Supplemental Appendix to
"A Pairwise Differencing Distribution Regression Approach for Network Models"

Gabriela M. Miyazato Szini[†]

August 5, 2026

This supplement contains additional simulation results (Appendix S.A) and additional empirical results (Appendix S.B) for "A Pairwise Differencing Distribution Regression Approach for Network Models." Numbered results referenced here without an "S" prefix refer to the main paper.

Appendix S.A Simulations

Appendix S.A.1 Monte Carlo Simulations for a Single Threshold

Building on the data generating processes (DGP) of Jochmans (2018), I consider additional specifications to better understand the implied growth rates of fixed effects that satisfy Assumption 3.4 and their connection to network sparsity. The outcome for a network with n nodes is generally generated as

$$y_{ij} = 1\{x_{ij}\theta_0 + \alpha_i + \gamma_j - \epsilon_{ij} \geq 0\},$$

where $\varepsilon_{ij} \sim$ i.i.d. Logistic$(0,1)$ across the $n(n-1)$ dyads. Following Section 3, the fixed effects take the general form $\alpha_i = a_{n,i}g(n)$, and $\gamma_j = b_{n,j}g(n)$, where $a_{n,i}, b_{n,j} \in [0,1]$ captures cross-node heterogeneity and $g(n)$ controls the growth rate of the fixed effects with n. The distribution of $(a_{n,i}, b_{n,j})$ and the rate $g(n)$ together determine the degree of network sparsity. The DGPs below differ in the specification of $(a_{n,i}, b_{n,j})$.

I set $\theta_0 = -1$ throughout, to ensure that covariates and fixed effects operate in the same direction, with the single regressor generated as $x_{ij} = -|u_i - u_j|$ where $u_i = \nu_i - 1/2$ for $\nu_i \sim Beta(2,2)$. Positive values of $g(n)$ push the fixed effects toward $+\infty$, generating

[†]Tilburg University, e-mail: g.m.miyazatoszini@tilburguniversity.edu

networks where most pairs are linked and sparsity arises through few non-links. As $g(n)$ grows, both p_n and $1 - q_n$ shrink to zero, reducing the share of informative quadruples available for estimation. This is the empirically relevant case for the trade application in Section 6: bilateral trade is bounded below at zero with a mass point at zero, so the estimable thresholds lie above the bound where $\Pr(y_{ij} \leq y)$ rises toward one, and sparsity arises only in the right tail. Jochmans (2018) studies the left-tail case (few links); the right-tail analysis is new to this paper.

Sample sizes are $n \in \{25, 50, 100, 150\}$, with 1000 Monte Carlo replications. The values of $g(n)$ range from 0 (the dense case) through $\log\log n$, $\sqrt{\log n}$, $\log n$, $2\log n$, $(\log n)^{3/2}$, and $(\log n)^2$, in order of increasing sparsity. Tables S.1, S.2 and S.3 report the complete set of results.

DGP 1: Uniform. Following Jochmans (2018), the fixed effects are a deterministic function of the sample size:

$$\alpha_i = \frac{n-i}{n-1} g(n), \quad \gamma_i = \alpha_i,$$

yielding a sequence $a_{n,i}$ uniformly spaced between $[0, 1]$. For large $g(n)$, the condition in Assumption 3.4 translates into the requirement that $g(n) = o(n^{1/4})$. Since all the growth rates $g(n)$ considered are logarithmic functions of n, they all satisfy Assumption 3.4, as any power of $\log n$ grows slower than $n^{1/4}$. However, the speed at which np_n grows with n varies substantially across specifications: for slower-growing $g(n)$ such as $\log\log n$ or $\sqrt{\log n}$, the growth of np_n is apparent at moderate sample sizes, while for faster-growing $g(n)$ such as $(\log n)^2$, it may require sample sizes far beyond those considered here.

DGP 2: Beta heterogeneity. The main change in this specification is regarding the sequence of fixed effects. I consider a deterministic sequence of fixed effects, that is normalized to be in the range $[0, 1]$, using a Beta quantile function such that

$$a_{n,i} = F^{-1}_{Beta}\left(\frac{n-i}{n-1}\right),$$

where F^{-1}_{Beta} is the inverse CDF of $Beta(\alpha, \beta)$.

For $\alpha = \beta < 1$, this specification produces a U-shaped distribution of $a_{n,i}$, concentrating mass near zero and one. Compared to the uniform distribution above, a higher fraction of nodes has $a_{n,i} \approx 0$ and thus maintains moderate link probabilities even for large $|g(n)|$, generating more informative quadruples. At the same time, also a higher fraction of nodes has $a_{n,i} \approx 1$ relative to the uniform case. This greater effective heterogeneity allows for higher magnitudes of $|g(n)|$ than under uniform spacing, for a given general level of sparsity.

More specifically, for large $g(n)$ and $\alpha = \beta = 0.5$, the condition in Assumption 3.4 translates into $g(n) = o(n^{1/2})$, which is less restrictive than in the uniform case. As in the uniform case, all logarithmic growth rates satisfy Assumption 3.4 asymptotically, as any power of $\log n$ grows slower than $n^{1/2}$. However, the faster growth of np_n under Beta heterogeneity means that it becomes empirically visible at smaller sample sizes and to $g(n)$ specifications that grow more rapidly. The sparsity level accommodated by the CMLE, as measured by p_n, is ultimately the same across both specifications; what differs is the magnitude of $g(n)$ required to reach a given level of sparsity.

Results. Table S.1 reports network statistics. The DGP predictions are reflected in finite-sample behavior: under uniform heterogeneity, growth in np_n is visible through $g(n) = \sqrt{\log n}$, becomes modest at $g(n) = \log n$ (from 0.12 to 0.15), and is effectively undetectable beyond. Under Beta heterogeneity, growth extends through $g(n) = (\log n)^{3/2}$ (from 0.03 to 0.05), with stagnation only at $g(n) = (\log n)^2$. Although Assumption 3.4 is satisfied in all cases asymptotically, the simulations illustrate that the practical feasibility of estimation depends on whether np_n grows with n at the sample sizes available.

Notice that as $g(n)$ increases, the theoretical and empirical np_n move closer together. The theoretical p_n in the table is computed using the tail approximation derived in Appendix D, which is valid when fixed effects tend to $\pm\infty$. In the dense case ($g(n) = 0$), fixed effects are at zero and this approximation is outside its regime of validity, producing $p_n = 1$ while only about 12% of quadruples are empirically informative. As $g(n)$ grows and the fixed effects move into the regime where the approximation applies, the theoretical and empirical measures align, both shrinking toward zero. This is also the regime where the rate restriction $\sqrt{n}\,(1-q_n) \to \infty$ becomes operative, with $1-q_n$ visibly shrinking under both heterogeneity patterns as $g(n)$ grows.

