \text{otherwise} \end{cases} \tag{3.15}$$

Then I solve for a_N, b_N and c_N, and check that the matrix $({\omega^*}_{ij,kl})_{ij,kl}$ given by 3.15 is indeed Ω_N^{-1}.

[8] Thanks to Python!

By definition of the inverse, for all i, j, p, m:

$$\sum_{kl} \omega_{ij,kl}\omega^*_{kl,pm} = \delta_{\{i,j\},\{p,m\}} = \begin{cases} 1 & \text{if } \{i,j\} = \{p,m\} \\ 0 & \text{otherwise} \end{cases}$$

On the other side, assuming 3.15:

$$\sum_{kl} \omega_{ij,kl}\omega^*_{kl,pm} = a_N \times \omega_{ij,pm} + b_N \left(\sum_{|\{k,l\} \cap \{p,m\}|=1} \omega_{ij,kl} \right) + c_N \left(\sum_{\{k,l\} \cap \{p,m\}=\emptyset} \omega_{ij,kl} \right)$$

So:

$$a_N \times \omega_{ij,pm} + b_N \left(\sum_{|\{k,l\} \cap \{p,m\}|=1} \omega_{ij,kl} \right) + c_N \left(\sum_{\{k,l\} \cap \{p,m\}=\emptyset} \omega_{ij,kl} \right) = \delta_{\{i,j\},\{p,m\}} \tag{3.16}$$

We have 3 cases:

- $\{i,j\} = \{p,m\}$: equation 3.16 is equivalent to:

$$3a_N + 2(N-2)b_N = 1 \tag{3.17}$$

- —$\{i,j\} \cap \{p,m\}| = 0$: equation 3.16 is equivalent to:

$$4b_N + (2N-5)c_N = 0 \tag{3.18}$$

- —$\{i,j\} \cap \{p,m\}| = 1$: equation 3.16 becomes:

$$a_N + (N+1)b_N + (N-3)c_N = 0 \tag{3.19}$$

Equations (3.17) - (3.19) yield:

$$\begin{cases} a_N = \frac{1}{3} - \frac{2(N-2)}{3} b_N \\ c_N = -\frac{4}{2n-5} b_N \\ a_N + (N+1)b_N + (N-3)c_N = 0 \end{cases}$$

$$\iff \begin{cases} a_N = \frac{1}{3} - \frac{2(N-2)}{3} b_N \\ c_N = -\frac{4}{2n-5} b_N \\ \frac{1}{3} - \frac{2(N-2)}{3} b_N + (N+1)b_N - \frac{4(N-3)}{2N-5} b_N = 0 \end{cases}$$

Finally:

$$\begin{cases} a_N = \frac{2N^2-7N+7}{(N-1)(2N-1)} \\ b_N = -\frac{2N-5}{(N-1)(2N-1)} \\ c_N = \frac{4}{(N-1)(2N-1)} \end{cases} \tag{3.20}$$

To check that the guess in 3.15 is correct, it suffices to check that a_N, b_N and c_N above satisfy equations 3.17 to 3.19, which they of course do.

A simple example

Take the example where:

$$X_{ij} = (1, X_i X_j)$$

for some i.i.d random variables $(X_i)_{1\leq i\leq N}$.

$$\begin{aligned}
X_N'\Omega_N^{-1}X_N &= \sum_{i<j;k<l} X_{ij}'\omega_{ij,kl}^* X_{kl} \\
&= a_N \sum_{i<j} X_{ij}'X_{ij} + b_N \times \sum_{|\{i,j\}\cap\{k,l\}|=1;i<j;k<l} X_{ij}'X_{kl} + c_N \times \sum_{\{i,j\}\cap\{k,l\}=\emptyset;i<j;k<l} X_{ij}'X_{kl} \\
&= a_N \sum_{i<j} \begin{pmatrix} 1 & X_iX_j \\ X_iX_j & X_i^2X_j^2 \end{pmatrix} + b_N \times \sum_{|\{i,j\}\cap\{k,l\}|=1;i<j;k<l} \begin{pmatrix} 1 & X_kX_l \\ X_iX_j & X_kX_lX_iX_j \end{pmatrix} + ... \\
&... + c_N \times \sum_{\{i,j\}\cap\{k,l\}=\emptyset;i<j;k<l} \begin{pmatrix} 1 & X_kX_l \\ X_iX_j & X_kX_lX_iX_j \end{pmatrix} \\
&=: \begin{pmatrix} \mathcal{A}_N & \mathcal{B}_N \\ \mathcal{C}_N & \mathcal{D}_N \end{pmatrix}
\end{aligned}$$

