

Supplemental Appendix: Shrinkage Estimation of High-Dimensional Factor Models with Structural Instabilities

Xu Cheng, Zhipeng Liao, and Frank Schorfheide

A Supplemental Tables

Tables S-1, S-2, S-3, and S-4 provide some additional Monte Carlo results.

Tables S-5 to S-7 provide a list of variables used in the empirical application.

Table S-1: KNOWN BREAK DATE, HOMOGENEOUS R^2 , $\pi_0 = 0.5$, I.I.D. ERRORS

DGP Configuration					$\widehat{\mathcal{M}}$	$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	\mathbf{w}	N	T		0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	100	0.88	(0.88	0.12	0.00)	(0.98	0.02	0.00)	1.00	1.04	0.90	1.42
3	3		150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.56
3	3		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.55
Panel B. Type-1 Change															
3	3	0.2	100	100	0.03	(0.94	0.06	0.00)	(0.97	0.03	0.00)	1.00	1.06	0.94	1.37
3	3	0.2	150	150	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	0.99	1.26
3	3	0.2	200	200	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.08	1.00	1.14
3	3	0.5	100	100	0.96	(0.96	0.04	0.00)	(0.97	0.03	0.00)	1.00	1.31	1.67	1.17
3	3	0.5	150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.34	2.44	1.01
3	3	0.5	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.25	3.14	1.00
3	3	1.0	100	100	0.71	(0.71	0.29	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.45	1.00
3	3	1.0	150	150	0.99	(0.99	0.01	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.34	1.00
3	3	1.0	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.33	1.00
Panel C. Type-2 Change															
1	2		100	100	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.99	1.16	1.00
1	2		150	150	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.15	1.00
1	2		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.15	1.00
3	4		100	100	0.36	(0.71	0.29	0.00)	(0.53	0.47	0.00)	1.00	1.17	1.43	0.79
3	4		150	150	0.96	(0.99	0.01	0.00)	(0.98	0.03	0.00)	1.00	1.26	1.26	0.98
3	4		200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	1.25	1.00

Notes: Parameters $\alpha = \beta = 0, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 1$.

Table S-2: KNOWN BREAK DATE, HOMOGENEOUS R^2 , $\pi_0 = 0.8, \zeta = 1$

DGP Configuration					$\widehat{\mathcal{M}}$	$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	\mathbf{w}	N	T		0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.80
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.73
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.72
Panel B. Type-1 Change															
3	3	0.2	100	200	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	2.08
3	3	0.2	150	300	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.57
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	1.30
3	3	0.5	100	200	0.50	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.80	1.26	0.90
3	3	0.5	150	300	0.90	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	3.61	2.25	0.91
3	3	0.5	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	4.49	3.41	0.99
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	2.84	1.01
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	2.03	1.00
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.98	1.38	1.00
Panel C. Type-2 Change															
1	2		100	200	0.97	(1.00	0.00	0.00)	(0.97	0.03	0.00)	1.00	1.27	1.61	0.95
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.10	1.19	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.14	1.00
3	4		100	200	0.00	(1.00	0.00	0.00)	(0.00	1.00	0.00)	1.00	1.05	2.22	0.75
3	4		150	300	0.07	(1.00	0.00	0.00)	(0.07	0.93	0.00)	1.00	1.04	1.63	0.55
3	4		200	400	0.40	(1.00	0.00	0.00)	(0.40	0.60	0.00)	1.00	1.22	1.29	0.57

Notes: Parameters $\alpha = \beta = 0.2, \rho_a = \rho_b = 0.5, \eta_a = \eta_b = 1, \zeta = 1$.

Table S-3: KNOWN BREAK DATE, HOMOGENEOUS R^2 , CHANGE IN FACTOR DYNAMICS

DGP Configuration						$\widehat{\mathcal{M}}$	$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
ρ_a	ρ_b	η_a	η_b	N	T		0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. $\pi_0 = 0.5$.																
0	0	1	2	100	100	0.95	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.13	0.86	1.01
0	0	1	2	150	150	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.06	0.85	1.00
0	0	1	2	200	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.85	1.00
0	0.8	1	1	100	100	0.53	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.18	0.90	1.16
0	0.8	1	1	150	150	0.62	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.22	0.88	1.12
0	0.8	1	1	200	200	0.72	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.23	0.86	1.09
0	0.8	1	0.6	100	100	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.99	1.64
0	0.8	1	0.6	150	150	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	1.63
0	0.8	1	0.6	200	200	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	0.99	1.61
Panel B. $\pi_0 = 0.8$																
0	0	1	2	100	200	0.90	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.54	0.63	1.06
0	0	1	2	150	300	0.98	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.35	0.61	1.01
0	0	1	2	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.17	0.61	1.00
0	0.8	1	1	100	200	0.38	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.49	0.78	1.53
0	0.8	1	1	150	300	0.49	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.70	0.72	1.39
0	0.8	1	1	200	400	0.58	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.79	0.68	1.30
0	0.8	1	0.6	100	200	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.04	0.99	2.79
0	0.8	1	0.6	150	300	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	1.00	2.74
0	0.8	1	0.6	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.72

Notes: Parameters: $\psi_i = \lambda_i, r_a = r_b = 1, \alpha = \beta = 0, \zeta = 1$. In the last three rows of each panel, the change from (ρ_a, η_a) to (ρ_b, η_b) does not result in a change in the factor variance, and such a change cannot be identified.

Table S-4: UNKNOWN BREAK DATE, HETEROGENEOUS R^2 , $\pi_0 = 0.8$

DGP Configuration					$\widehat{\mathcal{M}}$	$\widehat{r}_a - r_a$			$\widehat{r}_b - r_b$			MSE			
r_a	r_b	\mathbf{w}	N	T		0	-1	1	0	-1	1	PMS	PLS	Full	Sub
Panel A. No Change															
3	3		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.01	3.91
3	3		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.49
3	3		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.00	1.00	3.10
Panel B. Type-1 Change															
3	3	0.2	100	200	0.02	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.03	0.97	3.29
3	3	0.2	150	300	0.01	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.02	0.99	3.15
3	3	0.2	200	400	0.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.01	1.00	2.54
3	3	0.5	100	200	0.70	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.24	0.79	1.11
3	3	0.5	150	300	0.93	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.93	1.13	1.36
3	3	0.5	200	400	0.99	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	2.35	1.50	1.39
3	3	1.0	100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.19	2.63	1.40
3	3	1.0	150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.14	4.13	1.18
3	3	1.0	200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	1.09	4.71	1.03
Panel C. Type-2 Change															
1	2		100	200	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.94	2.78	1.01
1	2		150	300	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.96	1.57	1.00
1	2		200	400	1.00	(1.00	0.00	0.00)	(1.00	0.00	0.00)	1.00	0.97	1.18	1.00
3	4		100	200	0.21	(1.00	0.00	0.00)	(0.21	0.79	0.00)	1.00	0.98	2.88	1.31
3	4		150	300	0.71	(1.00	0.00	0.00)	(0.71	0.29	0.00)	1.00	1.05	3.47	1.04
3	4		200	400	0.97	(1.00	0.00	0.00)	(0.97	0.03	0.00)	1.00	1.03	3.45	1.00

Notes: Parameters $\alpha = \beta = 0.2$, $\rho_a = \rho_b = 0.5$, $\eta_a = \eta_b = 1$, $\zeta = 4$. The conjecture break date π_c is correctly specified.

Table S-5: LIST OF VARIABLES - PART 1

Name	Category	TC	Long Description
Cons: Dur	NIPA	5	Real Personal Consumption Expenditures: Durable Goods
Cons: Svc	NIPA	5	Real Personal Consumption Expenditures: Services
Cons: NonDur	NIPA	5	Real Personal Consumption Expenditures: Nondurable Goods
Real InvtCh	NIPA	1	Component for Change in Private Inventories, deflated by JCXFE
Real WageG	NIPA	5	Component for Government GDP: Wage and Salary Disbursements by Industry, Government, NIPA Tables 2.7A and 2.7B, deflated by JCXFE
IP: DurGds materials	IP	5	Industrial Production: Durable Materials
IP: NondurGds materials	IP	5	Industrial Production: Nondurable Materials
IP: DurConsGoods	IP	5	Industrial Production: Durable Consumer Goods
IP: Auto	IP	5	IP: Automotive products
IP: NonDurConsGoods	IP	5	Industrial Production: Nondurable Consumer Goods
IP: BusEquip	IP	5	Industrial Production: Business Equipment
IP: EnergyProds	IP	5	IP: Consumer Energy Products
CapU Tot	IP	1	Capacity Utilization: Total Industry
CapU Man	IP	1	Capacity Utilization: Manufacturing (FRED past 1972)
Emp: DurGoods	Emp	5	All Employees: Durable Goods Manufacturing
Emp: Const	Emp	5	All Employees: Construction
Emp: Edu&Health	Emp	5	All Employees: Education & Health Services
Emp: Finance	Emp	5	All Employees: Financial Activities
Emp: Infor	Emp	5	All Employees: Information Services
Emp: Bus Serv	Emp	5	All Employees: Professional & Business Services
Emp: Leisure	Emp	5	All Employees: Leisure & Hospitality
Emp: OtherSvcs	Emp	5	All Employees: Other Services
Emp: Mining/NatRes	Emp	5	All Employees: Natural Resources & Mining
Emp: Trade&Trans	Emp	5	All Employees: Trade, Transportation & Utilities
Emp: Retail	Emp	5	All Employees: Retail Trade
Emp: Wholesal	Emp	5	All Employees: Wholesale Trade
Emp: Gov(Fed)	Emp	5	All Employees: Government: Federal
Emp: Gov (State)	Emp	5	All Employees: Government: State Government
Emp: Gov (Local)	Emp	5	All Employees: Government: Local Government
URate: Age16-19	Emp	2	Unemployment Rate - 16-19 yrs
URate: Age > 20 Men	Emp	2	Unemployment Rate - 20 yrs. & over, Men
URate: Age > 20 Women	Emp	2	Unemployment Rate - 20 yrs. & over, Women
U: Dur < 5wks	Emp	5	Number Unemployed for Less than 5 Weeks
U: Dur 5-14wks	Emp	5	Number Unemployed for 5-14 Weeks
U: Dur > 15-26wks	Emp	5	Civilians Unemployed for 15-26 Weeks
U: Dur > 27wks	Emp	5	Number Unemployed for 27 Weeks & over
U: Job Losers	Emp	5	Unemployment Level - Job Losers
U: LF Reentry	Emp	5	Unemployment Level - Reentrants to Labor Force
U: Job Leavers	Emp	5	Unemployment Level - Job Leavers
U: New Entrants	Emp	5	Unemployment Level - New Entrants

Notes: TC is transformation code; see Stock and Watson (2012).

Table S-6: LIST OF VARIABLES - PART 2

Name	Category	TC	Long Description
Emp: SlackWk	Emp	5	Employment Level - Part-Time for Economic Reasons, All Industries
AWH Man	Emp	1	Average Weekly Hours: Manufacturing
AWH Privat	Emp	2	Average Weekly Hours: Total Private Industrie
AWH Overtime	Emp	2	Average Weekly Hours: Overtime: Manufacturing
HPermits	HSS	5	New Private Housing Units Authorized by Building Permit
Hstarts: MW	HSS	5	Housing Starts in Midwest Census Region
Hstarts: NE	HSS	5	Housing Starts in Northeast Census Region
Hstarts: S	HSS	5	Housing Starts in South Census Region
Hstarts: W	HSS	5	Housing Starts in West Census Region
Constr. Contracts	HSS	4	Construction contracts (mil. sq. ft.) (Copyright, McGraw-Hill)
Ret. Sale	Ord	5	Sales of retail stores (mil. Chain 2000 \$)
Orders (DurMfg)	Ord	5	Mfrs' new orders durable goods industries (bil. chain 2000 \$)
Orders (ConsumerGoods/Mat.)	Ord	5	Mfrs' new orders, consumer goods and materials (mil. 1982 \$)
UnfOrders (DurGds)	Ord	5	Mfrs' unfilled orders durable goods indus. (bil. chain 2000 \$)
Orders (NonDefCap)	Ord	5	Mfrs' new orders, nondefense capital goods (mil. 1982 \$)
VendPerf	Ord	1	Index of supplier deliveries – vendor performance (pct.)
MT Invent	Ord	5	Manufacturing and trade inventories (bil. Chain 2005 \$)
PCED-MotorVec	Pri	6	Motor vehicles and parts
PCED-DurHousehold	Pri	6	Furnishings and durable household equipment
PCED-Recreation	Pri	6	Recreational goods and vehicles
PCED-OthDurGds	Pri	6	Other durable goods
PCED-Food-Bev	Pri	6	Food and beverages purchased for off-premises consumption
PCED-Clothing	Pri	6	Clothing and footwear
PCED-Gas-Enrgy	Pri	6	Gasoline and other energy goods
PCED-OthNDurGds	Pri	6	Other nondurable goods
PCED-Housing-Utilities	Pri	6	Housing and utilities
PCED-HealthCare	Pri	6	Health care
PCED-TransSvg	Pri	6	Transportation services
PCED-RecServices	Pri	6	Recreation services
PCED-FoodServ-Acc.	Pri	6	Food services and accommodations
PCED-FIRE	Pri	6	Financial services and insurance
PCED-OtherServices	Pri	6	Other services
PPI: FinConsGds	Pri	6	Producer Price Index: Finished Consumer Goods
PPI: FinConsGds(Food)	Pri	6	Producer Price Index: Finished Consumer Foods
PPI: IndCom	Pri	6	Producer Price Index: Industrial Commodities
PPI: IntMat	Pri	6	Producer Price Index: Intermediate Materials: Supplies & Components
NAPM ComPrice	Pri	1	NAPM COMMODITY PRICES INDEX (PERCENT)
Real Price: NatGas	Pri	5	PPI: Natural Gas, deflated by PCEPILFE
Real Price: Oil	Pri	5	PPI: Crude Petroleum, deflated by PCEPILFE

Notes: TC is transformation code; see Stock and Watson (2012).