Table S.1: Network Statistics — Jochmans DGP, Right Tail ($\theta_0 = -1$)

		Uniform Heterogeneity						Beta Heterogeneity					
		Network Characteristics				np_n		Network Characteristics				np_n	
n	$g(n)$	% Quad	p_n	% Links	$1-q_n$	Theor.	Empir.	% Quad	p_n	% Links	$1-q_n$	Theor.	Empir.
25	0	12.02	1.0000	56.40	1.0000	25.00	3.01	12.02	1.0000	56.40	1.0000	25.00	3.01
50		12.04	1.0000	56.36	1.0000	50.00	6.02	12.04	1.0000	56.36	1.0000	50.00	6.02
100		12.04	1.0000	56.35	1.0000	100.00	12.04	12.04	1.0000	56.35	1.0000	100.00	12.04
150		12.04	1.0000	56.35	1.0000	150.00	18.06	12.04	1.0000	56.35	1.0000	150.00	18.06
25	$\log\log n$	4.90	0.1232	79.45	0.3509	3.08	1.23	4.90	0.1366	78.97	0.3696	3.42	1.22
50		3.95	0.0898	82.01	0.2996	4.49	1.97	3.98	0.1036	81.39	0.3219	5.18	1.99
100		3.29	0.0696	83.84	0.2638	6.96	3.29	3.36	0.0833	83.08	0.2885	8.33	3.36
150		2.99	0.0612	84.73	0.2474	9.18	4.48	3.07	0.0747	83.91	0.2733	11.20	4.60
25	$\sqrt{\log n}$	2.37	0.0486	86.44	0.2206	1.22	0.59	2.46	0.0613	85.57	0.2476	1.53	0.61
50		1.93	0.0369	87.95	0.1922	1.85	0.96	2.06	0.0491	86.92	0.2215	2.45	1.03
100		1.60	0.0291	89.15	0.1705	2.91	1.60	1.76	0.0406	87.99	0.2015	4.06	1.76
150		1.44	0.0256	89.76	0.1601	3.85	2.16	1.60	0.0368	88.56	0.1917	5.51	2.41
25	$\log n$	0.46	0.0089	94.10	0.0946	0.22	0.12	0.61	0.0166	92.66	0.1287	0.41	0.15
50		0.24	0.0043	95.80	0.0654	0.21	0.12	0.39	0.0101	94.19	0.1003	0.50	0.19
100		0.14	0.0023	96.89	0.0475	0.23	0.14	0.27	0.0067	95.23	0.0816	0.67	0.27
150		0.10	0.0016	97.36	0.0401	0.24	0.15	0.22	0.0054	95.70	0.0736	0.81	0.33
25	$2\log n$	0.03	0.0008	98.31	0.0288	0.02	0.01	0.09	0.0036	96.82	0.0596	0.09	0.02
50		0.02	0.0003	98.87	0.0184	0.02	0.01	0.07	0.0022	97.45	0.0464	0.11	0.03
100		0.01	0.0002	99.19	0.0127	0.02	0.01	0.05	0.0014	97.85	0.0380	0.14	0.05
150		0.01	0.0001	99.32	0.0105	0.02	0.01	0.04	0.0012	98.04	0.0344	0.18	0.06
25	$(\log n)^{3/2}$	0.05	0.0012	97.91	0.0348	0.03	0.01	0.12	0.0045	96.40	0.0668	0.11	0.03
50		0.02	0.0004	98.84	0.0187	0.02	0.01	0.07	0.0022	97.42	0.0469	0.11	0.03
100		0.01	0.0001	99.29	0.0111	0.01	0.01	0.04	0.0012	98.01	0.0353	0.12	0.04
150		0.00	0.0001	99.46	0.0085	0.01	0.01	0.03	0.0009	98.26	0.0306	0.14	0.05
25	$(\log n)^2$	0.00	0.0002	99.31	0.0130	0.00	0.00	0.03	0.0014	98.11	0.0372	0.03	0.01
50		0.00	0.0000	99.68	0.0056	0.00	0.00	0.01	0.0006	98.74	0.0237	0.03	0.01
100		0.00	0.0000	99.84	0.0027	0.00	0.00	0.01	0.0003	99.10	0.0163	0.03	0.01
150		0.00	0.0000	99.89	0.0018	0.00	0.00	0.01	0.0002	99.25	0.0135	0.03	0.01

Notes: Based on 1000 Monte Carlo replications. % Quad is the share of informative quadruples; p_n and $1-q_n$ are the theoretical non-link and quadruple density parameters; np_n Theor. uses the population p_n, Empir. uses the simulated share. The dense case ($g(n) = 0$) is identical across heterogeneity specifications.

Table S.2: Estimation Results — Jochmans DGP, Right Tail ($\theta_0 = -1$), Uniform Heterogeneity

		Mean Bias			Median Bias			Size			se/sd		
n	$g(n)$	MLE	BC	CMLE	MLE	BC	CMLE	MLE	BC	CMLE	MLE	BC	CMLE
25	0	−0.098	−0.004	−0.014	−0.103	−0.010	−0.016	0.063	0.037	0.032	0.93	1.02	1.05
50		−0.035	0.008	0.006	−0.035	0.007	0.006	0.051	0.044	0.038	0.99	1.04	1.06
100		−0.019	0.002	0.001	−0.019	0.002	0.001	0.060	0.057	0.053	0.97	0.99	1.00
150		−0.010	0.003	0.003	−0.011	0.003	0.003	0.049	0.045	0.043	1.01	1.02	1.03
25	$\log\log n$	−0.070	0.033	0.018	−0.046	0.052	0.034	0.091	0.050	0.044	0.88	0.98	1.01
50		−0.020	0.027	0.024	−0.005	0.041	0.042	0.059	0.039	0.036	0.97	1.02	1.05
100		−0.025	−0.002	−0.003	−0.023	0.000	−0.000	0.052	0.042	0.040	0.97	1.00	1.01
150		−0.009	0.006	0.006	−0.006	0.009	0.006	0.047	0.042	0.039	1.01	1.03	1.04
25	$\sqrt{\log n}$	−0.057	0.061	0.048	−0.073	0.045	0.048	0.073	0.044	0.041	0.89	1.01	1.03
50		−0.038	0.016	0.012	−0.041	0.013	−0.002	0.045	0.039	0.035	0.98	1.03	1.06
100		−0.028	−0.002	−0.003	−0.014	0.011	0.012	0.053	0.050	0.045	0.99	1.01	1.03
150		−0.009	0.008	0.008	−0.009	0.008	0.006	0.050	0.048	0.046	0.98	0.99	1.00
25	$\log n$	−0.091	0.105	0.061	−0.102	0.088	0.057	0.066	0.026	0.025	0.87	1.04	1.05
50		−0.082	0.011	−0.004	−0.100	−0.007	−0.001	0.064	0.042	0.030	0.92	1.00	1.04
100		−0.056	−0.009	−0.013	−0.059	−0.011	−0.016	0.058	0.049	0.042	0.97	1.01	1.03
150		−0.024	0.008	0.008	−0.029	0.002	0.010	0.057	0.044	0.043	0.97	1.00	1.01
25	$2\log n$	—	—	—	—	—	—	—	—	—	—	—	—
50		−0.183	0.070	0.037	−0.280	−0.053	−0.137	0.062	0.011	0.023	0.73	0.93	0.90
100		−0.093	0.015	−0.003	−0.046	0.058	0.042	0.060	0.032	0.031	0.93	1.03	1.05
150		−0.054	0.016	0.010	−0.076	−0.006	0.008	0.054	0.043	0.032	0.98	1.05	1.07
25	$(\log n)^{3/2}$	—	—	—	—	—	—	—	—	—	—	—	—
50		−0.152	0.090	0.059	−0.254	−0.019	−0.109	0.067	0.014	0.024	0.73	0.92	0.90
100		−0.114	0.006	−0.017	−0.097	0.025	−0.009	0.060	0.036	0.028	0.94	1.06	1.09
150		−0.044	0.036	0.029	−0.072	0.007	0.011	0.054	0.040	0.038	0.95	1.03	1.05
25	$(\log n)^2$	—	—	—	—	—	—	—	—	—	—	—	—
50		—	—	—	—	—	—	—	—	—	—	—	—
100		0.887	1.015	$0.950^{\dagger}$	−0.442	−0.040	−0.153	0.058	0.004	0.015	1.41	1.96	0.22
150		−0.219	0.057	−0.002	−0.220	0.021	−0.077	0.078	0.016	0.026	0.82	1.04	1.03

Notes: True parameter $\theta_0 = -1$; 1000 replications. MLE = maximum likelihood; BC = bias-corrected (Chernozhukov et al. 2024); CMLE = conditional maximum likelihood (pairwise differencing). Size = rejection frequency of the two-sided t-test at the 5% nominal level. se/sd = ratio of average estimated standard error to simulation standard deviation. Entries marked — indicate that a non-negligible share of replications failed to converge due to insufficient variation in the binary outcomes, rendering the row unreliable.