with:

$$\begin{aligned}
\mathcal{A}_N &= a_N \sum_{i<j} 1 + b_N \sum_{|\{i,j\}\cap\{k,l\}|=1;i<j;k<l} 1 + c_N \sum_{\{i,j\}\cap\{k,l\}=\emptyset;i<j;k<l} 1 \\
&= \frac{N(N-1)}{2} a_N + N(N-1)(N-2) b_N \\
&\quad + \left((\frac{N(N-1)}{2})^2 - \frac{N(N-1)}{2} - N(N-1)(N-2) \right) c_N \\
&= \frac{N(N-1)}{2} \left(a_N + 2(N-2)b_N + \left(\frac{N(N-1)}{2} - 1 - 2(N-2) \right) c_N \right) \\
&= \frac{N(N-1)}{2(3+2(N-2))} \\
&= \frac{N}{4} + o(N)
\end{aligned}$$

$$
\begin{aligned}
\mathcal{B}_N = \mathcal{C}_N &= a_N \sum_{i<j} X_i X_j + b_N \sum_{|\{i,j\}\cap\{k,l\}|=1; i<j; k<l} X_i X_j + c_N \sum_{\{i,j\}\cap\{k,l\}=\emptyset; i<j; k<l} X_i X_j \\
&= a_N \sum_{i<j} X_i X_j + b_N \sum_{i<j} X_i X_j \sum_{|\{i,j\}\cap\{k,l\}|=1; k<l} 1 + c_N \sum_{i<j} X_i X_j \sum_{\{i,j\}\cap\{k,l\}=\emptyset; k<l} 1 \\
&= a_N \sum_{i<j} X_i X_j + 2(N-2) b_N \sum_{i<j} X_i X_j \\
&\qquad + \left(\frac{N(N-1)}{2} - 1 - 2(N-2) \right) c_N \sum_{i<j} X_i X_j \\
&= \frac{N(N-1)}{2(3+2(N-2))} \left(\frac{2}{N(N-1)} \sum_{i<j} X_i X_j \right) \\
&= \frac{N}{4} E(X_1)^2 + o_p(N)
\end{aligned}
$$

where the last equality uses a U-statistic law of large numbers. Similarly:

$$
\begin{aligned}
\mathcal{D}_N &= a_N \sum_{i<j} X_i^2 X_j^2 + b_N \sum_{|\{i,j\}\cap\{k,l\}|=1; i<j; k<l} X_k X_l X_i X_j + c_N \sum_{\{i,j\}\cap\{k,l\}=\emptyset; i<j; k<l} X_k X_l X_i X_j \\
&= a_N \binom{N}{2} \left(E(X_1^2)^2 + o_p(1) \right) + N(N-1)(N-2) b_N \left(E(X_1)^2 E(X_1^2) + o_p(1) \right) \\
&+ \frac{N(N-1)(N-2)(N-3)}{4} c_N \left(E(X_1)^4 + o_p(1) \right) \\
&= \binom{N}{2} \left(a_N E(X_1^2)^2 + 2(N-1) b_N E(X_1)^2 E(X_1^2) + \frac{(N-2)(N-3)}{2} c_N E(X_1)^4 + o_p(1) \right)
\end{aligned}
$$

Assuming: $E(X_1^2)^2 - 2E(X_1)^2 E(X_1^2) + E(X_1)^4 \neq 0$, i.e. $Var(X_1) \neq 0$, then:

$$
\mathcal{D}_N = O_p(N^2)
$$

Let's now compute $(X_N' \Omega_N^{-1} X_N)^{-1}$:

$$
(X_N' \Omega_N^{-1} X_N)^{-1} = \frac{1}{\mathcal{A}_N \mathcal{D}_N - \mathcal{B}_N^2} \begin{pmatrix} \mathcal{D}_N & -\mathcal{B}_N \\ -\mathcal{B}_N & \mathcal{A}_N \end{pmatrix} \tag{3.21}
$$

So, denoting $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$ and the MLE estimator:$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$, equation (3.21) shows that:

$$\begin{aligned} Avar(\hat{\beta}_0) &= O(\frac{1}{N}) \\ Avar(\hat{\beta}_1) &= O(\frac{1}{N^2}) \end{aligned} \tag{3.22}$$

Bibliography

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.