Table S-7: LIST OF VARIABLES - PART 3

Name	Category	TC	Long Description
FedFunds	IntL	2	Effective Federal Funds Rate
TB-3Mth	IntL	2	3-Month Treasury Bill: Secondary Market Rate
BAA-GS10	IntS	1	BAA-GS10 Spread
MRTG-GS10	IntS	1	Mortg-GS10 Spread
TB6m-TB3m	IntS	1	tb6m-tb3m
GS1-TB3m	IntS	1	GS1-Tb3m
GS10-TB3m	IntS	1	GS10-Tb3m
CP-TB Spread	IntS	1	CP-Tbill Spread: CP3FM-TB3MS
Ted-Spread	IntS	1	MED3-TB3MS (Version of TED Spread)
Real C&I Loan	Mon	5	Commercial and Industrial Loans at All Commercial BanksDefl by PCEPILFE
Real ConsLoans	Mon	5	Consumer (Individual) Loans at All Commercial Banks Outlier Code because of change in data in April 2010 see FRB H8 ReleasDefl by PCEPILFE
Real NonRevCredit	Mon	5	Total Nonrevolving Credit Owned and Securitized, OutstandingDefl by PCEPILFE
Real LoansRealEst	Mon	5	Real Estate Loans at All Commercial BanksDefl by PCEPILFE
Real RevolvCredit	Mon	5	Total Revolving Credit OutstandingDefl by PCEPILFE
S&P500	StPr	5	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
DJIA	StPr	5	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
VXO	StPr	1	VXO (Linked by N. Bloom) .. Average daily VIX from 2009
Ex rate: Major	ExR	5	FRB Nominal Major Currencies Dollar Index (Linked to EXRUS in 1973:1)
Ex rate: Switz	ExR	5	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER USD)
Ex rate: Japan	ExR	5	FOREIGN EXCHANGE RATE: JAPAN (YEN PER USD)
Ex rate: UK	ExR	5	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
EX rate: Canada	ExR	5	FOREIGN EXCHANGE RATE: CANADA (CAD PER USD)
Cons. Expectations	Others	1	Consumer expectations NSA (Copyright, University of Michigan)

Notes: TC is transformation code; see Stock and Watson (2012).

B Some Auxiliary Results

We first present a lemma on the transformation matrices R_a and R_b defined in (2.8) and (2.10) of the main text. This lemma is used in the proof of Theorem 1. Let $\tilde{F}_a^r \in R^{T_0 \times r_a}$ and $\tilde{F}_b^r \in R^{(T-T_0) \times r_b}$ denote the first r_a and r_b columns of \tilde{F}_a and \tilde{F}_b , respectively. The $r_a \times r_a$ diagonal matrix \tilde{V}_a consists of the first r_a largest eigenvalues of $(T_0 N)^{-1} X_a X_a'$ in a decreasing order, and the $r_b \times r_b$ diagonal matrix \tilde{V}_b consists of the first r_b largest eigenvalues of $(T_1 N)^{-1} X_b X_b'$ in a decreasing order. Under Assumptions A-D, Theorem 1 of Bai and Ng (2002) shows that

$$T_0^{-1} \|\tilde{F}_a^r - F_a H_a\|^2 = O_p(C_{NT_0}^{-2}) \text{ and } T_1^{-1} \|\tilde{F}_b^r - F_b H_b\|^2 = O_p(C_{NT_1}^{-2}), \quad (\text{B.1})$$

where

$$H_a = \Sigma_a \frac{F_a' \tilde{F}_a^r}{T_0} \tilde{V}_a^{-1} \text{ and } H_b = \Sigma_b \frac{F_b' \tilde{F}_b^r}{T_1} \tilde{V}_b^{-1}. \quad (\text{B.2})$$

Lemma 3 *Suppose that Assumptions A-D hold. Then,*

$$H_a - R_a = O_p(C_{NT}^{-1}) \text{ and } H_b - R_b = O_p(C_{NT}^{-1}).$$

Proof of Lemma 3. Note that R_a is invertible w.p.a.1. Hence, we can write

$$F_a \Lambda^{0'} = F_a R_a R_a^{-1} \Lambda^{0'} = F_a^R \Lambda^{R'}, \text{ where } F_a^R = F_a R_a \text{ and } \Lambda^{R'} = R_a^{-1} \Lambda^{0'}. \quad (\text{B.3})$$

The transformed factors satisfy

$$\begin{aligned} \frac{F_a^R \Lambda^{R'}}{T_0} &= V_a^{-1/2} \Upsilon_a' \Sigma_a^{1/2} \frac{F_a' F_a}{T_0} \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2} \\ &= V_a^{-1/2} (\Upsilon_a' \Sigma_a^{1/2} \Sigma_F \Sigma_a^{1/2} \Upsilon_a) V_a^{-1/2} + O_p(T_0^{-1/2}) \\ &= V_a^{-1/2} (V_a) V_a^{-1/2} + O_p(T_0^{-1/2}) = I_{r_a} + O_p(T_0^{-1/2}), \end{aligned} \quad (\text{B.4})$$

where the first equality follows from $F_a^R = F_a R_a$ and $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$, the second equality follows from $F_a' F_a / T_0 - \Sigma_F = O_p(T_0^{-1/2})$ in Assumption A, and the third equality follows from (2.7). The transformed loadings satisfy

$$\frac{\Lambda^{R'} \Lambda^{R'}}{N} = V_a^{1/2} \Upsilon_a^{-1} \Sigma_a^{-1/2} \frac{\Lambda^{0'} \Lambda^{0'}}{N} \Sigma_a^{-1/2} \Upsilon_a^{-1} V_a^{1/2} = V_a^{1/2} \Upsilon_a^{-1} \Upsilon_a^{-1} V_a^{1/2} = V_a, \quad (\text{B.5})$$

where the first equality follows from $\Lambda^{R'} = R_a^{-1} \Lambda^{0'}$ and $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$, the second equality follows from $\Sigma_a = \Lambda^{0'} \Lambda^{0'} / N$ by definition, the third equality holds because $\Upsilon_a' \Upsilon_a = I_{r_a}$.

Let L_a be a $r_a \times r_a$ matrix defined as

$$L_a = \frac{\Lambda^{R'} \Lambda^R}{N} \frac{F_a^{R'} \tilde{F}_a^r}{T_0} \tilde{V}_a^{-1}, \quad (\text{B.6})$$

which is a transformation matrix analogous to H_a but with F_a and Λ^0 replaced by F_a^R and Λ^R , respectively. Stock and Watson (2002) and Bai and Ng (2002) show that L_a is invertible w.p.a.1 and \tilde{F}_a^r is a consistent estimator of $F_a^R L_a$. The transformation matrix H_a and the new transformation matrix L_a satisfy

$$\begin{aligned} H_a &= R_a \frac{R_a^{-1} \Lambda^{0'} \Lambda^0 R_a'^{-1}}{N} \frac{R_a' F_a' \tilde{F}_a^r}{T_0} \tilde{V}_a^{-1} \\ &= R_a \frac{\Lambda^{R'} \Lambda^R}{N} \frac{F_a^{R'} \tilde{F}_a^r}{T_0} \tilde{V}_a^{-1} = R_a L_a, \end{aligned} \quad (\text{B.7})$$

where the first equality follows from the definition of H_a in (B.2), the second equality follows from $F_a^R = F_a R_a$ and $\Lambda^{R'} = R_a^{-1} \Lambda^{0'}$, the third equality follows from the definition of L_a in (B.6).

Equation (2) of Bai and Ng (2013) shows that $L_a = I_{r_a}$ if the underlying factor matrix F_a^R satisfies $F_a^{R'} F_a^R / T_0 = I_r$, and the underlying loading matrix Λ^R satisfies that $\Lambda^{R'} \Lambda^R$ is a diagonal matrix with distinct elements. By (B.4) and (B.5), we know that these conditions are satisfied asymptotically by the transformation above. Following the arguments for equation (2) of Bai and Ng (2013), we obtain

$$L_a = I_{r_a} + O_p(C_{NT_0}^{-1}), \quad (\text{B.8})$$

with two modifications to the proof in Bai and Ng (2013): (i) $T_0^{-1}(\tilde{F}_a^r - F_a^R L_a)' F_a^R = O_p(C_{NT_0}^{-2})$ in Bai and Ng (2013) is changed to $T_0^{-1}(\tilde{F}_a^r - F_a^R L_a)' F_a^R = O_p(C_{NT_0}^{-1})$, which follows from $F_a^R L_a = F_a H_a$, (B.1), and the Cauchy-Schwarz inequality, and (ii) $F_a^{R'} F_a^R / T_0 = I_{r_a}$ is changed to $F_a^{R'} F_a^R / T_0 = I_{r_a} + O_p(T_0^{-1/2})$ and the $O_p(T_0^{-1/2})$ term is absorbed in $O_p(C_{NT_0}^{-1})$ in (B.8). The reason for the first change is that Assumptions A-D in this paper are comparable to Assumptions A – D of Bai and Ng (2002), which are weaker than similar assumptions in Bai and Ng (2013). The Assumptions in Bai and Ng (2013) are needed to obtain asymptotic distributions of the estimated factors and loadings, which is not the purpose here. After making these two modifications above, the rest of the arguments for equation (2) of Bai and Ng (2013) follow directly to yield the result in (B.8).

Combining the results in (B.7) and (B.8), we obtain $H_a - R_a = O_p(C_{NT}^{-1})$ because $T_0/T \rightarrow \pi_0 \in (0, 1)$. Similar arguments give $H_b - R_b = O_p(C_{NT}^{-1})$. \square

C Proof of Results in Section 4

Recall that we have defined

$$\Lambda^R = \Lambda^0(R_a^{-1})' \in R^{N \times r_a}, \Psi^R = \Psi^0(R_b^{-1})' \in R^{N \times r_b} \text{ and } \Gamma^R = (\Psi_1^R - \Lambda^R, \Psi_2^R) \quad (\text{C.1})$$

in (2.9), (2.11), and (2.12), respectively. For the ease of notation, we also define

$$\Lambda^* = (\Lambda^R, \mathbf{0}_{N \times (k-r_a)}), \Psi^* = (\Psi^R, \mathbf{0}_{N \times (k-r_b)}) \text{ and } \Gamma^* = \Psi^* - \Lambda^*. \quad (\text{C.2})$$

If $N^{-1} \|\Psi_\ell^R - \Lambda_\ell^R\|^2 \rightarrow 0$ as $N \rightarrow \infty$ for some ℓ , we replace the definition of Γ_ℓ^R and Γ_ℓ^* above with 0. The augmented matrices Λ^* and Ψ^* are transformed from Λ^+ and $\Psi^+ = \Lambda^+ + \Gamma^+$ defined in (3.1). Generally speaking, for the rest of the proof, the superscript 0 represents the true factor loadings, the superscript R represents transformed factor loadings, and the superscript asterisk represents augmented transformed factor loadings.

Following the definition of \mathcal{Z} in (5.11) and the definition of Γ^* ,

$$\mathcal{Z} = \{\ell = 1, \dots, k : \Gamma_\ell^* \neq 0\} \text{ and } \mathcal{Z}^C = \{\ell = 1, \dots, k : \Gamma_\ell^* = 0\}. \quad (\text{C.3})$$

By the definition of Γ^* , $\{r_b + 1, \dots, k\} \subseteq \mathcal{Z}^C$ and $\mathcal{Z} \subseteq \{1, \dots, r_b\}$. We allow $\ell \in \mathcal{Z}^C$ for some $\ell \leq r_b$ in the proofs below.

Recall $\widehat{\Lambda}$ and $\widehat{\Gamma}$ are the PLS estimators. Write $\widehat{\Psi} = \widehat{\Lambda} + \widehat{\Gamma}$. Define

$$Z_\lambda^2 = N^{-1} \left\| \widehat{\Lambda} - \Lambda^* \right\|^2, Z_\psi^2 = N^{-1} \left\| \widehat{\Psi} - \Psi^* \right\|^2, Z_\gamma^2 = N^{-1} \left\| \widehat{\Gamma} - \Gamma^* \right\|^2. \quad (\text{C.4})$$

Proof of Theorem 1. The criterion function for the shrinkage estimator can be written as

$$\begin{aligned} Q(\Lambda, \Gamma) &= M_a(\Lambda, \widetilde{F}_a) + M_b(\Psi, \widetilde{F}_b) + P_1(\Lambda) + P_2(\Gamma), \text{ where} \\ M_a(\Lambda, F_a) &= (NT)^{-1} \|X_a - F_a \Lambda'\|^2, \\ M_b(\Psi, F_b) &= (NT)^{-1} \|X_b - F_b(\Lambda + \Gamma)'\|^2, \\ P_1(\Lambda) &= \alpha_{NT} \sum_{\ell=1}^k \omega_\ell^\lambda \|\Lambda_\ell\| \text{ and } P_1(\Gamma) = \beta_{NT} \sum_{\ell=1}^k \omega_\ell^\gamma \|\Gamma_\ell\|, \end{aligned} \quad (\text{C.5})$$

with $\Psi = \Lambda + \Gamma$. For notational simplicity, the dependence on N and T is suppressed. Because the shrinkage estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ minimize the criterion function $Q(\Lambda, \Gamma)$, we have $Q(\widehat{\Lambda}, \widehat{\Gamma}) \leq Q(\Lambda^*, \Gamma^*)$, i.e.,

$$\begin{aligned} & \left[M_a(\widehat{\Lambda}, \widetilde{F}_a) - M_a(\Lambda^*, \widetilde{F}_a) \right] + \left[M_b(\widehat{\Psi}, \widetilde{F}_b) - M_b(\Psi^*, \widetilde{F}_b) \right] \\ & \leq \left[P_1(\Lambda^*) - P_1(\widehat{\Lambda}) \right] + \left[P_2(\Gamma^*) - P_2(\widehat{\Gamma}) \right], \end{aligned} \quad (\text{C.6})$$

where $\widehat{\Psi} = \widehat{\Lambda} + \widehat{\Gamma}$.