Table S.3: Estimation Results — Jochmans DGP, Right Tail ($\theta_0 = -1$), Beta Heterogeneity

		Mean Bias			Median Bias			Size			se/sd		
n	$g(n)$	MLE	BC	CMLE	MLE	BC	CMLE	MLE	BC	CMLE	MLE	BC	CMLE
25	0	−0.098	−0.004	−0.014	−0.103	−0.010	−0.016	0.063	0.037	0.032	0.93	1.02	1.05
50		−0.035	0.008	0.006	−0.035	0.007	0.006	0.051	0.044	0.038	0.99	1.04	1.06
100		−0.019	0.002	0.001	−0.019	0.002	0.001	0.060	0.057	0.053	0.97	0.99	1.00
150		−0.010	0.003	0.003	−0.011	0.003	0.003	0.049	0.045	0.043	1.01	1.02	1.03
25	$\log\log n$	−0.071	0.033	0.017	−0.048	0.055	0.019	0.078	0.046	0.041	0.89	0.98	1.01
50		−0.022	0.025	0.022	−0.002	0.045	0.040	0.051	0.036	0.035	0.99	1.03	1.06
100		−0.023	0.000	−0.000	−0.024	−0.000	−0.000	0.063	0.050	0.049	0.97	1.00	1.01
150		−0.009	0.007	0.006	−0.007	0.009	0.008	0.048	0.044	0.042	1.01	1.03	1.04
25	$\sqrt{\log n}$	−0.080	0.043	0.024	−0.081	0.040	0.034	0.061	0.035	0.033	0.92	1.03	1.06
50		−0.033	0.022	0.018	−0.044	0.010	0.010	0.052	0.046	0.036	0.98	1.03	1.07
100		−0.025	0.002	0.001	−0.013	0.013	0.010	0.052	0.044	0.042	0.99	1.02	1.04
150		−0.012	0.006	0.006	−0.009	0.008	0.009	0.060	0.055	0.054	0.97	0.99	1.00
25	$\log n$	−0.083	0.101	0.069	−0.073	0.093	0.046	0.062	0.021	0.021	0.88	1.06	1.05
50		−0.083	0.007	−0.003	−0.096	−0.007	−0.014	0.056	0.028	0.027	0.95	1.04	1.07
100		−0.056	−0.010	−0.012	−0.051	−0.005	−0.001	0.065	0.048	0.042	0.96	1.00	1.02
150		−0.029	0.001	0.001	−0.026	0.005	0.010	0.063	0.056	0.047	0.97	1.00	1.02
25†	$2\log n$	−0.039	0.397	0.286	−0.179	0.148	0.109	0.069	0.002	0.016	0.60	0.99	0.85
50		−0.095	0.060	0.044	−0.147	0.014	0.008	0.070	0.027	0.029	0.85	0.99	1.01
100		−0.071	0.002	−0.004	−0.050	0.020	0.013	0.059	0.043	0.036	0.93	1.00	1.02
150		−0.042	0.006	0.005	−0.047	0.000	0.003	0.049	0.036	0.035	1.00	1.05	1.07
25†	$(\log n)^{3/2}$	−0.003	0.353	0.247	−0.151	0.153	0.102	0.066	0.006	0.018	0.68	0.94	0.88
50		−0.096	0.058	0.041	−0.130	0.021	−0.004	0.073	0.029	0.031	0.85	0.99	1.01
100		−0.076	0.001	−0.008	−0.060	0.016	0.014	0.057	0.044	0.037	0.93	1.00	1.03
150		−0.047	0.005	0.003	−0.054	−0.004	−0.016	0.039	0.030	0.028	0.99	1.04	1.07
25	$(\log n)^2$	—	—	—	—	—	—	—	—	—	—	—	—
50		−0.182	0.102	0.047	−0.272	0.017	−0.088	0.074	0.012	0.019	0.49	0.62	0.59
100		−0.120	0.009	−0.019	−0.126	0.001	−0.035	0.066	0.034	0.037	0.90	1.02	1.05
150		−0.026	0.055	0.046	−0.038	0.045	0.027	0.050	0.034	0.028	0.99	1.07	1.10

Notes: True parameter $\theta_0 = -1$; 1000 replications. MLE = maximum likelihood; BC = bias-corrected (Chernozhukov et al. 2024); CMLE = conditional maximum likelihood (pairwise differencing). Size = rejection frequency of the two-sided t-test at the 5% nominal level. se/sd = ratio of average estimated standard error to simulation standard deviation. †Results in this row are based on 999 out of 1000 replications, where all estimators produced finite estimates. Entries marked — indicate that a non-negligible share of replications failed to converge due to insufficient variation in the binary outcomes, rendering the row unreliable.

Tables S.2 and S.3 report estimation results across the same sparsity configurations, for uniform heterogeneity and for Beta. The MLE exhibits bias away from zero across most configurations, as expected with two-way fixed effects. While the bias decreases with n, it remains non-negligible at $n = 150$ in sparser configurations. In the two most extreme sparsity rows at $n = 25$ under Beta heterogeneity, the mean bias is near zero but the median bias remains substantial (-0.179 and -0.151). The MLE's rejection frequencies do not account for this bias, since the estimator is centered at the wrong value.

The bias correction removes most of the incidental-parameter bias at moderate sparsity levels but deteriorates in sparse networks. At $g(n) = \log n$ and $n = 25$, the BC mean bias is 0.105 under uniform heterogeneity and 0.101 under Beta, comparable in magnitude to the MLE bias it is designed to remove. Under Beta heterogeneity at $g(n) = 2 \log n$ and $(\log n)^{3/2}$ with $n = 25$, the BC mean bias increases further to 0.397 and 0.353, while the CMLE remains at 0.286 and 0.247 with smaller median bias throughout. The gap between BC and CMLE widens as sparsity increases, and BC's size becomes extremely conservative (rejection rates near zero) at the most extreme sparsity levels.

The CMLE performs well across configurations where np_n continues to grow with n. Under uniform heterogeneity, it delivers reliable inference through $g(n) = \log n$ (over 97% of pairs linked); under Beta heterogeneity, through $g(n) = (\log n)^{3/2}$ (over 98% of pairs linked). The boundary of reliable CMLE performance aligns with the range where np_n grows visibly with n, established above. Beyond this range, the CMLE remains reasonable at larger n even where np_n growth is slow. Throughout, the CMLE is slightly conservative: rejection frequencies at or below nominal and se/sd ratios above one, indicating mild overestimation of variability. Overall, the two estimators perform comparably at moderate sparsity levels, but BC deteriorates more than the CMLE at the sparse extremes, particularly at small sample sizes.

Appendix S.A.2 Additional Tables and Figures for Subsection 5.1

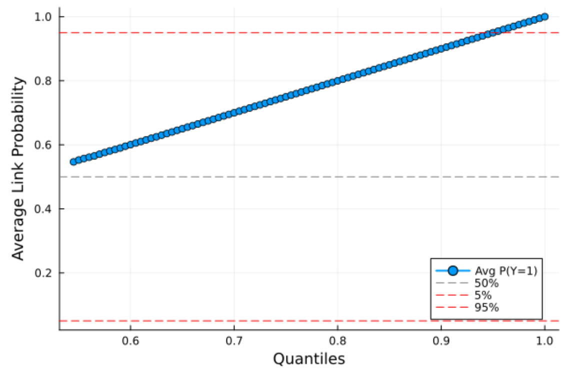
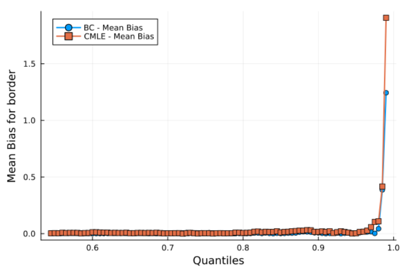
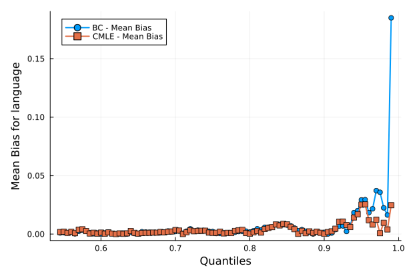


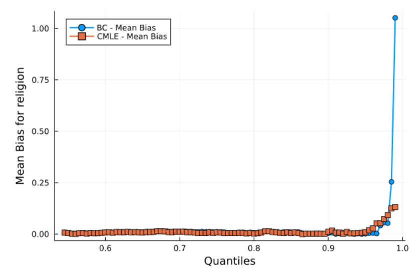
Figure S.1: Average probability $\Pr(\tilde{y}_{ij,k} = 1)$ at each threshold y_k in the empirically calibrated DGP. Dashed lines indicate reference levels at 5%, 50%, and 95%. Thresholds correspond to empirical quantiles in [0.545, 0.990] at intervals of 0.005.