Pat Bajari, Brian Burdick, Guido Imbens, Lorenzo Masoero, James McQueen, Thomas Richardson, and Ido Rosen. Experimental design in marketplaces. *Statistical Science*, 2023. URL https://www.amazon.science/publications/experimental-design-in-marketplaces.

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, 1987a. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(87)90015-7. URL https://www.sciencedirect.com/science/article/pii/0304407687900157.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34:305–334, 1987b.

Arun G. Chandrasekhar and Matthew O. Jackson. Tractable and consistent random graph models. Available at SSRN: https://ssrn.com/abstract=2150428 or http://dx.doi.org/10.2139/ssrn.2150428, September 21 2012.

Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *Ann. Statist.*, 41(5):2428–2461, 2013. doi: 10.1214/13-AOS1155. URL https://projecteuclid.org/euclid.aos/1383661269.

Harry Crane and Henry Towsner. Relatively exchangeable structures. *The Journal of Symbolic Logic*, 83(2):416–442, 2018. doi: 10.1017/jsl.2017.61.

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.

A. Genz. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1(2):141–149, 1992. doi: 10.1080/10618600.1992.10477010.

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.

J. Hahn and W.K. Newey. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319, 2004.

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.

R. Hofstad. *Random Graphs and Complex Networks, Volume 1*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016.

J. Hájek. A characterization of limiting distributions of regular estimators. *Z. Wahrsch. Verw. Gebiete*, 14:323–330, 1970.

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.

R.W. Keener. *Theoretical Statistics: Topics for a Core Course*. Springer Texts in Statistics. Springer New York, 2010. ISBN 9780387938394. URL https://books.google.co.in/books?id=aVJmcega44cC.

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.

L. LeCam. On some asymptotic properties of maximum likelihood estimates and related bayes' estimates. *Univ. California Publ. Statist.*, 1:277–329, 1953.

L. Lovász and M.D. Plummer. *Matching Theory*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2009. ISBN 9780821847596. URL https://books.google.com/books?id=OaoJBAAAQBAJ.

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.

A. Mele. A structural model of dense network formation. *Econometrica*, 85:825–850, 2017. doi: 10.3982/ECTA10400.

A. Mele and L. Zhu. Approximate variational estimation for a model of network formation. Available at SSRN: https://ssrn.com/abstract=2909829 or http://dx.doi.org/10.2139/ssrn.2909829, April 4 2019.

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.

Kevin A. O'Neil and Richard A. Redner. Asymptotic Distributions of Weighted U-Statistics of Degree 2. *The Annals of Probability*, 21(2):1159 – 1169, 1993. doi: 10.1214/aop/1176989286. URL https://doi.org/10.1214/aop/1176989286.

Patrick Eugene O'Neil. Asymptotics in random (0, 1)-matrices. *Proceedings of the American Mathematical Society*, 25(1):39–45, 1970. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/2036523.

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.

Cosma Rohilla Shalizi and Alessandro Rinaldo. Consistency under sampling of exponential random graph models. *Ann. Statist.*, 41(2):508–535, 2013. doi: 10.1214/12-AOS1044. URL https://projecteuclid.org/euclid.aos/1366980556.

T. A. B. Snijders. Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 2002.

Andrew T. Stephen and Olivier Toubia. Explaining the power-law degree distribution in a social commerce network. *Social Networks*, 31:262–270, 2009. ISSN 0378-8733. doi: 10.1016/j.socnet.2009.07.002.

M. S. Stigle. The epic story of maximum likelihood. *Statistical Science*, 22(4):598–620, November 2007.

Takeaki Uno. Algorithms for enumerating all perfect, maximum and maximal matchings in bipartite graphs. In *Proceedings of the 8th International Symposium on Algorithms and Computation*, ISAAC '97, page 92–101, Berlin, Heidelberg, 1997. Springer-Verlag. ISBN 3540638903.

A. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998. doi: 10.1017/CBO9780511802256.