We start with the right-hand side of (C.6). Define

$$p_1 = P_1(\Lambda^*) - P_1^r(\widehat{\Lambda}) \text{ and } p_2 = \begin{cases} P_2(\Gamma^*) - P_2^r(\widehat{\Gamma}) & \text{if } \Gamma^0 \neq 0 \\ 0 & \text{if } \Gamma^0 = 0 \end{cases}, \text{ where}$$

$$P_1^r(\widehat{\Lambda}) = \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda \|\widehat{\Lambda}_\ell\| \leq \alpha_{NT} \sum_{\ell=1}^k \omega_\ell^\lambda \|\widehat{\Lambda}_\ell\| = P_1(\widehat{\Lambda}),$$

$$P_2^r(\widehat{\Gamma}) = \beta_{NT} \sum_{\ell \in \mathcal{Z}} \omega_\ell^\gamma \|\widehat{\Gamma}_\ell\| \leq \beta_{NT} \sum_{\ell=1}^k \omega_\ell^\gamma \|\widehat{\Gamma}_\ell\| = P_2(\widehat{\Gamma}). \quad (\text{C.7})$$

If $\Gamma^0 = 0$, we have $\Gamma^* = 0$ and $P_2(\Gamma^*) - P_2^r(\widehat{\Gamma}) \leq 0$ because $P_2(\Gamma^*) = 0$ and $P_2^r(\Gamma) \geq 0$. The penalty terms on the right-hand side of (C.6) satisfy

$$P_1(\Lambda^*) - P_1(\widehat{\Lambda}) \leq p_1 \text{ and } P_2(\Gamma^*) - P_2(\widehat{\Gamma}) \leq p_2 \quad (\text{C.8})$$

following the inequalities in (C.7).

We have $\Lambda_\ell^* = 0$ for $\ell = r_a + 1, \dots, k$ and $\Gamma_\ell^* = 0$ for $\ell \in \mathcal{Z}^C$, which implies that

$$P_1(\Lambda^*) = \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda \|\Lambda_\ell^*\| \text{ and } P_2(\Gamma^*) = \beta_{NT} \sum_{\ell \in \mathcal{Z}} \omega_\ell^\gamma \|\Gamma_\ell^*\|. \quad (\text{C.9})$$

Following (C.7), (C.9), the triangle inequality, and the Cauchy-Schwarz inequality, we have

$$p_1 \leq \alpha_{NT} \sum_{\ell=1}^{r_a} \omega_\ell^\lambda \left\| \widehat{\Lambda}_\ell - \Lambda_\ell^* \right\| \leq b_\Lambda Z_\lambda, \text{ where } b_\Lambda = N^{1/2} \alpha_{NT} \left[\sum_{\ell=1}^{r_a} (\omega_\ell^\lambda)^2 \right]^{1/2} \quad (\text{C.10})$$

and Z_λ is defined in (C.4). By the same arguments,

$$p_2 \leq b_\Gamma Z_\gamma, \text{ where } b_\Gamma = \begin{cases} N^{1/2} \beta_{NT} \left[\sum_{\ell \in \mathcal{Z}} (\omega_\ell^\gamma)^2 \right]^{1/2} & \text{if } \Gamma^0 \neq 0 \\ 0 & \text{if } \Gamma^0 = 0 \end{cases} \quad (\text{C.11})$$

and Z_γ is in (C.4). Combining (C.6) and (C.8)-(C.11), we obtain

$$\left[M_a(\widehat{\Lambda}, \widetilde{F}_a) - M_a(\Lambda^*, \widetilde{F}_a) \right] + \left[M_b(\widehat{\Psi}, \widetilde{F}_b) - M_b(\Psi^*, \widetilde{F}_b) \right] \leq b_\Lambda Z_\lambda + b_\Gamma Z_\gamma. \quad (\text{C.12})$$

Next, we consider the left-hand side of (C.12). To this end, we first show some useful equalities. Write $\widetilde{F}_a = (\widetilde{F}_a^r, \widetilde{F}_a^\perp) \in R^{T_0 \times k}$, where \widetilde{F}_a is partitioned into a $T_0 \times r_a$ submatrix \widetilde{F}_a^r and a $T_0 \times (k - r_a)$ submatrix \widetilde{F}_a^\perp . Replacing \widetilde{F}_a^r with $F_a^R = F_a R_a$, we define

$$F_a^* = (F_a^R, \widetilde{F}_a^\perp) = (F_a R_a, \widetilde{F}_a^\perp) \in R^{T_0 \times k}. \quad (\text{C.13})$$

Some equivalent relationships are useful in the calculation below

$$F_a^* \Lambda^{*'} = F_a^R \Lambda^{R'} = F_a \Lambda^{0'} \text{ and } \tilde{F}_a \Lambda^{*'} = \tilde{F}_a^r \Lambda^{R'}, \quad (\text{C.14})$$

because $\Lambda^* = (\Lambda^R, \mathbf{0}_{N \times (k-r_a)})$. It follows that

$$\begin{aligned} F_a \Lambda^{0'} - \tilde{F}_a \hat{\Lambda}' &= F_a^* \Lambda^{*'} - \tilde{F}_a \hat{\Lambda}' \\ &= (F_a^* - \tilde{F}_a) \Lambda^{*'} - \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \\ &= (F_a R_a - \tilde{F}_a^r) \Lambda^{R'} - \tilde{F}_a (\hat{\Lambda} - \Lambda^*)', \end{aligned} \quad (\text{C.15})$$

where the first equality follows from (C.14), the second equality follows from adding and subtracting $\tilde{F}_a \Lambda^{*'}$, and the third equality follows from (C.14). The difference between the true common component $F_a \Lambda^{0'}$ and the estimated common component $\tilde{F}_a \hat{\Lambda}'$ are decomposed into two pieces by the calculation in (C.15), where the first piece focuses on the factor estimation error and the second piece focuses on the factor loading estimation error.

The first term on the left-hand side of (C.12) satisfies

$$\begin{aligned} M_a(\hat{\Lambda}, \tilde{F}_a) &= (NT)^{-1} \left\| X_a - \tilde{F}_a \hat{\Lambda}' \right\|^2 \\ &= (NT)^{-1} \left\| e_a + (F_a \Lambda^{0'} - \tilde{F}_a \hat{\Lambda}') \right\|^2 \\ &= (NT)^{-1} \left\| \left(e_a + (F_a R_a - \tilde{F}_a^r) \Lambda^{R'} \right) - \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right\|^2 \\ &= M_1 + M_2 + M_3 + M_4, \end{aligned} \quad (\text{C.16})$$

where the first equality follows from the definition of $M_a(\Lambda, F_a)$ in (C.5), the second equality follows from $X_a = e_a + F_a \Lambda^{0'}$, the third equality holds by the decomposition in (C.15), and M_1, M_2, M_3 and M_4 are defined as follows. The first term M_1 is

$$\begin{aligned} M_1 &= (NT)^{-1} \left\| e_a + (F_a R_a - \tilde{F}_a^r) \Lambda^{R'} \right\|^2 \\ &= (NT)^{-1} \left\| X_a - \tilde{F}_a \Lambda^{*'} \right\|^2 = M_a(\Lambda^*, \tilde{F}_a), \end{aligned} \quad (\text{C.17})$$

following $X_a = e_a + F_a^R \Lambda^{R'}$, $\tilde{F}_a^r \Lambda^{R'} = \tilde{F}_a \Lambda^{*'}$ in (C.14) and the definition of $M_a(\Lambda, F)$ in (C.5). The second term M_2 is

$$\begin{aligned} M_2 &= (NT)^{-1} \left\| \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right\|^2 \\ &= (NT)^{-1} \text{tr} \left((\hat{\Lambda} - \Lambda^*) \tilde{F}_a' \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right) \\ &= \frac{T_0}{T} N^{-1} \left\| \hat{\Lambda} - \Lambda^* \right\|^2 = \frac{T_0}{T} Z_\lambda^2, \end{aligned} \quad (\text{C.18})$$

following $\tilde{F}'_a \tilde{F}_a / T_0 = I_{r_a}$ and the definition of Z_λ . The third term M_3 is

$$M_3 = -2(NT)^{-1} \text{tr} \left(e'_a \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right). \quad (\text{C.19})$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} (NT)^{-1} \left| \text{tr} \left(e'_a \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right) \right| &\leq (NT)^{-1} \left| \text{tr} \left(e'_a \tilde{F}_a \tilde{F}'_a e_a \right) \right|^{1/2} \left\| \hat{\Lambda} - \Lambda^* \right\| \\ &= N^{-1/2} T^{-1} \left| T_0 \text{tr} \left(P_{\tilde{F}_a} e_a e'_a \right) \right|^{1/2} Z_\lambda \\ &\leq N^{-1/2} T^{-1} \left| NT_0^2 k \rho_1 \left((NT_0)^{-1} e_a e'_a \right) \right|^{1/2} Z_\lambda \\ &= \frac{C_{3,n} Z_\lambda}{2}. \end{aligned} \quad (\text{C.20})$$

The first equality holds because $P_{\tilde{F}_a} = T_0^{-1} \tilde{F}_a \tilde{F}'_a$, $\text{tr}(AB) = \text{tr}(BA)$ for two matrices, and because of the definition of Z_λ . The second inequality follows from von Neumann's trace inequality and the fact that the eigenvalues of $P_{\tilde{F}_a}$ consist of k ones and $T - k$ zeros. By Assumption C(vi) and simple calculations,

$$\begin{aligned} C_{3,n} &= 2N^{-1/2} T^{-1} \left| NT_0^2 k \rho_1 \left((NT_0)^{-1} e_a e'_a \right) \right|^{1/2} \\ &= 2N^{-1/2} T^{-1} \left| NT_0^2 O_p(C_{NT}^{-2}) \right|^{1/2} \\ &= \frac{T_0}{T} O_p(C_{NT}^{-1}) = O_p(C_{NT}^{-1}), \end{aligned} \quad (\text{C.21})$$

which together with (C.19) and (C.20) implies

$$|M_3| \leq C_{3,n} Z_\lambda, \text{ where } C_{3,n} = O_p(C_{NT}^{-1}). \quad (\text{C.22})$$

The fourth term M_4 is

$$M_4 = -2(NT)^{-1} \text{tr} \left(\Lambda^R (F_a R_a - \tilde{F}_a^r)' \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right). \quad (\text{C.23})$$

To investigate M_4 , we note that

$$\frac{(F_a R_a - \tilde{F}_a^r)' \tilde{F}_a}{T_0} = \frac{(F_a H_a - \tilde{F}_a^r)' \tilde{F}_a}{T_0} + \frac{(F_a (R_a - H_a))' \tilde{F}_a}{T_0} = O_p(C_{NT}^{-1}) \quad (\text{C.24})$$

by the Cauchy-Schwarz inequality, (B.1), and Lemma 3. Applying the Cauchy-Schwarz inequality, we have

$$\begin{aligned} &(NT)^{-1} \left| \text{tr} \left(\Lambda^R (F_a R_a - \tilde{F}_a^r)' \tilde{F}_a (\hat{\Lambda} - \Lambda^*)' \right) \right| \\ &\leq (NT)^{-1} \left\| \Lambda^R \right\| \left\| (F_a R_a - \tilde{F}_a^r)' \tilde{F}_a \right\| \left\| \hat{\Lambda} - \Lambda^* \right\| \\ &= \frac{T_0}{T} \left(N^{-1} \left\| \Lambda^R \right\|^2 \right)^{1/2} \left\| \frac{(F_a R_a - \tilde{F}_a^r)' \tilde{F}_a}{T_0} \right\| \left(N^{-1} \left\| \hat{\Lambda} - \Lambda^* \right\|^2 \right)^{1/2} = \frac{C_{4,n} Z_\lambda}{2}. \end{aligned} \quad (\text{C.25})$$

Using $\Lambda^{R'} = R_a^{-1}\Lambda'$, $R_a^{-1} = O_p(1)$, $\|N^{-1}\Lambda'\Lambda - \Sigma_\Lambda\| \rightarrow 0$ and (C.24), we deduce that

$$C_{4,n} = \frac{2T_0}{T} \left(N^{-1} \|\Lambda^R\|^2 \right)^{1/2} \left\| \frac{(F_a R_a - \tilde{F}_a^r)' \tilde{F}_a}{T_0} \right\| = \frac{T_0}{T} O_p(C_{NT}^{-1}) = O_p(C_{NT}^{-1}), \quad (\text{C.26})$$

which together with (C.23) and (C.25) yields

$$|M_4| \leq C_{4,n} Z_\lambda, \text{ where } C_{4,n} = O_p(C_{NT}^{-1}). \quad (\text{C.27})$$

Putting the four terms in (C.17), (C.18), (C.22), and (C.27) into (C.16), we obtain

$$M_a(\widehat{\Lambda}, \tilde{F}_a) - M_a(\Lambda^*, \tilde{F}_a) \geq \frac{T_0}{T} Z_\lambda^2 - C_{a,n} Z_\lambda, \text{ where } C_{a,n} = C_{3,n} + C_{4,n} = O_p(C_{NT}^{-1}). \quad (\text{C.28})$$

Replacing the first subsample with the second subsample and the factor loadings Λ with Ψ , we also have

$$M_b(\widehat{\Psi}, \tilde{F}_b) - M_b(\Psi^*, \tilde{F}_b) \geq \frac{T_1}{T} Z_\psi^2 - C_{b,n} Z_\psi, \text{ where } C_{b,n} = O_p(C_{NT}^{-1}). \quad (\text{C.29})$$

Plugging (C.28) and (C.29) into the left-hand side of (C.12), we obtain

$$\frac{T_0}{T} Z_\lambda^2 - C_{a,n} Z_\lambda + \frac{T_1}{T} Z_\psi^2 - C_{b,n} Z_\psi \leq b_\Lambda Z_\lambda + b_\Gamma Z_\gamma \leq (b_\Lambda + b_\Gamma) Z_\lambda + b_\Gamma Z_\psi, \quad (\text{C.30})$$

following the triangle inequality. Rearranging (C.30) gives

$$\begin{aligned} & \pi_0 \left(Z_\lambda - \frac{C_{a,n} + b_\Lambda + b_\Gamma}{2\pi_0} \right)^2 + \pi_1 \left(Z_\psi - \frac{C_{b,n} + b_\Gamma}{2\pi_1} \right)^2 \\ & \leq \pi_0 \left(\frac{C_{a,n} + b_\Lambda + b_\Gamma}{2\pi_0} \right)^2 + \pi_1 \left(\frac{C_{b,n} + b_\Gamma}{2\pi_1} \right)^2, \end{aligned} \quad (\text{C.31})$$

where $\pi_0 = T_0/T \in (0, 1)$ and $\pi_1 = 1 - \pi_0$. It follows from (C.31), $C_{a,n} = O_p(C_{NT}^{-1})$, $C_{b,n} = O_p(C_{NT}^{-1})$, and the triangle inequality that

$$\begin{aligned} Z_\lambda &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}), \\ Z_\psi &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}), \\ Z_\gamma &= O_p(b_\Lambda + b_\Gamma + C_{NT}^{-1}). \end{aligned} \quad (\text{C.32})$$

Assumptions P1 and P2 imply that

$$\omega_\ell^\lambda = O_p(1) \text{ for } \ell = 1, \dots, r_a, \quad \omega_\ell^\gamma = O_p(1) \text{ for } \ell \in \mathcal{Z}. \quad (\text{C.33})$$

Assumption T(i) implies that $b_\Lambda = O_p(C_{NT}^{-1})$ and $b_\Gamma = O_p(C_{NT}^{-1})$, following (C.33). It follows from (C.32) that

$$Z_\lambda = O_p(C_{NT}^{-1}) \text{ and } Z_\gamma = O_p(C_{NT}^{-1}). \quad (\text{C.34})$$

Theorems 1(a) and 1(c) follow from the definitions of Z_λ and Z_γ in (C.4) and the results in (C.34).