(a) Border

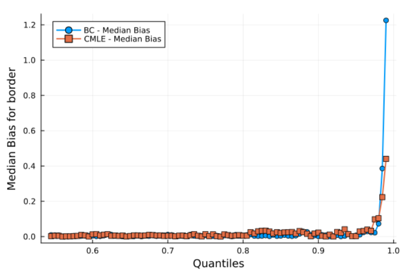


(b) Language

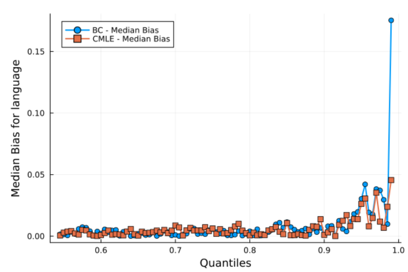


(c) Religion

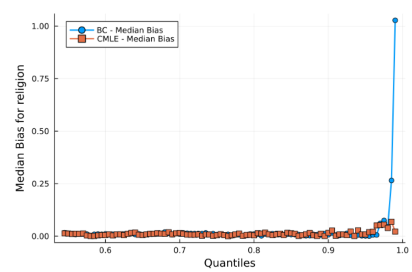
Figure S.2: Mean bias of the bias-corrected estimator (BC) and the conditional maximum likelihood estimator (CMLE) across quantiles of the trade distribution, based on 500 replications using the empirically calibrated DGP. Thresholds correspond to empirical quantiles in [0.545, 0.990] at intervals of 0.005.



(a) Border

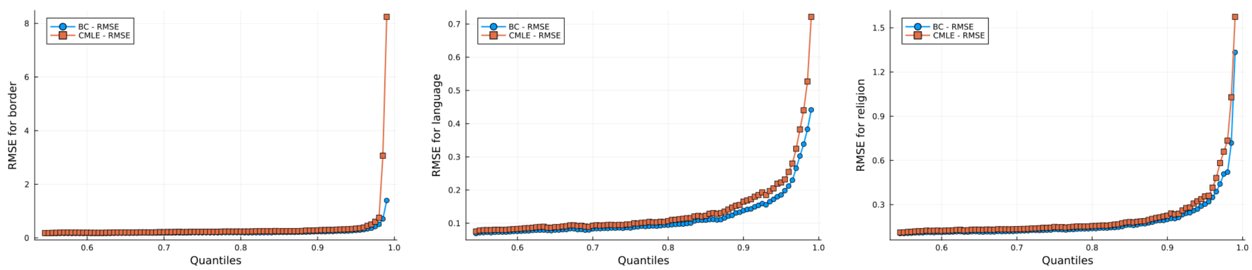


(b) Language



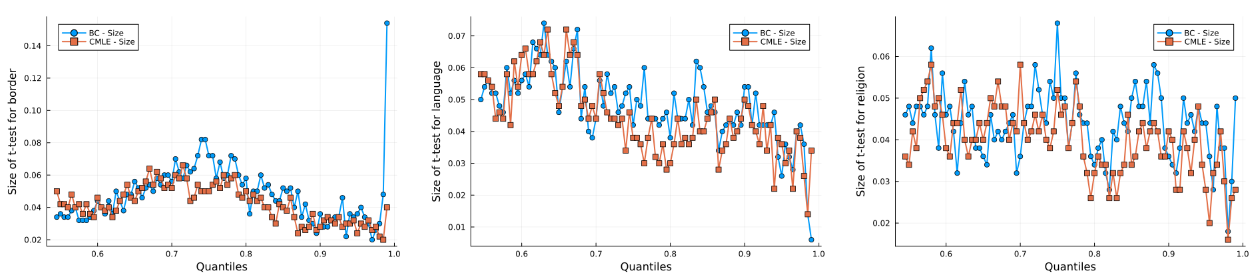
(c) Religion

Figure S.3: Median bias of the bias-corrected estimator (BC) and the conditional maximum likelihood estimator (CMLE) across quantiles of the trade distribution, based on 500 replications using the empirically calibrated DGP. Thresholds correspond to empirical quantiles in [0.545, 0.990] at intervals of 0.005.



(a) Border (b) Language (c) Religion

Figure S.4: RMSE for the bias-corrected estimator (BC) and the conditional maximum likelihood estimator (CMLE) across quantiles of the trade distribution, based on 500 replications using the empirically calibrated DGP. Thresholds correspond to empirical quantiles in [0.545, 0.990] at intervals of 0.005.



(a) Border (b) Language (c) Religion

Figure S.5: Rejection frequency of the two-sided t-test at the 5% nominal level for the bias-corrected estimator (BC) and the conditional maximum likelihood estimator (CMLE) across quantiles of the trade distribution, based on 500 replications using the empirically calibrated DGP. Thresholds correspond to empirical quantiles in [0.545, 0.990] at intervals of 0.005.

Appendix S.A.3 Additional Tables and Figures for Subsection 5.2

Table S.4: Coverage Comparison: Pointwise vs Sup-t Bands ($\tau_{\max} = 0.99$, Equally Spaced)

		Distance		Legal		Border		Language		Religion	
K	DGP	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t
5	Varying	82.80	97.60	85.60	97.00	83.60	93.20	80.80	95.00	83.60	96.40
5	Constant	85.80	97.00	84.80	96.80	82.60	96.40	82.00	95.80	84.80	96.00
15	Varying	66.40	96.20	69.60	96.80	71.20	92.60	64.80	96.00	69.20	95.40
15	Constant	70.60	96.60	70.40	96.40	69.00	96.40	65.60	95.80	67.80	96.60
50	Varying	49.60	96.20	50.40	97.40	56.00	93.00	46.20	95.60	51.40	96.80
50	Constant	52.20	96.00	51.80	97.20	54.80	95.20	47.40	94.80	53.20	97.40

Notes: Empirical coverage rates (in %) of 95% simultaneous confidence bands. "PW" (Pointwise) uses $z_{0.975} = 1.96$ at each threshold; "Sup-t" uses simulated critical values. Target coverage is 95%. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). Based on 500 Monte Carlo simulations.

Table S.5 provides the results for the coverage of the pointwise confidence intervals and the sup-t simultaneous confidence bands for the specifications *first* K and *last* K when considering the maximum quantile to be $\tau_{max} = 0.95$, and Tables S.4 and S.6 provide the results for all the designs considering the maximum quantile to be $\tau_{max} = 0.99$.

Again, while the results for the pointwise confidence intervals show under-coverage, the sup-t bands maintain coverage at or slightly above 95% level across all configurations. Moreover, the pointwise coverage varies substantially across threshold selection schemes. First K tends to achieve higher pointwise coverage, particularly at larger K, while equally spaced and last K show more pronounced under-coverage, reflecting increased estimation uncertainty when thresholds extend into sparser regions of the distribution. In all cases, however, pointwise confidence bands exhibit substantial under-coverage that worsens with K. For the sup-t simultaneous bands, while *last* K shows slightly more conservative coverage compared to *equally spaced* and *first* K, this deviation from the nominal target is small, whereas pointwise under-coverage is far more pronounced.

Importantly, the choice of the maximum quantile τ_{max} has negligible impact on sup-t coverage. At $\tau_{\max} = 0.99$, Border shows mild undercoverage (approximately 93%) in

Table S.5: Coverage Comparison: Pointwise vs Sup-t Bands ($\tau_{\max} = 0.95$, First K and Last K)

			Distance		Legal		Border		Language		Religion	
K	Selection	DGP	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t
5	First K	Varying	90.80	95.20	92.20	96.20	91.80	96.00	90.80	94.60	92.60	96.60
5	First K	Constant	91.60	95.20	92.00	96.00	91.20	96.20	90.60	95.60	93.20	95.80
5	Last K	Varying	90.60	95.60	91.00	96.60	90.80	97.40	89.80	97.20	90.00	97.00
5	Last K	Constant	91.60	95.80	91.60	97.40	91.60	97.80	90.20	97.40	90.40	97.60
15	First K	Varying	85.40	95.40	87.00	95.20	87.00	96.60	82.80	93.80	86.40	95.80
15	First K	Constant	86.60	94.80	87.80	96.20	86.80	97.20	86.60	95.40	88.60	95.60
15	Last K	Varying	80.60	97.40	81.40	97.60	82.80	97.40	80.20	95.60	79.60	97.60
15	Last K	Constant	80.60	96.60	80.60	97.20	82.00	98.40	78.40	96.40	77.40	96.80
50	First K	Varying	66.60	95.60	69.20	96.60	75.80	94.40	67.00	94.60	69.20	96.20
50	First K	Constant	69.40	95.40	73.80	97.00	73.60	95.20	70.40	95.00	73.60	97.40
50	Last K	Varying	61.80	96.00	61.40	96.80	64.40	95.60	59.80	96.80	63.40	96.20
50	Last K	Constant	60.00	96.00	59.20	95.60	62.80	96.60	59.00	97.40	64.20	97.20

Notes: Empirical coverage rates (in %) of 95% simultaneous confidence bands. "PW" (Pointwise) uses $z_{0.975} = 1.96$ at each threshold; "Sup-t" uses simulated critical values. Target coverage is 95%. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). First K selects the lowest K thresholds; Last K selects the highest K thresholds up to $\tau_{\max}$. Based on 500 Monte Carlo simulations.

some configurations, which may reflect higher estimation uncertainty at the most extreme thresholds for this covariate; all other covariates maintain coverage at or above 95%. For pointwise coverage, however, extending to more extreme quantiles has noticeable effects. *First* K results are identical across different maximal thresholds since this scheme always selects the lowest thresholds. *Equally Spaced* shows deterioration at large K, while *last* K shows consistent deterioration across all K values because this scheme selects thresholds closer to the more extreme boundary ($\tau_{\max} = 0.99$). These patterns reflect increased estimation uncertainty at extreme quantiles, which the sup-t simulated critical values accommodate but pointwise inference does not.

Table S.8 provides the results for size and power of the joint equality testing procedures for $\tau_{max} = 0.95$ for the specifications *first* K and *last* K when considering the maximum quantile to be $\tau_{max} = 0.95$, and Tables S.4 and S.6 provide the results for all the designs considering the maximum quantile to be $\tau_{max} = 0.99$.

Table S.6: Coverage Comparison: Pointwise vs Sup-t Bands ($\tau_{\max} = 0.99$, First K and Last K)

			Distance		Legal		Border		Language		Religion	
K	Selection	DGP	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t	PW	Sup-t
5	First K	Varying	90.80	95.20	92.20	96.20	91.80	96.00	90.80	94.60	92.60	96.60
5	First K	Constant	91.60	95.20	92.00	96.00	91.20	96.20	90.60	95.60	93.20	95.80
5	Last K	Varying	90.00	98.20	87.80	98.40	89.00	94.80	88.00	98.40	87.80	97.00
5	Last K	Constant	91.00	98.00	86.00	98.20	92.00	98.40	86.00	98.40	84.80	97.60
15	First K	Varying	85.40	95.40	87.00	95.20	87.00	96.60	82.80	93.80	86.40	95.80
15	First K	Constant	86.60	94.80	87.80	96.20	86.80	97.20	86.60	95.40	88.60	95.60
15	Last K	Varying	76.00	97.80	76.60	98.00	76.80	94.40	73.40	96.60	72.40	97.80
15	Last K	Constant	75.80	98.00	77.40	98.00	72.60	96.80	75.00	98.20	70.80	98.20
50	First K	Varying	66.60	95.60	69.20	96.60	75.80	94.40	67.00	94.60	69.20	96.20
50	First K	Constant	69.40	95.40	73.80	97.00	73.60	95.20	70.40	95.00	73.60	97.40
50	Last K	Varying	57.40	97.40	54.80	97.00	57.80	93.40	52.40	97.60	55.40	95.80
50	Last K	Constant	56.20	96.80	52.40	94.20	55.60	97.60	52.00	97.60	58.00	96.20

Notes: Empirical coverage rates (in %) of 95% simultaneous confidence bands. "PW" (Pointwise) uses $z_{0.975} = 1.96$ at each threshold; "Sup-t" uses simulated critical values. Target coverage is 95%. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). First K selects the lowest K thresholds; Last K selects the highest K thresholds up to $\tau_{\max}$. Based on 500 Monte Carlo simulations.

Overall, the sup-t test maintains valid size across the configurations, with rejection rates typically between 2% and 6% under the null hypothesis for $\tau_{max} = 0.95$. However, when thresholds are concentrated in the upper tail (*last* K), the sup-t test tends to be conservative, with size below the nominal level, particularly at $\tau_{\max} = 0.99$ where the selected thresholds lie in sparser regions. Moreover, power is reduced for covariates whose coefficients vary primarily at lower quantiles. This reflects both the wider confidence bands at extreme thresholds, and the limited heterogeneity across the coefficients in the selected range. However, in comparison, the Wald test shows more substantial size distortion that varies by covariate and worsens with K. For Border, the Wald test over-rejects in all configurations when $K = 50$, with the distortion most severe for *first* K at 22.6%, exceeding the 5% nominal level substantially. The Wald test's poor performance at large K is consistent with its known theoretical limitations when testing many restrictions jointly. Furthermore, note that the choice of $\tau_{\max}$ (0.95 vs 0.99) does not affect size for either test substantially

Table S.7: Size and Power by Covariate: Sup-t vs Wald ($\tau_{\max} = 0.99$, Equally Spaced)

	Distance		Legal		Border		Language		Religion	
K	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald
Panel A: Constant Coefficients (Size, target: 5%)										
5	2.00	2.20	3.40	3.00	2.60	3.00	3.20	3.80	4.20	3.40
15	2.40	3.00	3.80	2.00	3.20	3.80	2.80	2.40	3.60	1.40
50	3.40	2.40	2.40	1.00	4.60	14.20	4.60	2.20	3.40	2.60
Panel B: Varying Coefficients (Power)										
5	99.80	100.00	100.00	100.00	97.60	99.20	26.20	24.40	67.20	77.20
15	100.00	100.00	100.00	100.00	99.00	98.60	58.80	55.40	69.20	88.60
50	100.00	100.00	100.00	100.00	99.60	100.00	53.80	77.80	88.80	93.40

Notes: Rejection rates (in %) for testing $H_0 : \theta_d(\tau_1) = \cdots = \theta_d(\tau_K)$ at the 5% level. Sup-t = Sup-t test; Wald = Wald test. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). Based on 500 Monte Carlo simulations.

under the equally spaced configuration.