Next, we show the superefficiency results in Theorems 1(b), 1(d), and 1(e). To this end, first define

$$\mathcal{L}_a = \{\ell : (\omega_\ell^\lambda)^{-1} = O_p(C_{NT}^{-2d})\} \text{ and } \mathcal{L}_b = \{\ell : (\omega_\ell^\gamma)^{-1} = O_p(C_{NT}^{-2d})\}. \quad (\text{C.35})$$

Under Assumptions P1 and P2,

$$\{r_a + 1, \dots, k\} \subseteq \mathcal{L}_a, \{r_b + 1, \dots, k\} \subseteq \mathcal{L}_b, \text{ and if } \Gamma^0 = 0, \{1, \dots, k\} = \mathcal{L}_b. \quad (\text{C.36})$$

Define the residual matrices

$$e_a(\widehat{\Lambda}) = X_a - \widetilde{F}_a \widehat{\Lambda}' \in R^{T_0 \times N} \text{ and } e_b(\widehat{\Lambda} + \widehat{\Gamma}) = X_b - \widetilde{F}_b(\widehat{\Lambda} + \widehat{\Gamma})' \in R^{T_1 \times N}. \quad (\text{C.37})$$

Let $e_t^a(\widehat{\Lambda})$ for $t = 1, \dots, T_0$ be the rows of $e_a(\widehat{\Lambda})$ and $e_t^b(\widehat{\Lambda} + \widehat{\Gamma})$ for $t = T_0 + 1, \dots, T$ be the rows of $e_b(\widehat{\Lambda} + \widehat{\Gamma})$. Let $\widetilde{F}_\ell = (\widetilde{F}'_{a,\ell}, \widetilde{F}'_{b,\ell})' \in R^{T \times 1}$, where $\widetilde{F}_{a,\ell}$ and $\widetilde{F}_{b,\ell}$ are the ℓ -th columns of \widetilde{F}_a and \widetilde{F}_b , respectively, and let $\widetilde{F}_{t,\ell}$ denote the t -th row of \widetilde{F}_ℓ . By Lemma 4.2 of Bühlmann and van de Geer (2011), a sufficient condition for $\widehat{\Lambda}_\ell = 0$ is

$$2(NT)^{-1} \left\| \sum_{t=1}^{T_0} e_t^a(\widehat{\Lambda}) \widetilde{F}_{t,\ell} + \sum_{t=T_0+1}^T e_t^b(\widehat{\Lambda} + \widehat{\Gamma}) \widetilde{F}_{t,\ell} \right\| < \alpha_{NT} \omega_\ell^\lambda, \quad (\text{C.38})$$

where the left-hand side is associated with the partial derivative of $M_a(\Lambda, \widetilde{F}_a) + M_b(\Psi, \widetilde{F}_b)$, with respect to Λ_ℓ evaluated at the PLS estimators, and the right-hand side is the marginal penalty once $\widehat{\Lambda}_\ell$ deviates from 0. Intuitively, the optimal solution is $\widehat{\Lambda}_\ell = 0$ when the marginal penalty on the right-hand side of (C.38) is larger than the marginal gain on the left-hand side of (C.38). The inequality in (C.38) can be equivalently written as

$$\left\| e^a(\widehat{\Lambda})' \widetilde{F}_{a,\ell} + e^b(\widehat{\Lambda} + \widehat{\Gamma})' \widetilde{F}_{b,\ell} \right\| < \frac{NT}{2} \alpha_{NT} \omega_\ell^\lambda, \quad (\text{C.39})$$

which holds provided that

$$\left\| e^a(\widehat{\Lambda})' \widetilde{F}_{a,\ell} \right\| + \left\| e^b(\widehat{\Lambda} + \widehat{\Gamma})' \widetilde{F}_{b,\ell} \right\| < \frac{NT}{2} \alpha_{NT} \omega_\ell^\lambda. \quad (\text{C.40})$$

Next, we study the two terms on the left-hand side of (C.40). The first term satisfies

$$\begin{aligned}
\left\| e^a(\widehat{\Lambda})' \widetilde{F}_{a,\ell} \right\| &= \left\| (e_a + F_a \Lambda^{0l} - \widetilde{F}_a \widehat{\Lambda}')' \widetilde{F}_{a,\ell} \right\| \\
&= \left\| e'_a \widetilde{F}_{a,\ell} + (F_a R_a - \widetilde{F}_a^r) \Lambda^{Rl} \widetilde{F}_{a,\ell} - \widetilde{F}_a (\widehat{\Lambda} - \Lambda^*)' \widetilde{F}_{a,\ell} \right\| \\
&\leq \left\| e'_a \widetilde{F}_{a,\ell} \right\| + \left\| F_a R_a - \widetilde{F}_a^r \right\| \left\| \Lambda^{Rl} \right\| \left\| \widetilde{F}_{a,\ell} \right\| + \left\| \widetilde{F}_a \right\| \left\| \widehat{\Lambda} - \Lambda^* \right\| \left\| \widetilde{F}_{a,\ell} \right\| \quad (\text{C.41})
\end{aligned}$$

where the second equality follows from (C.15) and the inequality follows from the Cauchy-Schwarz inequality and the triangle inequality. The terms in the last line of (C.41) are:

(i)

$$\begin{aligned}
\|e'_a \widetilde{F}_{a,\ell}\| &= (NT)^{1/2} \sqrt{\widetilde{F}'_{a,\ell} \frac{e_a e'_a}{NT} \widetilde{F}_{a,\ell}} \\
&\leq (NT)^{1/2} T_0^{1/2} \sqrt{\rho_1 ((NT)^{-1} e_a e'_a)} \sqrt{\frac{\widetilde{F}'_{a,\ell} \widetilde{F}_{a,\ell}}{T_0}} \\
&= (NT)^{1/2} T_0^{1/2} O_p(C_{NT}^{-1}) = O_p(N^{1/2} T C_{NT}^{-1}), \quad (\text{C.42})
\end{aligned}$$

where the second equality is by $T_0^{-1} \widetilde{F}'_{a,\ell} \widetilde{F}_{a,\ell} = 1$ and Assumption C(vi); (ii) $\|F_a R_a - \widetilde{F}_a^r\| = O_p(T^{1/2} C_{NT}^{-1})$ by (B.1); (iii) $\|\Lambda^{Rl}\| = O_p(N^{1/2})$ because $R_a = O_p(1)$ and $\|\Lambda' \Lambda / N - \Sigma_\Lambda\| \rightarrow 0$; (iv) $\|\widetilde{F}_{a,\ell}\| = O(T^{1/2})$ and $\|\widetilde{F}_a\| = O(T^{1/2})$ because $T_0^{-1} \widetilde{F}'_a \widetilde{F}_a = I_{r_a}$; (v) $\|\widehat{\Lambda} - \Lambda^*\| = O_p(N^{1/2} C_{NT}^{-1})$ by the definition of Z_λ and (C.34). Putting them together with (C.41), we have

$$\left\| e^a(\widehat{\Lambda})' \widetilde{F}_{a,\ell} \right\| = O_p(N^{1/2} T C_{NT}^{-1}). \quad (\text{C.43})$$

By the same arguments, we have

$$\left\| e^b(\widehat{\Lambda} + \widehat{\Gamma})' \widetilde{F}_{b,\ell} \right\| = O_p(N^{1/2} T C_{NT}^{-1}). \quad (\text{C.44})$$

Equations (C.43) and (C.44) imply that for the inequality in (C.40) to hold, it suffices to have

$$N^{-1/2} C_{NT}^{-1} = o_p(\alpha_{NT} \omega_\ell^\lambda), \quad (\text{C.45})$$

which is satisfied for all $\ell \in \mathcal{L}_a$ under Assumption T(ii).

To prove Theorems 1(d) and 1(e), note that a sufficient condition for $\widehat{\Gamma}_\ell = 0$ is

$$2(NT)^{-1} \left\| \sum_{t=T_0+1}^T e_t^b(\widehat{\Lambda} + \widehat{\Gamma}) \widetilde{F}_{t,\ell} \right\| < \beta_{NT} \omega_\ell^\gamma. \quad (\text{C.46})$$

Following (C.44), the inequality in (C.46) holds provided that

$$N^{-1/2}C_{NT}^{-1} = o_p(\beta_{NT}\omega_\ell^\gamma), \quad (\text{C.47})$$

which is satisfied for all $\ell \in \mathcal{L}_b$ under Assumption T(ii). Therefore, Theorems 1(b), 1(d), and 1(e) follow from (C.36).

Some remarks on the proof of Theorem 1 and its relationship to the proofs of Corollaries 1 and 2 below are in order. First, in the proof of Theorem 1, we give general definition of \mathcal{Z} , \mathcal{L}_a and \mathcal{L}_b without imposing Assumptions P1 and P2 so that the the proof can be recycled when these assumptions are relaxed. Specifically, Theorem 1 can be proved as above without Assumptions P1 and P2 as long as (C.33) and (C.36) can be verified for a given preliminary estimator, as we shall do in the proofs below. Second, Assumptions P1 and P2 are slightly stronger than needed to prove Theorem 1, however, we present them as is for the simplicity of the presentation to convey the idea. These assumptions can be relaxed as follows: Assumption P1(ii) assumes that $\Pr(N^{-1}|\tilde{\Gamma}_\ell|^2 \geq C) \rightarrow 1$ for $\ell = 1, \dots, r_b$, while we only need this to hold for $\ell \in \mathcal{Z}$ rather than for all $\ell = 1, \dots, r_b$ in order to verify (C.33). The set \mathcal{Z} , associated with the nonzero columns of Γ^R , could be a subset of $\{1, \dots, r_b\}$ to identify a type-1 or type-2 change. For this reason, the proofs of Corollaries 1 and 2 do not verify Assumptions P1 and P2 but rather show Theorem 1 directly. \square

Proof of Lemma 1. Because $\Lambda^R = \Lambda^0 R_a^{-1'}$ and $\Psi^R = \Psi^0 R_b^{-1'}$ with $R_a = \Sigma_a^{1/2} \Upsilon_a V_a^{-1/2}$ and $R_b = \Sigma_b^{1/2} \Upsilon_b V_b^{-1/2}$, we have

$$\frac{\Lambda^{R'} \Lambda^R}{N} = V_a^{1/2} \Upsilon_a' \Sigma_a^{-1/2} \frac{\Lambda^{0'} \Lambda^0}{N} \Sigma_a^{-1/2} \Upsilon_a V_a^{1/2} = V_a \text{ and } \frac{\Psi^{R'} \Psi^R}{N} = V_b. \quad (\text{C.48})$$

By definition, V_a is a diagonal matrix and its ℓ -th diagonal element is the ℓ -th largest eigenvalue of $\Sigma_a^{1/2} \Sigma_F \Sigma_a^{1/2}$, which is the same as the ℓ -th largest eigenvalue of $\Sigma_a \Sigma_F$. Following Assumption B and the continuity of the eigenvalue (with respect to the matrix), it converges to the ℓ -th largest eigenvalue of $\Sigma_\Lambda \Sigma_F$, denoted by $\rho_\ell(\Sigma_\Lambda \Sigma_F)$. Similarly, the ℓ -th diagonal element of V_b converges to the ℓ -th largest eigenvalue of $\Sigma_\Psi \Sigma_{\bar{F}}$, denoted by $\rho_\ell(\Sigma_\Psi \Sigma_{\bar{F}})$.

Let a_ℓ be a selection vector that selects the ℓ -th column of a matrix. Part (a) holds because

$$N^{-1} \|\Lambda_\ell^R\|^2 = a_\ell' (N^{-1} \Lambda^{R'} \Lambda^R) a_\ell = a_\ell' V_a a_\ell = \rho_\ell(\Sigma_\Lambda \Sigma_F) + o(1). \quad (\text{C.49})$$

To prove part (b), note that for $r_a < \ell \leq r_b$, the ℓ -th column of Γ^R is equivalent to the ℓ -th column of Ψ^R . Hence,

$$N^{-1} \|\Gamma_\ell^R\|^2 = a'_\ell (N^{-1} \Psi^{R'} \Psi^R) a_\ell = a'_\ell V_b a_\ell = \rho_\ell(\Sigma_\Psi \Sigma_F) + o(1). \quad (\text{C.50})$$

To show part (c), first note that if $r_a = r_b$, we have

$$N^{-1} \Gamma^{R'} \Gamma^R = N^{-1} (\Psi^R - \Lambda^R)' (\Psi^R - \Lambda^R) = \mathbf{e}' \Sigma_{\Lambda\Psi}^+ \mathbf{e} + o(1), \quad (\text{C.51})$$

where $\mathbf{e} = \lim_{N \rightarrow \infty} (R_b^{-1}, -R_a^{-1})'$ has full rank following Assumptions A and B and $\Sigma_{\Lambda\Psi}^+$ is defined in (2.6). By a Cholesky decomposition, write $\Sigma_{\Lambda\Psi}^+ = (\Sigma_{\Lambda\Psi}^+)^{1/2} (\Sigma_{\Lambda\Psi}^+)^{1/2}$ with $\text{rank}((\Sigma_{\Lambda\Psi}^+)^{1/2}) = \text{rank}(\Sigma_{\Lambda\Psi}^+) > r_a$. For a $2r_a \times 2r_a$ matrix $(\Sigma_{\Lambda\Psi}^+)^{1/2}$, the rank of the null space of $(\Sigma_{\Lambda\Psi}^+)^{1/2}$ is smaller than r_a . It follows that $(\Sigma_{\Lambda\Psi}^+)^{1/2} \mathbf{e} \neq 0$ because $\text{rank}(\mathbf{e}) = r_a$, and this immediately implies that part (c) holds with $\Sigma_\Gamma = \mathbf{e}' \Sigma_{\Lambda\Psi}^+ \mathbf{e} \neq 0$.