Turning to more details on the results for power, the sup-t test exhibits high rejection rates for Distance, Legal, and Border across all specifications for the *equally spaced* design, which is consistent with the larger heterogeneity across the coefficients for these covariates in Figures S.6 and S.7. Language and Religion show more moderate power, reflecting their smaller coefficient variation across thresholds. For Religion, power increases substantially from $K = 5$ to $K = 15$ and remains high at $K = 50$. For Language, power increases from $K = 5$ to $K = 15$ but declines slightly at $K = 50$, suggesting that the additional thresholds introduce estimation noise without capturing further coefficient variation. The choice of $\tau_{\max}$ (0.95 vs 0.99) does not yield a consistent pattern for power; the effect is covariate-specific and depends on where each covariate's coefficients vary most.

Across the different specifications for threshold selection, power depends on which region of the distribution captures the most coefficient variation. For instance, Distance varies mostly at lower quantiles, yielding high power for *first* K but substantially lower power for *last* K. Legal shows the opposite pattern at larger K, with higher power for *last* K where its coefficients vary most (Figures S.8 and S.9). Comparing the two tests, Wald

Table S.8: Size and Power by Covariate: First K and Last K ($\tau_{\max} = 0.95$)

		Distance		Legal		Border		Language		Religion	
K	Selection	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald
Panel A: Constant Coefficients (Size, target: 5%)											
5	First K	3.40	3.60	4.20	4.20	5.20	6.00	5.80	5.40	3.20	3.40
5	Last K	1.80	1.60	2.20	1.60	2.40	1.80	2.40	1.80	2.40	2.80
15	First K	2.80	1.80	3.40	2.80	5.20	8.20	5.60	3.40	2.80	3.00
15	Last K	2.40	2.60	2.40	0.40	1.80	2.60	2.40	2.20	2.40	1.60
50	First K	3.20	1.60	2.00	1.20	5.60	22.60	4.20	3.20	3.60	2.60
50	Last K	3.00	1.80	2.40	0.60	2.60	9.40	2.00	2.00	1.80	1.80
Panel B: Varying Coefficients (Power)											
5	First K	87.60	83.60	14.20	14.40	10.00	4.80	53.80	38.80	6.00	7.20
5	Last K	5.40	4.80	3.60	3.60	16.20	11.00	58.00	45.20	10.80	9.00
15	First K	100.00	97.60	40.80	44.40	20.60	11.60	65.60	45.40	12.80	26.20
15	Last K	5.80	29.60	55.20	84.20	33.80	31.00	51.40	43.00	9.40	21.60
50	First K	100.00	98.60	87.40	95.00	91.40	90.40	64.80	58.00	16.00	82.40
50	Last K	13.60	55.00	99.40	100.00	62.20	90.40	47.00	67.60	48.00	87.60

Notes: Rejection rates (in %) for testing $H_0 : \theta_d(\tau_1) = \cdots = \theta_d(\tau_K)$ at the 5% level. Sup-t = Sup-t test; Wald = Wald test. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). First K selects the lowest K thresholds; Last K selects the highest K thresholds up to $\tau_{\max}$. Based on 500 Monte Carlo simulations.

shows higher power than sup-t in several configurations, particularly for Language and Religion at large K. However, given the Wald test's size distortion for other covariates at the same levels of K and its known theoretical limitations when testing many restrictions jointly, the sup-t test offers more reliable inference overall.

Table S.9: Size and Power by Covariate: First K and Last K ($\tau_{\max} = 0.99$)

		Distance		Legal		Border		Language		Religion	
K	Selection	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald	Sup-t	Wald
Panel A: Constant Coefficients (Size, target: 5%)											
5	First K	3.40	3.60	4.20	4.20	5.20	6.00	5.80	5.40	3.20	3.40
5	Last K	0.40	0.80	2.60	1.80	1.00	1.20	2.20	1.20	3.00	1.60
15	First K	2.80	1.80	3.40	2.80	5.20	8.20	5.60	3.40	2.80	3.00
15	Last K	1.40	0.40	1.80	0.40	1.20	1.40	1.20	0.60	3.80	1.80
50	First K	3.20	1.60	2.00	1.20	5.60	22.60	4.20	3.20	3.60	2.60
50	Last K	1.80	1.40	2.20	1.60	2.20	9.00	2.40	1.00	3.40	1.00
Panel B: Varying Coefficients (Power)											
5	First K	87.60	83.60	14.20	14.40	10.00	4.80	53.80	38.80	6.00	7.20
5	Last K	2.60	9.80	7.40	24.60	12.80	22.00	3.40	2.40	2.80	3.20
15	First K	100.00	97.60	40.80	44.40	20.60	11.60	65.60	45.40	12.80	26.20
15	Last K	7.60	23.60	14.40	20.20	61.20	84.80	3.80	38.20	6.60	15.40
50	First K	100.00	98.60	87.40	95.00	91.40	90.40	64.80	58.00	16.00	82.40
50	Last K	13.20	72.00	67.00	100.00	76.60	98.00	4.00	65.80	15.20	92.00

Notes: Rejection rates (in %) for testing $H_0 : \theta_d(\tau_1) = \cdots = \theta_d(\tau_K)$ at the 5% level. Sup-t = Sup-t test; Wald = Wald test. "Constant" = coefficients identical across thresholds (H_0 true); "Varying" = coefficients follow empirical trade data pattern (H_0 false). First K selects the lowest K thresholds; Last K selects the highest K thresholds up to $\tau_{\max}$. Based on 500 Monte Carlo simulations.

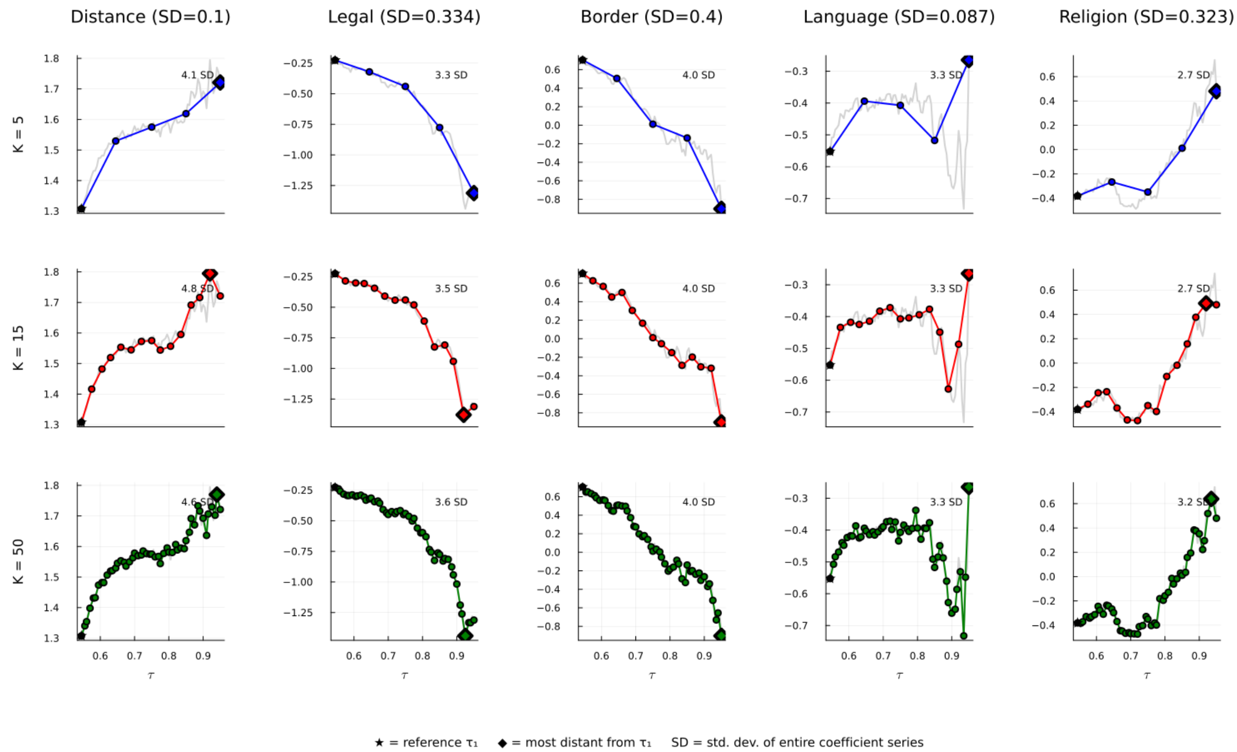


Figure S.6: True coefficient paths from empirical trade data with equally spaced threshold selection ($\tau_{\max} = 0.95$). Gray line: full coefficient series $\{\theta_d(\tau)\}_{\tau \leq \tau_{\max}}$. Colored markers: selected thresholds for $K = 5$ (blue), $K = 15$ (red), $K = 50$ (green). Star (★) marks reference threshold τ_1; diamond (♦) marks most distant coefficient from τ_1. Each panel reports the maximum deviation in standard deviation units, where SD is computed over the entire coefficient series.