To prove part (d), write

$$\begin{aligned} N^{-1} \|\Gamma_\ell^R\|^2 &= N^{-1} \|\Gamma^R a_\ell\|^2 = N^{-1} \|\Psi^R a_\ell - \Lambda^R a_\ell\|^2 \\ &\geq (N^{-1/2} \|\Psi^R a_\ell\| - N^{-1/2} \|\Lambda^R a_\ell\|)^2 \\ &= [(\rho_\ell(\Sigma_\Psi \Sigma_F))^{1/2} - (\rho_\ell(\Sigma_\Lambda \Sigma_F))^{1/2}]^2 + o(1), \end{aligned} \quad (\text{C.52})$$

where the first two equalities follow from the definition of a_ℓ and Γ^R , the inequality follows from the triangle inequality, and the last equality holds by (C.48). \square

Proof of Theorem 2. First, Theorem 1(a) for $\ell = r_a$ and Lemma 1(a) imply that $\Pr(\|\widehat{\Lambda}_\ell\| > 0) \rightarrow 1$ for $\ell = r_a$ and thus $\Pr(\widehat{r}_a \geq r_a) \rightarrow 1$. Theorem 1(b) implies that $\Pr(\widehat{r}_a \leq r_a) \rightarrow 1$. Thus, $\Pr(\widehat{r}_a = r_a) \rightarrow 1$.

Second, for a type-2 change where $r_b > r_a$, Theorem 1(c) for $\ell = r_b$ and Lemma 1(b) imply that $\Pr(\|\widehat{\Gamma}_\ell\| > 0) \rightarrow 1$ for $\ell = r_b$ and thus $\Pr(\widehat{r}_b \geq r_b) \rightarrow 1$. Theorem 1(e) implies that $\Pr(\widehat{r}_b \leq r_b) \rightarrow 1$. Hence, $\Pr(\widehat{r}_b = r_b) \rightarrow 1$ for a type-2 change, which, together with part (a), also implies $\Pr(\widehat{\mathcal{S}} = 1) \rightarrow 1$ for a type-2 change because by definition, $\widehat{\mathcal{S}} = 1$ if $\widehat{r}_b > \widehat{r}_a$.

Third, for a type-1 change where $r_b = r_a$ and $\mathcal{S}_0 = 1$, Theorem 1(c), Lemmas 1(c) and 1(d), and Assumption ID imply that $\Pr(\|\widehat{\Gamma}_\ell\| > 0) \rightarrow 1$ for some $\ell \leq r_a$ and thus $\Pr(\widehat{\mathcal{S}} = 1) \rightarrow 1$. Note that by definition in (3.5), we have $\widehat{r}_b \geq \widehat{r}_a$. Thus, part (a) and $r_a = r_b$ imply that $\Pr(\widehat{r}_b \geq r_b) \rightarrow 1$. On the other hand, Theorem 1(e) implies that $\Pr(\widehat{r}_b \leq r_b) \rightarrow 1$. Hence, $\Pr(\widehat{r}_b = r_b) \rightarrow 1$ for a type-1 change.

Finally, for the case where there is no change, i.e., $r_a = r_b$ and $\mathcal{S}_0 = 0$, Theorems 1(d) and 1(e) imply that $\Pr(\widehat{\Gamma} = 0) \rightarrow 1$. Thus, $\Pr(\widehat{\mathcal{S}} = 0) \rightarrow 1$ by (3.6) and $\Pr(\widehat{r}_b = r_b) \rightarrow 1$ by (3.5) and part (a). \square

Proof of Corollary 1. We first study the properties of the unrestricted least square estimator $\widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma}_{LS}$. Note that the unrestricted least squares estimator is a special case of the PLS estimator when $\alpha_{NT} = \beta_{NT} = 0$. Therefore, following (C.32),

$$N^{-1} \|\widetilde{\Lambda}_{LS} - \Lambda^*\|^2 = O_p(C_{NT}^{-2}) \text{ and } N^{-1} \|\widetilde{\Gamma}_{LS} - \Gamma^*\|^2 = O_p(C_{NT}^{-2}), \quad (\text{C.53})$$

which combined with the definitions of Λ^* and Γ^* and Lemma 1 imply that

$$\Pr(N^{-1} \|\widetilde{\Lambda}_{LS,\ell}\|^2 \geq C) \rightarrow 1 \text{ for } \ell = 1, \dots, r_a, \quad \Pr(N^{-1} \|\widetilde{\Gamma}_{LS,\ell}\|^2 \geq C) \rightarrow 1 \text{ for } \ell \in \mathcal{Z} \quad (\text{C.54})$$

and

$$N^{-1} \|\widetilde{\Lambda}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2}) \text{ for } \ell > r_a \text{ and } N^{-1} \|\widetilde{\Gamma}_{LS,\ell}\|^2 = O_p(C_{NT}^{-2}) \text{ for } \ell \in \mathcal{Z}^C. \quad (\text{C.55})$$

Next, we show that (C.33) and (C.36) hold without imposing Assumptions P1 and P2, so that the proof of Theorem 1 follows without these two assumptions. The definition of weights in (3.4) and (C.54) imply that (C.33) holds for the case $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. The definition of \mathcal{L}_a and \mathcal{L}_b together with (C.55) imply that $\mathcal{L}_a = \{r_a + 1, \dots, k\}$ and $\mathcal{L}_b = \mathcal{Z}^C$. By definition, $\{r_b + 1, \dots, k\} \subseteq \mathcal{Z}^C$ and, if $\Gamma^0 = 0$, then $\{1, \dots, k\} = \mathcal{Z}^C$, which implies that (C.36) holds for the case $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. Therefore, Theorem 1 holds without imposing Assumptions P1 and P2 for the one-step estimator $\widetilde{\Lambda} = \widetilde{\Lambda}_{LS}$ and $\widetilde{\Gamma} = \widetilde{\Gamma}_{LS}$. Applying Theorem 1, model selection consistency follows from the proof for Theorem 2. \square

Proof of Corollary 2. We first study the preliminary estimators $\widetilde{\Lambda}^{(2)}$, $\widetilde{\Psi}^{(2)}$, and $\widetilde{\Gamma}^{(2)}$, and the weights ω_ℓ^λ and ω_ℓ^γ in the second step. Because $\widetilde{\Lambda}^{(2)} = \widehat{\Lambda}_{PMS}^{(1)}$, whose first $\widehat{r}_a^{(1)}$ columns are the same as those of $\widetilde{\Lambda}_{LS}$ and whose last $k - \widehat{r}_a^{(1)}$ columns are zeros, it follows from (3.4) that

$$\omega_\ell^\lambda = (N^{-1} \|\widetilde{\Lambda}_{LS,\ell}\|^2)^{-d} \text{ for } \ell = 1, \dots, k, \quad (\text{C.56})$$

which is the same for the first- and second-step estimators. If there is a type-2 change, $\widehat{r}_b^{(1)} > \widehat{r}_a^{(1)}$ w.p.a.1 by Corollary 1, and

$$\omega_\ell^\gamma = (N^{-1} \|\widetilde{\Gamma}_{LS,\ell}\|^2)^{-d} \text{ for } \ell = 1, \dots, k, \quad (\text{C.57})$$

which is the same for the first and second step estimations.

If there are no structural instabilities or there is a type-1 change, $\hat{r}_b^{(1)} = \hat{r}_a^{(1)} = r_b = r_a$ w.p.a.1 by Corollary 1. Let $\tilde{\Psi}_{LS}^-$ and $\tilde{\Lambda}_{LS}^-$ denote the first r_a columns of $\tilde{\Psi}_{LS}$ and $\tilde{\Lambda}_{LS}$, respectively. Given $\hat{r}_b^{(1)} = \hat{r}_a^{(1)} = r_a = r_b$, we have $\bar{\Psi}^{(1)} = \tilde{\Psi}_{LS}^-$, $\bar{\Lambda}^{(1)} = \tilde{\Lambda}_{LS}^-$, and the second-step preliminary estimator $\tilde{\Gamma}^{(2)}$ can be written as

$$\tilde{\Gamma}^{(2)} = \left(\tilde{\Psi}_{LS}^- Q - \tilde{\Lambda}_{LS}^-, 0_{N \times (k-r_a)} \right), \quad (\text{C.58})$$

following from $\tilde{\Gamma}^{(2)} = \tilde{\Psi}^{(2)} - \tilde{\Lambda}^{(2)}$ and steps 1d, 1e, and 2a in the algorithm to construct the two-step estimator.

Define

$$\Gamma^Q = \left(\Psi^R Q - \Lambda^R, 0_{N \times (k-r_a)} \right). \quad (\text{C.59})$$

Recall that Ψ^R and Λ^R are the transformed factor loadings. In addition, Γ^R and Λ^R are the first r_a columns of Γ^* and Λ^* , respectively, given $r_a = r_b$. By (C.58) and (C.59), w.p.a.1,

$$\begin{aligned} N^{-1} \|\tilde{\Gamma}^{(2)} - \Gamma^Q\|^2 &= N^{-1} \left\| (\tilde{\Psi}_{LS}^- - \Psi^R) Q - (\tilde{\Lambda}_{LS}^- - \Lambda^R) \right\|^2 \\ &= N^{-1} \left\| \left(\tilde{\Gamma}_{LS}^- - \Gamma^R \right) Q + \left(\tilde{\Lambda}_{LS}^- - \Lambda^R \right) (Q - I_{r_a}) \right\|^2 \\ &= O_p(C_{NT}^{-2}), \end{aligned} \quad (\text{C.60})$$

where the last equality follows from the triangle inequality and (C.53). To analyze $\tilde{\Gamma}^{(2)}$ for the second-step estimation, we first discuss the centering term Γ^Q when there is a type-1 change. Assumption R implies that

$$N^{-1} \|\Gamma_\ell^Q\|^2 \geq C \text{ if } \ell \in \mathcal{Z} \quad (\text{C.61})$$

because $\Gamma_\ell^Q = \Psi^R Q_\ell - \Lambda_\ell^R$ and $\|Q_\ell\| = 1$. Therefore, (C.60) and (C.61) imply that

$$\omega_\ell^\gamma = O_p(1) \text{ for } \ell \in \mathcal{Z} \text{ when there is a type-1 change.} \quad (\text{C.62})$$

If there is no structural change, by (C.53), $N^{-1} \|\tilde{\Lambda}_{LS}^- - \Lambda^R\|^2 = O_p(C_{NT}^{-2})$ and $N^{-1} \|\tilde{\Psi}_{LS}^- - \Psi^R\|^2 = O_p(C_{NT}^{-2})$. Because $\Lambda^R = \Psi^R$ in this case, we have $N^{-1} \|\tilde{\Lambda}_{LS}^- - \tilde{\Psi}_{LS}^-\|^2 = O_p(C_{NT}^{-2})$, which further implies that

$$N^{-1} \left\| \tilde{\Psi}_{LS}^- Q - \tilde{\Lambda}_{LS}^- \right\|^2 \leq N^{-1} \left\| \tilde{\Psi}_{LS}^- - \tilde{\Lambda}_{LS}^- \right\|^2 = O_p(C_{NT}^{-2}), \quad (\text{C.63})$$

where the inequality holds because the choice of Q solves the orthogonal procrustes problem by minimizing $\|\tilde{\Psi}_{LS}^- Q - \tilde{\Lambda}_{LS}^-\|^2$ among all orthogonal matrices (Schönemann (1966)). Combining (C.58) and (C.63), we obtain

$$N^{-1}\|\tilde{\Gamma}^{(2)}\|^2 = O_p(C_{NT}^{-2}) \text{ when } \Gamma^0 = 0, \quad (\text{C.64})$$

which together with (C.53) and $\Gamma^* = 0$ implies that

$$(\omega_\ell^\gamma)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1, \dots, k \text{ when there is no structural change.} \quad (\text{C.65})$$

Next, we show that (C.33) and (C.36) hold without imposing Assumptions P1 and P2, so that the proof of Theorem 1 follows without these two assumptions. To show (C.33), note that $\omega_\ell^\lambda = O_p(1)$ for $\ell = 1, \dots, r_a$ is implied by (C.54) and (C.56), $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is implied by (C.54) and (C.57) for a type-2 change, and $\omega_\ell^\gamma = O_p(1)$ for $\ell \in \mathcal{Z}$ is proved in (C.62) for a type-1 change.

To show (C.36), note that: (i) $\{r_a + 1, \dots, k\} \subseteq \mathcal{L}_a$ holds by (C.55) and (C.56); (ii) $\{r_b + 1, \dots, k\} \subseteq \mathcal{L}_b$ holds by (C.55) and (C.57); and (iii) if $\Gamma^0 = 0$, $\{1, \dots, k\} = \mathcal{L}_b$ follows from (C.53) and (C.65).

Because (C.33) and (C.36) hold without imposing Assumptions P1 and P2, Theorem 1 holds without imposing Assumptions P1 and P2 for the two-step estimator. Applying Theorem 1, model selection consistency follows from the proof for Theorem 2. \square

D Proof of Results in Section 6

Proof of Lemma 2. For $\pi \leq \pi_0$, the result follows from the representation in (6.3) and Assumptions A-D. Analogous arguments yield results for $\pi > \pi_0$. \square

Proof of Corollary 3. This corollary is implied by Lemma 2.