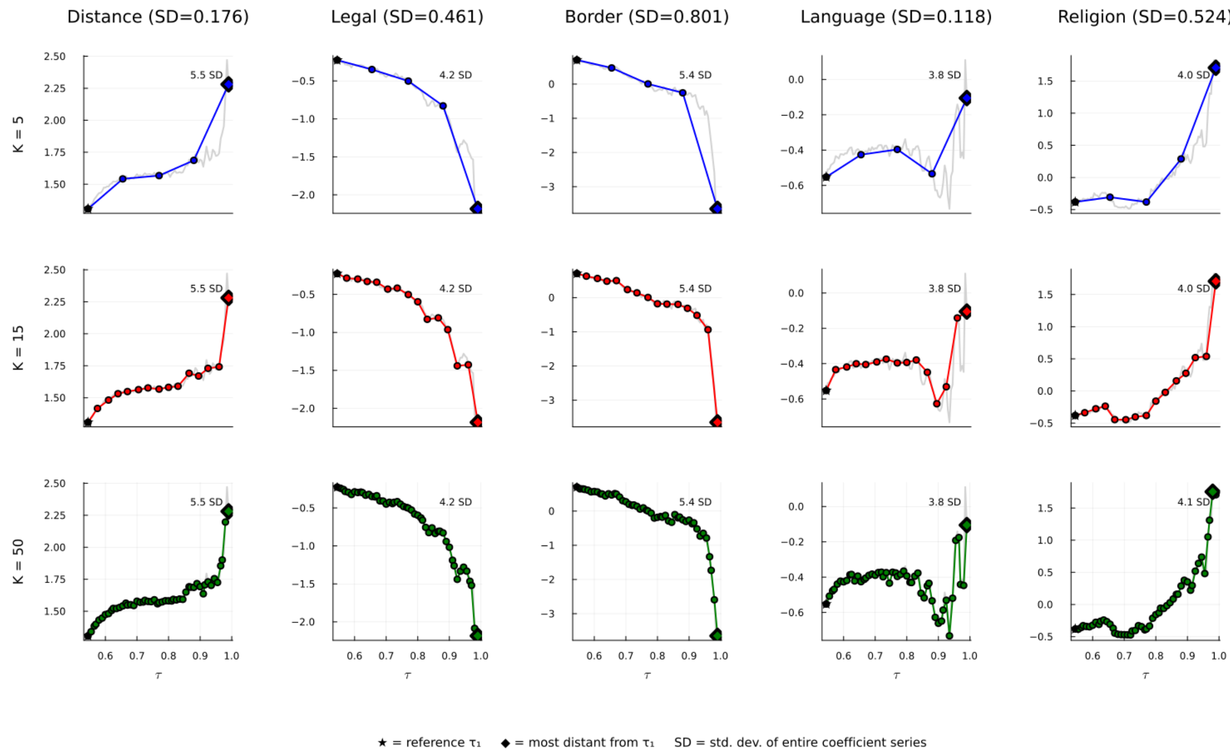


Figure S.7: True coefficient paths from empirical trade data with equally spaced threshold selection ($\tau_{\max} = 0.99$). Gray line: full coefficient series $\{\theta_d(\tau)\}_{\tau \leq \tau_{\max}}$. Colored markers: selected thresholds for $K = 5$ (blue), $K = 15$ (red), $K = 50$ (green). Star (★) marks reference threshold τ_1; diamond (♦) marks most distant coefficient from τ_1. Each panel reports the maximum deviation in standard deviation units, where SD is computed over the entire coefficient series.

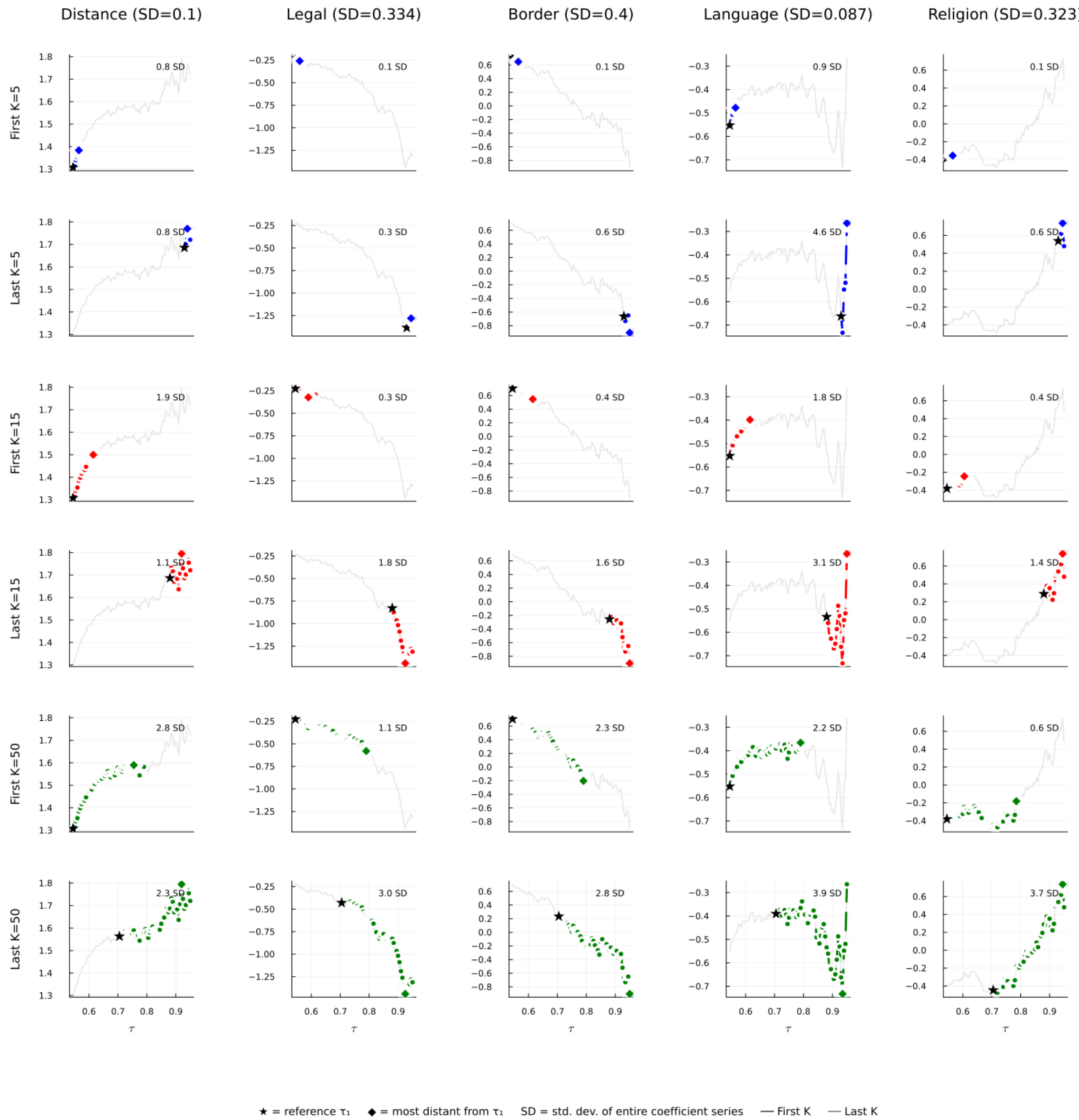


Figure S.8: True coefficient paths from empirical trade data with First K and Last K threshold selection ($\tau_{\max} = 0.95$). Gray line: full coefficient series $\{\theta_d(\tau)\}_{\tau \leq \tau_{\max}}$. First K selects the lowest K thresholds; Last K selects the highest K thresholds. Colored markers indicate $K = 5$ (blue), $K = 15$ (red), $K = 50$ (green). Star (★) marks reference threshold τ_1; diamond (♦) marks most distant coefficient. Lower power in these configurations reflects reduced coefficient variation within the selected threshold ranges compared to equally spaced selection.