Proof of Theorem 3. In the proof below, we use $o_{p\pi}(\cdot)$ and $O_{p\pi}(\cdot)$ to represent $o_p(\cdot)$ and $O_p(\cdot)$ that hold uniformly over $\pi \in \Pi$.

Define $r^+ = \text{rank}(\Sigma_{\Lambda\Psi}^+)$, $T_a = \lfloor T\pi \rfloor$, and $T_b = T - T_a$. First, consider the second subsample $X_b(\pi)$. When $\pi < \pi_0$, following the model in (6.2), the variance of the factor loadings is

$$\Sigma_{ab}^+ = N^{-1} (\Lambda^0, \Psi^0)' (\Lambda^0, \Psi^0). \quad (\text{D.1})$$

By Assumption B and the continuous mapping theorem, we know that Σ_{ab}^+ has rank r^+ w.p.a.1, which implies that the rank of (Λ^0, Ψ^0) is r^+ w.p.a.1. Thus, there is a $(r_a + r_b) \times (r_a + r_b)$ orthogonal matrix S such that the first r^+ columns of $(\Lambda^0, \Psi^0)S$ have full rank and the last $(r_a + r_b - r_+)$ columns are 0 w.p.a.1. As such, the model in (6.2) can be written as an approximate factor model with r^+ factors, and the factors and their loadings both have full ranks asymptotically. With a transformation analogous to that in (2.10) to standardize the factors and diagonalize the loadings, the DGP in (6.2) can be written as

$$X_b(\pi) = F_b^R(\pi)\Psi^R(\pi)' + e_b(\pi), \quad (\text{D.2})$$

where $F_b^R(\pi)$ is $T_b \times r^+$, $\Psi^R(\pi)$ is $N \times r^+$, and

$$\begin{aligned} T_b^{-1}F_b^R(\pi)'F_b^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1}\Psi^R(\pi)'\Psi^R(\pi) &= \Lambda_b(\pi), \end{aligned} \quad (\text{D.3})$$

where $\Lambda_b(\pi)$ is a $r^+ \times r^+$ diagonal matrix whose diagonal elements are the positive eigenvalues of $\Sigma_F^+(\pi)\Sigma_{ab}^+$ in a decreasing order. This is analogous to the transformation considered in (B.3)-(B.5) in the proof of Lemma 3 except $\pi < \pi_0$ rather than $\pi = \pi_0$. When $\pi \geq \pi_0$, the DGP in (6.2) can be written as in (D.2) and (D.3) but with $r^+ = r_b$ and $\Psi^R(\pi) = \Psi^R$, where $\Psi^R = \Psi^0(R_b^{-1})'$.

Next, we consider the first subsample $X_a(\pi)$. Following the transformation discussed above, when $\pi > \pi_0$, the DGP in (6.1) can be written as

$$X_a(\pi) = F_a^R(\pi)\Lambda^R(\pi)' + e_a(\pi), \quad (\text{D.4})$$

where $F_a^R(\pi)$ is $T_a \times r^+$, $\Lambda^R(\pi)$ is $N \times r^+$, and

$$\begin{aligned} T_a^{-1}F_a^R(\pi)'F_a^R(\pi) &= I_{r^+} + O_{p\pi}(T^{-1/2}), \\ N^{-1}\Lambda^R(\pi)'\Lambda^R(\pi) &= \Lambda_a(\pi), \end{aligned} \quad (\text{D.5})$$

where $\Lambda_a(\pi)$ is a $r^+ \times r^+$ diagonal matrix with positive eigenvalues. When $\pi \leq \pi_0$, the DGP in (6.1) can be written as that in (D.4) and (D.5) but with $r^+ = r_a$ and $\Lambda^R(\pi) = \Lambda^R = \Lambda^0(R_a^{-1})'$.

For any $\pi \in \Pi$, $X_a(\pi)$ contains at least the r_a factors in $X_a(\pi_0)$ and $X_b(\pi)$ contains at least the r_b factors in $X_b(\pi_0)$. Therefore,

$$N^{-1}\|\Lambda_\ell^R(\pi)\|^2 \geq C \text{ for } \ell = 1, \dots, r_a, \quad N^{-1}\|\Psi_\ell^R(\pi)\|^2 \geq C \text{ for } \ell = 1, \dots, r_b. \quad (\text{D.6})$$

Note that in the proof of Theorem 1 above, the magnitudes of the approximation errors are developed under Assumptions A-D. After Assumptions A and C are replaced by Assumptions A* and C*, Assumptions A*, B, C*, and D are all uniform over $\pi \in \Pi$. As a result, replacing π_0 with π , asymptotic results as those in Theorem 1 hold uniformly over $\pi \in \Pi$. We use such uniform convergence in the analysis below.

Below we analyze model selection based on the two-step procedure. Recall that $\widehat{r}_a^{(i)}(\pi)$ for $i = 1$ and 2 denotes the estimator of $r_a(\pi)$ by the first- and second-step PLS estimator. Let $\omega_\ell^{\lambda^{*(i)}}(\pi)$ and $\omega_\ell^{\gamma^{*(i)}}(\pi)$ denote the weights in step i . Let $\widetilde{\Psi}_{LS}^-(\pi)$ denote the first $\widehat{r}_a^{(1)}$ columns of $\widetilde{\Psi}_{LS}(\pi)$. By construction, the adaptive weights in (6.12) satisfy

$$\begin{aligned} \omega_\ell^{\lambda^{*(i)}}(\pi) &= \left(N^{-1} \|\widetilde{\Lambda}_{\ell,LS}(\pi)\|^2 \right)^{-d} \text{ for } i = 1 \text{ and } 2, \\ \omega_\ell^{\gamma^{*(1)}}(\pi) &= \max \left\{ \left(N^{-1} \|\widetilde{\Gamma}_{\ell,LS}(\pi)\|^2 \right)^{-d}, \left(N^{-1} \|\widetilde{\Psi}_{\ell,LS}(\pi)\|^2 \right)^{-d} \right\}, \\ \omega_\ell^{\gamma^{*(2)}}(\pi) &= \omega_\ell^{\gamma^{*(1)}}(\pi) \text{ if (i) } \widehat{r}_a^{(1)} < \widehat{r}_b^{(1)} \text{ or (ii) } \widehat{r}_a^{(1)} = \widehat{r}_b^{(1)} \text{ and } \ell > \widehat{r}_a^{(1)}, \\ \omega_\ell^{\gamma^{*(2)}}(\pi) &= \max \left\{ \left(N^{-1} \|\widetilde{\Psi}_{\ell,LS}^-(\pi)w(\pi) - \widetilde{\Lambda}_{\ell,LS}(\pi)\|^2 \right)^{-d}, \left(N^{-1} \|\widetilde{\Psi}_{\ell,LS}(\pi)\|^2 \right)^{-d} \right\} \text{ otherwise,} \end{aligned} \quad (\text{D.7})$$

where the vector $w(\pi)$ satisfies $\|w(\pi)\| = 1$ and is obtained by the orthogonal transformation to minimize the difference between the first $\widehat{r}_a^{(1)}$ columns of $\widetilde{\Lambda}_{LS}(\pi)$ and $\widetilde{\Psi}_{LS}(\pi)$.

In the proof below, if notations and results are not specified to be the first step or the second step, they apply to both. We typically do not distinguish between them until discussing the penalties.

Step 1. We show

$$\Pr(\min_{\pi \in \Pi} \widehat{r}_a^{(i)}(\pi) \geq r_a) \rightarrow 1 \text{ for } i = 1 \text{ and } 2. \quad (\text{D.8})$$

To this end, it is sufficient to show $N^{-1} \|\widehat{\Lambda}_\ell(\pi) - \Lambda_\ell^R(\pi)\|^2 = o_{p\pi}(1)$ for $\ell = r_a$ in both steps. The proof strategy is different from that in Theorem 1 because here we do not require the convergence of $\widehat{\Lambda}_\ell(\pi)$ to $\Lambda_\ell^R(\pi)$ for $\ell > r_a$. Let $X_{a:b}$ denote a submatrix of X that contains the columns from a to b . For any $\pi \in \Pi$, define

$$\Lambda^\dagger(\pi) = \left(\Lambda_{1:r_a}^R(\pi), \widehat{\Lambda}(\pi)_{r_a+1:k} \right), \Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi), \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi). \quad (\text{D.9})$$

For notational simplicity, define $\Lambda^r(\pi) = \Lambda_{1:r_a}^R(\pi)$. Note that the definition of $\Lambda^\dagger(\pi)$ is different from that of Λ^* used in the proof of Theorem 1 even when $\pi = \pi_0$, because the

former involves the PLS estimator but the latter does not. Define

$$Z_\lambda^2(\pi) = N^{-1} \left\| \widehat{\Lambda}(\pi) - \Lambda^\dagger(\pi) \right\|^2, \quad Z_\psi^2(\pi) = N^{-1} \left\| \widehat{\Psi}(\pi) - \Psi^\dagger(\pi) \right\|^2, \quad Z_\gamma^2(\pi) = N^{-1} \left\| \widehat{\Gamma}(\pi) - \Gamma^\dagger(\pi) \right\|^2. \quad (\text{D.10})$$

The criterion function for the shrinkage estimator can be written as

$$Q(\Lambda, \Gamma; \pi) = M_a(\Lambda, \widetilde{F}_a(\pi)) + M_b(\Psi, \widetilde{F}_b(\pi)) + P_1^*(\Lambda) + P_2^*(\Gamma), \quad (\text{D.11})$$

where $\Psi = \Lambda + \Gamma$,

$$\begin{aligned} M_a(\Lambda, F_a) &= (NT)^{-1} \|X_a(\pi) - F_a \Lambda'\|^2, \text{ and} \\ M_b(\Psi, F_b) &= (NT)^{-1} \|X_b(\pi) - F_b(\Lambda + \Gamma)'\|^2. \end{aligned} \quad (\text{D.12})$$

For notational simplicity, we do not write $M_a(\Lambda, F_a)$ and $M_b(\Psi, F_b)$ indexed by π , although they are by definition. Define

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda*}(\xi)] \text{ and } \phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma*}(\xi)], \quad (\text{D.13})$$

where ξ has a uniform distribution on Π and $\mathbb{E}_\xi[\cdot]$ is taken w.r.t. ξ . As such, $P_1^*(\Lambda) = \sum_{\ell=1}^k \phi_\ell^\lambda \|\Lambda_\ell\|$ and $P_2^*(\Gamma) = \sum_{\ell=1}^k \phi_\ell^\gamma \|\Gamma_\ell\|$.

Because the shrinkage estimators $\widehat{\Lambda}(\pi)$ and $\widehat{\Gamma}(\pi)$ minimize the criterion function $Q(\Lambda, \Gamma; \pi)$, we have $Q(\widehat{\Lambda}(\pi), \widehat{\Gamma}(\pi)) \leq Q(\Lambda^\dagger(\pi), \Gamma^\dagger(\pi))$, i.e.,

$$\begin{aligned} & \left[M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) \right] + \left[M_b(\widehat{\Psi}(\pi), \widetilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \widetilde{F}_b(\pi)) \right] \\ & \leq \left[P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) \right] + \left[P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\widehat{\Gamma}(\pi)) \right], \end{aligned} \quad (\text{D.14})$$

where $\widehat{\Psi}(\pi) = \widehat{\Lambda}(\pi) + \widehat{\Gamma}(\pi)$. We start with the right-hand side of (D.14). Because the last $(k - r_a)$ columns of $\Lambda^\dagger(\pi)$ and $\widehat{\Lambda}(\pi)$ are the same, by the triangle inequality and the Cauchy-Schwarz inequality, we have

$$P_1^*(\Lambda^\dagger(\pi)) - P_1^*(\widehat{\Lambda}(\pi)) = \sum_{\ell=1}^{r_a} \phi_\ell^\lambda \left(|\Lambda_\ell^\dagger(\pi)| - |\widehat{\Lambda}_\ell(\pi)| \right) \leq \bar{b}_\Lambda Z_\lambda(\pi), \text{ where } \bar{b}_\Lambda = N^{1/2} \left(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2 \right)^{1/2}. \quad (\text{D.15})$$

Because $\Gamma^\dagger(\pi) = \widehat{\Gamma}(\pi)$, the second term on the right-hand side of (D.14) is 0.

Next, we consider the left-hand side of (D.14). Write $\widetilde{F}_a(\pi) = (\widetilde{F}_a^r(\pi), \widetilde{F}_a^\perp(\pi)) \in R^{T_a \times k}$, where $\widetilde{F}_a(\pi)$ is partitioned into the $T_a \times r_a$ and $T_a \times (k - r_a)$ submatrices $\widetilde{F}_a^r(\pi)$ and $\widetilde{F}_a^\perp(\pi)$. Similarly, write $\widehat{\Lambda}(\pi) = (\widehat{\Lambda}^r(\pi), \widehat{\Lambda}^\perp(\pi))$, where $\widehat{\Lambda}(\pi)$ is partitioned into the $N \times r_a$ and

$N \times (k - r_a)$ submatrices $\widehat{\Lambda}^r(\pi)$ and $\widehat{\Lambda}^\perp(\pi)$. With this partition, we can write $\Lambda^\dagger(\pi) = (\Lambda^r(\pi), \widehat{\Lambda}^\perp(\pi))$. Define $e_a(\Lambda(\pi), F(\pi)) = X_a(\pi) - F(\pi)\Lambda(\pi)'$. For the calculation below, we first show two expansions. The first is

$$\begin{aligned} e_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) &= X_a(\pi) - \widetilde{F}_a(\pi)\widehat{\Lambda}(\pi)' \\ &= X_a(\pi) - \widetilde{F}_a^r(\pi)\widehat{\Lambda}^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \\ &= \left(X_a(\pi) - \widetilde{F}_a^r(\pi)\Lambda^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \right) - \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)' \\ &= e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) - \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)', \end{aligned} \quad (\text{D.16})$$

where the first and last equalities hold by definition, the second equality follows from the partition of $\widetilde{F}_a(\pi)$ and $\widehat{\Lambda}(\pi)$, and the third equality follows from subtracting and adding $\widetilde{F}_a^r(\pi)\Lambda^r(\pi)'$. Because $r_a(\pi) \geq r_a$, we write $F_a^R(\pi) = (F_a^r(\pi), F_a^{r+}(\pi))$, where $F_a^R(\pi)$ is partitioned into the $T_a \times r_a$ and $T_a \times (r_a(\pi) - r_a)$ submatrices $F_a^r(\pi)$ and $F_a^{r+}(\pi)$. Similarly, write $\Lambda^R(\pi) = (\Lambda^r(\pi), \Lambda^{r+}(\pi))$, where $\Lambda^R(\pi)$ is partitioned into the $N \times r_a$ and $N \times (r_a(\pi) - r_a)$ submatrices $\Lambda^r(\pi)$ and $\Lambda^{r+}(\pi)$. Following the partition, we can write

$$X_a(\pi) = e_a(\pi) + F_a^r(\pi)\Lambda^r(\pi)' + F_a^{r+}(\pi)\Lambda^{r+}(\pi)'. \quad (\text{D.17})$$

The second expansion is

$$\begin{aligned} e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) &= X_a(\pi) - \widetilde{F}_a^r(\pi)\Lambda^r(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)' \\ &= e_a(\pi) + \left(F_a^r(\pi) - \widetilde{F}_a^r(\pi) \right) \Lambda^r(\pi)' + F_a^{r+}(\pi)\Lambda^{r+}(\pi)' - \widetilde{F}_a^\perp(\pi)\widehat{\Lambda}^\perp(\pi)', \end{aligned} \quad (\text{D.18})$$

where first equality holds by definition and the second equality follows from (D.17). With the first expansion in (D.16), we have

$$\begin{aligned} M_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) &= (NT)^{-1} \left\| e_a(\widehat{\Lambda}(\pi), \widetilde{F}_a(\pi)) \right\|^2 \\ &= (NT)^{-1} \left\| e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) \right\|^2 + (NT)^{-1} \left\| \widetilde{F}_a^r \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)' \right\|^2 \\ &\quad - 2(NT)^{-1} \text{tr} \left[e_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi))' \widetilde{F}_a^r(\pi) \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \right] \\ &= M_a(\Lambda^\dagger(\pi), \widetilde{F}_a(\pi)) + K_0 + K_1 + K_2 + K_3 + K_4, \end{aligned} \quad (\text{D.19})$$

where

$$\begin{aligned} K_0 &= (NT)^{-1} \left\| \widetilde{F}_a^r \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)' \right\|^2 \\ &= \frac{T_a}{T} \frac{1}{N} \text{tr} \left[\left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \frac{\widetilde{F}_a^{r'} \widetilde{F}_a^r}{T_a} \left(\widehat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right)' \right] \\ &= \frac{T_a}{T} Z_\lambda^2(\pi) \end{aligned} \quad (\text{D.20})$$

by definition and the fact that $T_a^{-1}(\tilde{F}_a^{r'}\tilde{F}_a^r) = I_{r_a \times r_a}$. The terms K_1 to K_4 follow from the second expansion in (D.18), and they are specified below. The first term is

$$K_1 = -2(NT)^{-1}tr \left[e_a(\pi)' \tilde{F}_a^r(\pi) \left(\hat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \right] = \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \quad (\text{D.21})$$

following calculations analogous to those in (C.20) and (C.21). The second term is

$$\begin{aligned} K_2 &= -2(NT)^{-1}tr \left(\Lambda^r(\pi) (F_a^r(\pi) - \tilde{F}_a^r(\pi))' \tilde{F}_a^r(\pi) (\hat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) \end{aligned} \quad (\text{D.22})$$

following calculations analogous to those in (C.25) and (C.26). The third term is

$$\begin{aligned} K_3 &= -2(NT)^{-1}tr \left(\Lambda^{r+}(\pi) F_a^{r+}(\pi)' \tilde{F}_a^r(\pi) (\hat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= -2(NT)^{-1}tr \left(\Lambda^{r+}(\pi) \left(F_a^{r+}(\pi) - \tilde{F}_a^{r+}(\pi) \right)' \tilde{F}_a^r(\pi) (\hat{\Lambda}^r(\pi) - \Lambda^r(\pi))' \right) \\ &= \frac{T_a}{T} O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi), \end{aligned} \quad (\text{D.23})$$

where $\tilde{F}_a^{r+}(\pi)$ is a submatrix of $\tilde{F}_a(\pi)$ with columns associated with those in $F_a^{r+}(\pi)$, the second equality holds because $\tilde{F}_a^{r+}(\pi)$ and $\tilde{F}_a^r(\pi)$ are orthogonal by construction, and the third equality holds by arguments analogous to those in (C.25) and (C.26). The fourth term is

$$K_4 = 2(NT)^{-1}tr \left[\hat{\Lambda}^\perp(\pi) \tilde{F}_a^\perp(\pi)' \tilde{F}_a^r(\pi) \left(\hat{\Lambda}^r(\pi) - \Lambda^r(\pi) \right) \right] = 0 \quad (\text{D.24})$$

because $\tilde{F}_a^\perp(\pi)' \tilde{F}_a^r(\pi) = 0$ by construction. Combining (D.19)-(D.24), we obtain

$$M_a(\hat{\Lambda}(\pi), \tilde{F}_a(\pi)) - M_a(\Lambda^\dagger(\pi), \tilde{F}_a(\pi)) = \frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi). \quad (\text{D.25})$$

Replacing the first subsample with the second subsample and applying similar arguments, we also have

$$M_b(\hat{\Psi}(\pi), \tilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \tilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \quad (\text{D.26})$$

Plugging (D.25) and (D.26) into the left-hand side of (D.14), we obtain

$$\frac{T_a}{T} Z_\lambda^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\lambda(\pi) + \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi \leq \bar{b}_\Lambda Z_\lambda(\pi), \quad (\text{D.27})$$

which further implies that

$$Z_\lambda(\pi) = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}). \quad (\text{D.28})$$

The unrestricted least square estimator for any $\pi \in \Pi$ can be viewed as a PLS estimator with 0 penalty. Therefore, $N^{-1}|\tilde{\Lambda}_{LS,\ell}(\pi) - \Lambda_\ell^R(\pi)|^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell = 1, \dots, r_a$ by (D.28), which together with (D.6) implies that $N^{-1}|\tilde{\Lambda}_{LS,\ell}(\pi)|^2 \geq C^{-1}$ w.p.a.1. for $\ell = 1, \dots, r_a$. For $i = 1$ and 2, we have $\omega_\ell^{\lambda^*(i)}(\pi) = (N^{-1}|\tilde{\Lambda}_{LS,\ell}(\pi)|^2)^{-d} \leq C^d$ w.p.a.1 for $\ell = 1, \dots, r_a$. Following the specification in (6.11), $\alpha_{NT}(\pi) = \kappa_1(\pi)N^{-1/2}C_{NTa}^{-d-1}$, where $\kappa_1(\pi) \leq \bar{\kappa}_1$. Thus, we have

$$N^{1/2}\phi_\ell^\lambda = N^{1/2}\mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda^*}(\xi)] = O_p(C_{NT}^{-1}) \quad (\text{D.29})$$

for $\ell = 1, \dots, r_a$, which implies

$$\bar{b}_\Lambda = O_p(C_{NT}^{-1}) \quad (\text{D.30})$$

for both the first- and second-step PLS estimation. It follows from (D.28) that $Z_\lambda(\pi) = O_{p\pi}(C_{NT}^{-1})$. This completes the proof of $\Pr(\min_{\pi \in \Pi} \hat{r}_a^{(i)}(\pi) \geq r_a) \rightarrow 1$ for $i = 1, 2$.

Step 2. We show for $i = 1$ and 2,

$$\Pr(\min_{\pi \in \Pi} \hat{r}_b^{(i)}(\pi) \geq r_b) \rightarrow 1 \text{ if } r_b > r_a. \quad (\text{D.31})$$

In this case, $N^{-1}|\Gamma_\ell^R(\pi)|^2 \geq C$ by Assumption R*(ii) and $N^{-1}|\Psi_\ell^R(\pi)|^2 \geq C$ by (D.6) for $\ell = r_b$. To show (D.31), it is sufficient to prove $N^{-1}|\hat{\Gamma}_\ell(\pi) - \Gamma_\ell^R(\pi)|^2 = o_{p\pi}(1)$ for $\ell = r_b$ for both the first and second step estimators. To this end, we redefine $\Lambda^\dagger(\pi)$ and $\Gamma^\dagger(\pi)$ in (D.9) as

$$\Lambda^\dagger(\pi) = \hat{\Lambda}(\pi), \Gamma^\dagger(\pi) = \left(\hat{\Gamma}(\pi)_{1:r_b-1}, \Gamma_{r_b}^R(\pi), \hat{\Gamma}(\pi)_{r_b+1:k} \right) \text{ and } \Psi^\dagger(\pi) = \Lambda^\dagger(\pi) + \Gamma^\dagger(\pi) \quad (\text{D.32})$$

and keep the definitions of $Z_\lambda(\pi)$, $Z_\psi(\pi)$, $Z_\gamma(\pi)$ in (D.10) unchanged. Now consider the inequality in (D.14). Because $\Lambda^\dagger(\pi) = \hat{\Lambda}(\pi)$, the right-hand side of (D.14) becomes for $\ell = r_b$,

$$P_2^*(\Gamma^\dagger(\pi)) - P_2^*(\hat{\Gamma}(\pi)) = \phi_\ell^\gamma \left(|\Gamma_\ell^R(\pi)| - |\hat{\Gamma}_\ell(\pi)| \right) \leq \bar{b}_{\Gamma_b} Z_\gamma(\pi), \text{ where } \bar{b}_{\Gamma_b} = N^{1/2}\phi_\ell^\gamma. \quad (\text{D.33})$$

By arguments analogous to those used to show (D.25) and (D.26), the left-hand side of (D.14) becomes

$$M_b(\hat{\Psi}(\pi), \tilde{F}_b(\pi)) - M_b(\Psi^\dagger(\pi), \tilde{F}_b(\pi)) = \frac{T_b}{T} Z_\psi^2(\pi) + O_{p\pi}(C_{NT}^{-1}) Z_\psi(\pi). \quad (\text{D.34})$$

Putting (D.33) and (D.34) together with (D.14), we get

$$Z_\psi(\pi) = O_{p\pi}(\bar{b}_{\Gamma_b} + C_{NT}^{-1}). \quad (\text{D.35})$$

Note that we can show the consistency of $\widehat{\Lambda}(\pi)$ and $\widehat{\Psi}(\pi)$ column by column because $\widetilde{F}_a(\pi)$ and $\widetilde{F}_b(\pi)$ both have orthogonal regressors by construction. Now following the arguments used to show (D.29), we have $\bar{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the first-step estimator, which immediately implies that $Z_\psi(\pi) = O_{p\pi}(C_{NT}^{-1})$ and

$$N^{-1} \|\widehat{\Gamma}_\ell(\pi) - \Gamma_\ell^R(\pi)\|^2 = O_{p\pi}(C_{NT}^{-2}) \text{ for } \ell = r_b. \quad (\text{D.36})$$

This proof (D.31) holds for $i = 1$ and also implies that $\widehat{r}_b^{(1)} = \min_{\pi \in \Pi} \widehat{r}_b^{(1)}(\pi) \geq r_b > r_a$ w.p.a.1. Thus, for the second-step estimator, $\omega_\ell^{\gamma^*(2)}(\pi)$ takes the form in (D.7) with $r_b > r_a$ w.p.a.1, which is the same as that for the first-step estimator. Hence, $\bar{b}_{\Gamma b} = O_p(C_{NT}^{-1})$ for the second-step estimator and it follows that (D.31) holds for $i = 2$ as well.

Step 3. We prove

$$\Pr(\widehat{r}_a^{(1)} = r_a) \rightarrow 1 \quad (\text{D.37})$$

by showing that the inequalities in (D.8) become equalities when $\pi = \pi_0$. To this end, it is sufficient to show $\Pr(\widehat{\Lambda}_\ell(\pi_0) = 0) \rightarrow 1$ for $\ell > r_a$ in the first-step estimation. (We use generic notation below without superscript (1) for notational simplicity.) By the proof of Theorem 1, to obtain $\Pr(\widehat{\Lambda}_\ell(\pi_0) = 0) \rightarrow 1$, it is sufficient to show

$$\left\| e^a(\widehat{\Lambda}(\pi_0))' \widetilde{F}_{a,\ell}(\pi_0) \right\| + \left\| e^b(\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))' \widetilde{F}_{b,\ell}(\pi_0) \right\| < \frac{NT}{2} \phi_\ell^\lambda, \quad (\text{D.38})$$

which is similar to (C.40). Replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ in the proof of Theorem 1 with $\widehat{\Lambda}(\pi_0)$ and $\widehat{\Gamma}(\pi_0)$, respectively, we have

$$N^{-1/2} \|\widehat{\Lambda}(\pi_0) - \Lambda^*\| = O_p(\bar{b}_\Lambda + \bar{b}_\Gamma + C_{NT}^{-1}) \text{ and } N^{-1/2} \|\widehat{\Gamma}(\pi_0) - \Gamma^*\| = O_p(\bar{b}_\Lambda + \bar{b}_\Gamma + C_{NT}^{-1}), \quad (\text{D.39})$$

where

$$\bar{b}_\Lambda = N^{1/2} \left(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2 \right)^{1/2} \text{ and } \bar{b}_\Gamma = N^{1/2} \left(\sum_{\ell \in \mathcal{Z}} (\phi_\ell^\gamma)^2 \right)^{1/2}. \quad (\text{D.40})$$

We have shown $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ in (D.30) for both the first- and second-step estimators. By similar arguments under Assumption R*(i) and (D.6), we also have $\bar{b}_\Gamma = O_p(C_{NT}^{-1})$ for the first step estimator. Because $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ and $\bar{b}_\Gamma = O_p(C_{NT}^{-1})$,

$$N^{-1/2} \|\widehat{\Lambda}^{(1)}(\pi_0) - \Lambda^*\| = O_p(C_{NT}^{-1}) \text{ and } N^{-1/2} \|\widehat{\Gamma}^{(1)}(\pi_0) - \Gamma^*\| = O_p(C_{NT}^{-1}). \quad (\text{D.41})$$

Following the arguments used to show (C.43) and (C.44), (D.38) holds provided that

$$N^{-1/2} C_{NT}^{-1} = o_p(\phi_\ell^\lambda), \quad (\text{D.42})$$

where $\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda^*(1)}(\xi)]$. Using $\alpha_{NT}(\pi) = \kappa_1(\pi)N^{-1/2}C_{NTa}^{-d-1}$, we have

$$\phi_\ell^\lambda = \mathbb{E}_\xi[\alpha_{NT}(\xi)\omega_\ell^{\lambda^*(1)}(\xi)] \geq \underline{\kappa}_1 N^{-1/2} C_{NT}^{-d-1} \mathbb{E}_\xi[\omega_\ell^{\lambda^*(1)}(\xi)\mathcal{I}_{\{\xi \leq \pi_0\}}], \quad (\text{D.43})$$

where $\underline{\kappa}_1$ is the lower bound of $\kappa_1(\pi)$. For $\pi \leq \pi_0$, $X_a(\pi)$ has r_a factors. Thus, the unrestricted least square estimator $N^{-1}||\tilde{\Lambda}_{LS,\ell}(\pi)||^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell > r_a$, by arguments analogous to (C.55). Therefore,

$$\sup_{\pi \leq \pi_0} \left(\omega_\ell^{\lambda^*(1)}(\pi) \right)^{-1} = \sup_{\pi \leq \pi_0} [N^{-1}||\tilde{\Lambda}_{LS,\ell}(\pi)||^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_a. \quad (\text{D.44})$$

Thus, for $\ell > r_a$,

$$\begin{aligned} N^{-1/2} C_{NT}^{-1} (\phi_\ell^\lambda)^{-1} &\leq \underline{\kappa}_1^{-1} C_{NT}^d \left(\mathbb{E}_\xi[\omega_\ell^{\lambda^*(1)}(\xi)\mathcal{I}_{\{\xi \leq \pi_0\}}] \right)^{-1} \\ &\leq \underline{\kappa}_1^{-1} C_{NT}^d \left(\inf_{\pi \leq \pi_0} [\omega_\ell^{\lambda^*(1)}(\pi)] \mathbb{E}_\xi[\mathcal{I}_{\{\xi \leq \pi_0\}}] \right)^{-1} \\ &= \frac{C_{NT}^d \sup_{\pi \leq \pi_0} \left(\omega_\ell^{\lambda^*(1)}(\pi) \right)^{-1}}{\underline{\kappa}_1 \mathbb{E}_\xi[\mathcal{I}_{\{\xi \leq \pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \quad (\text{D.45})$$

where the last equality is by (D.44) and $\underline{\kappa}_1 \mathbb{E}_\xi[\mathcal{I}_{\{\xi \leq \pi_0\}}] > C > 0$ for some fixed constant C . It follows that $\Pr(\widehat{\Lambda}_\ell^{(1)}(\pi_0) = 0) \rightarrow 1$ for $\ell > r_a$, which implies that

$$\Pr(\widehat{r}_a^{(1)}(\pi_0) \leq r_a) \rightarrow 1. \quad (\text{D.46})$$

Combining (D.8) with the result above, we obtain $\Pr(\min_{\pi \in \Pi} \widehat{r}_a^{(1)}(\pi) = \widehat{r}_a^{(1)}(\pi_0) = r_a) \rightarrow 1$. This proves (D.37).

Step 4. We prove

$$\Pr(\widehat{r}_b^{(1)} = r_b) \rightarrow 1 \quad (\text{D.47})$$

by showing that the inequalities in (D.31) become equalities when $\pi = \pi_0$. To this end, it is sufficient to show $\Pr(\widehat{\Gamma}_\ell^{(1)}(\pi_0) = 0) \rightarrow 1$ for $\ell > r_b$. (We use generic notation below without superscript (1) for notational simplicity.) By the proof of Theorem 1, to obtain $\Pr(\widehat{\Gamma}_\ell(\pi_0) = 0) \rightarrow 1$, it is sufficient to show

$$\left\| e^b (\widehat{\Lambda}(\pi_0) + \widehat{\Gamma}(\pi_0))' \tilde{F}_{b,\ell}(\pi_0) \right\| < \frac{NT}{2} \phi_\ell^\gamma. \quad (\text{D.48})$$

To this end, it is sufficient to show $N^{-1/2} C_{NT}^{-1} = o_p(\phi_\ell^\gamma)$. Using $\beta_{NT}(\pi) = \kappa_2(\pi)N^{-1/2}C_{NTb}^{-d-1}$, we have

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi)\omega_\ell^{\gamma^*(1)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2} C_{NT}^{-d-1} \mathbb{E}_\xi[\omega_\ell^{\gamma^*(1)}(\xi)\mathcal{I}_{\{\xi \geq \pi_0\}}], \quad (\text{D.49})$$

where $\underline{\kappa}_2$ is the lower bound of $\kappa_2(\pi)$. For $\pi \geq \pi_0$, $X_b(\pi)$ has r_b factors, thus $N^{-1}|\|\tilde{\Psi}_{LS,\ell}(\pi)\|^2 = O_{p\pi}(C_{NT}^{-2})$ for $\ell > r_b$ by arguments analogous to (C.55). Therefore,

$$\sup_{\pi > \pi_0} \left(\omega_\ell^{\gamma^*(1)}(\pi) \right)^{-1} \leq \sup_{\pi > \pi_0} [N^{-1}|\|\tilde{\Psi}_{LS,\ell}(\pi)\|^2]^d = O_p(C_{NT}^{-2d}) \text{ for } \ell > r_b. \quad (\text{D.50})$$

Thus, for $\ell > r_b$,

$$\begin{aligned} N^{-1/2} C_{NT}^{-1} (\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\mathbb{E}_\xi [\omega_\ell^{\gamma^*(1)}(\xi) \mathcal{I}_{\{\xi \geq \pi_0\}}] \right)^{-1} \\ &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\inf_{\pi > \pi_0} \left(\omega_\ell^{\gamma^*(1)}(\pi) \right) \mathbb{E}_\xi [\mathcal{I}_{\{\xi \geq \pi_0\}}] \right)^{-1} \\ &= \frac{C_{NT}^d \sup_{\pi > \pi_0} \left(\omega_\ell^{\gamma^*(1)}(\pi) \right)^{-1}}{\underline{\kappa}_2 \mathbb{E}_\xi [\mathcal{I}_{\{\xi \geq \pi_0\}}]} = O_p(C_{NT}^{-d}), \end{aligned} \quad (\text{D.51})$$

following from (D.50) and $\underline{\kappa}_2 \mathbb{E}_\xi [\mathcal{I}_{\{\xi \geq \pi_0\}}] > C > 0$ for some fixed constant C . It follows that $\Pr(\hat{\Gamma}_\ell^{(1)}(\pi_0) = 0) \rightarrow 1$ for $\ell > r_b$, which implies

$$\Pr(\hat{r}_b^{(1)}(\pi_0) \leq r_b) \rightarrow 1. \quad (\text{D.52})$$

When $r_b > r_a$, (D.31) and (D.52) imply that

$$\Pr(\hat{r}_b^{(1)} = \min_{\pi \in \Pi} \hat{r}_b^{(1)}(\pi) = r_b) \rightarrow 1. \quad (\text{D.53})$$

On the other hand, if $r_b = r_a$, we can use (D.52) to deduce that

$$\Pr(\min_{\pi \in \Pi} \hat{r}_b^{(1)}(\pi) \leq r_a) \rightarrow 1, \quad (\text{D.54})$$

which together with the definition of $\hat{r}_b^{(1)}$ and (D.37) implies that

$$\Pr(\hat{r}_b^{(1)} = \hat{r}_a^{(1)} = r_b) \rightarrow 1. \quad (\text{D.55})$$

This completes the proof of Step 4.

Step 5. We show

$$\Pr(\hat{r}_a^{(2)} = r_a) \rightarrow 1 \text{ and } \Pr(\hat{r}_b^{(2)} = r_b) \rightarrow 1. \quad (\text{D.56})$$

Following Steps 3 and 4, we know that the event $\{\hat{r}_a^{(1)} = r_a \text{ and } \hat{r}_b^{(1)} = r_b\}$ has probability approaching 1. If $r_b > r_a$, $\omega_\ell^{\lambda^*(i)}$ and $\omega_\ell^{\gamma^*(i)}$ are the same for $i = 1, 2$ following (D.7). Hence, all arguments in Steps 3 and 4 apply to the second-step estimator, which completes the proof immediately.

Next, we consider $r_a = r_b$. Conditioning on the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(1)} = r_b\}$, the proofs in Step 3 and Step 4 apply to the second-step estimator as well, and this gives the desired results.

Step 6. We show that when there is a type-1 change,

$$\Pr(\widehat{\Gamma}^{(2)}(\pi_0) \neq 0) \rightarrow 1. \quad (\text{D.57})$$

To this end, it is sufficient to show $N^{-1} \|\widehat{\Gamma}_\ell^{(2)}(\pi_0) - \Gamma_\ell^R(\pi_0)\|^2 \rightarrow_p 0$ for some $\ell \in \mathcal{Z}$. This follows from (D.39) for the second-step estimator, which holds by the same arguments as in Step 3 conditioning on the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(1)} = r_b\}$. Following Steps 3 and 4, this event occurs w.p.a.1.

The result in (D.57) and Step 5 together imply that $\Pr(\widehat{\mathcal{S}} = 1) \rightarrow 1$ if $\mathcal{S}_0 = 1$.

Step 7. When there is no structural instability, i.e., $\Gamma^0 = 0$, we show

$$\Pr(\sup_{\pi \in \Pi} \|\widehat{\Gamma}^{(2)}(\pi)\| = 0) \rightarrow 1. \quad (\text{D.58})$$

Replacing $\widehat{\Lambda}$ and $\widehat{\Gamma}$ in the proof of Theorem 1 with $\widehat{\Lambda}^{(2)}(\pi)$ and $\widehat{\Gamma}^{(2)}(\pi)$, we have uniform consistency

$$N^{-1/2} \|\widehat{\Lambda}^{(2)}(\pi) - \Lambda^*\| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}) \text{ and } N^{-1/2} \|\widehat{\Gamma}^{(2)}(\pi) - \Gamma^*\| = O_{p\pi}(\bar{b}_\Lambda + C_{NT}^{-1}), \quad (\text{D.59})$$

where $\bar{b}_\Lambda = N^{1/2}(\sum_{\ell=1}^{r_a} (\phi_\ell^\lambda)^2)^{1/2}$. We have shown $\bar{b}_\Lambda = O_p(C_{NT}^{-1})$ in (D.30). Revoking the proof of Theorem 1 with π_0 replaced by π , a sufficient condition for (D.58) is

$$N^{-1/2} C_{NT}^{-1} = o_p(\phi_\ell^\gamma) \text{ for } \ell = 1, \dots, k, \quad (\text{D.60})$$

where the left-hand side follows from uniform convergence rate of the criterion function and the right-hand side is based on the averaging penalty. Following Steps 3 and 4, we know that the event $\{\widehat{r}_a^{(1)} = r_a \text{ and } \widehat{r}_b^{(2)} = r_b\}$ has probability approaching 1. Using $\beta_{NT}(\pi) = \kappa_2(\pi) N^{-1/2} C_{NT_b}^{-d-1}$, we have

$$\phi_\ell^\gamma = \mathbb{E}_\xi[\beta_{NT}(\xi) \omega_\ell^{\gamma^*(2)}(\xi)] \geq \underline{\kappa}_2 N^{-1/2} C_{NT}^{-d-1} \mathbb{E}_\xi[\omega_\ell^{\gamma^*(2)}(\xi)]. \quad (\text{D.61})$$

Using the formula of $\omega_\ell^{\gamma^*(2)}(\pi)$ in (D.7), for $\ell > r_a$,

$$\left(\omega_\ell^{\gamma^*(2)}(\pi)\right)^{-1} = \left(\omega_\ell^{\gamma^*(1)}(\pi)\right)^{-1} \leq \left(N^{-1} \|\widetilde{\Gamma}_{\ell, LS}(\pi)\|^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \quad (\text{D.62})$$

w.p.a.1, where the last equality holds by arguments analogous to (C.55). On the other hand, for $\ell \leq r_a$,

$$\left(\omega_\ell^{\gamma^{*(2)}}(\pi)\right)^{-1} \leq \left(N^{-1} \|\tilde{\Psi}_{\ell,LS}^-(\pi) \mathbf{w}(\pi) - \tilde{\Lambda}_{\ell,LS}(\pi)\|^2\right)^d = O_{p\pi}(C_{NT}^{-2d}) \quad (\text{D.63})$$

w.p.a.1, where the equality follows from arguments analogous to (C.63) under Assumption R*(i). Combining the results in (D.62) and (D.63), we deduce that

$$\sup_{\pi \in \Pi} \left(\omega_\ell^{\gamma^{*(2)}}(\pi)\right)^{-1} = O_p(C_{NT}^{-2d}) \text{ for } \ell = 1, \dots, k. \quad (\text{D.64})$$

Thus, for $\ell = 1, \dots, k$,

$$\begin{aligned} N^{-1/2} C_{NT}^{-1} (\phi_\ell^\gamma)^{-1} &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\mathbb{E}_\xi[\omega_\ell^{\gamma^{*(2)}}(\xi)]\right)^{-1} \\ &\leq \underline{\kappa}_2^{-1} C_{NT}^d \left(\inf_{\pi \in \Pi} \left(\omega_\ell^{\gamma^{*(2)}}(\pi)\right)\right)^{-1} \\ &= \underline{\kappa}_2^{-1} C_{NT}^d \sup_{\pi \in \Pi} \left(\omega_\ell^{\gamma^{*(2)}}(\pi)\right)^{-1} = O_p(C_{NT}^{-d}), \end{aligned} \quad (\text{D.65})$$

following from (D.61) and (D.64). The condition in (D.60) follows from (D.65), and it is sufficient for the desired result. Therefore, if $\mathcal{S}_0 = 0$, we have $\Pr(\widehat{\mathcal{S}}_0 = 0) \rightarrow 1$. This completes the proof. \square