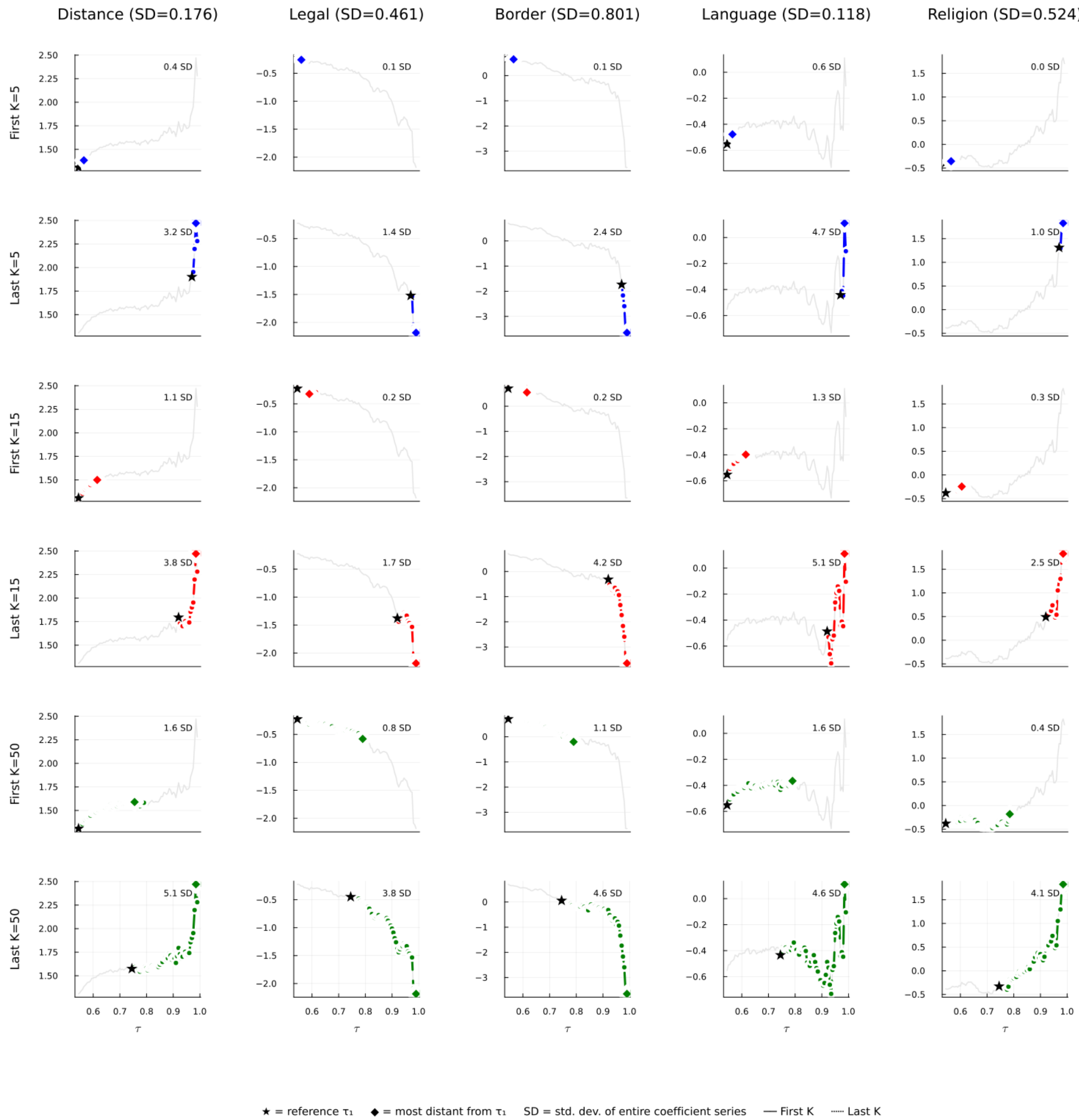
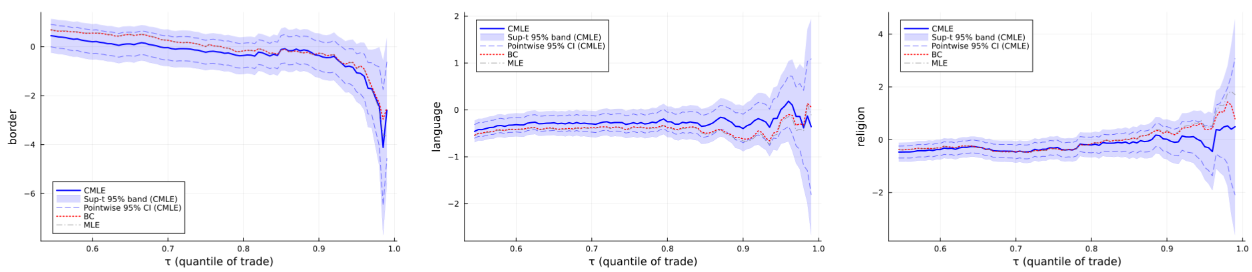


Figure S.9: True coefficient paths from empirical trade data with First K and Last K threshold selection ($\tau_{\max} = 0.99$). Gray line: full coefficient series $\{\theta_d(\tau)\}_{\tau \leq \tau_{\max}}$. First K selects the lowest K thresholds; Last K selects the highest K thresholds. Colored markers indicate $K = 5$ (blue), $K = 15$ (red), $K = 50$ (green). Star (★) marks reference threshold τ_1; diamond (⧫) marks most distant coefficient. Lower power in these configurations reflects reduced coefficient variation within the selected threshold ranges compared to equally spaced selection.

Appendix S.A.4 Additional Tables and Figures for Subsection 6

Table S.10: Descriptive statistics. Source: Helpman et al. (2008).

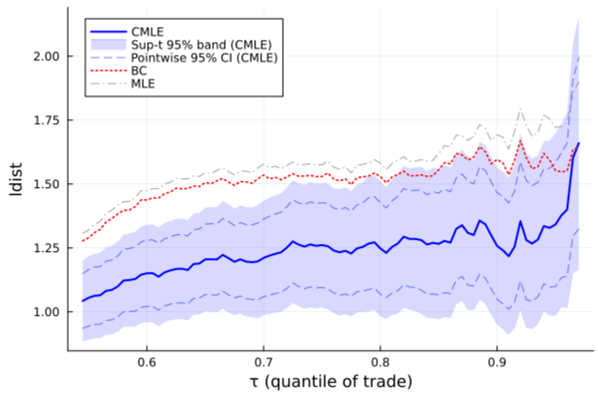
	Mean	Std. Dev.
Trade	0.45	0.50
Trade volume	84.54	1,082,219
Log distance	4.18	0.78
Legal	0.37	0.48
Language	0.29	0.45
Religion	0.17	0.25
Border	0.02	0.13
Currency	0.01	0.09
FTA	0.01	0.08
Colony	0.01	0.10



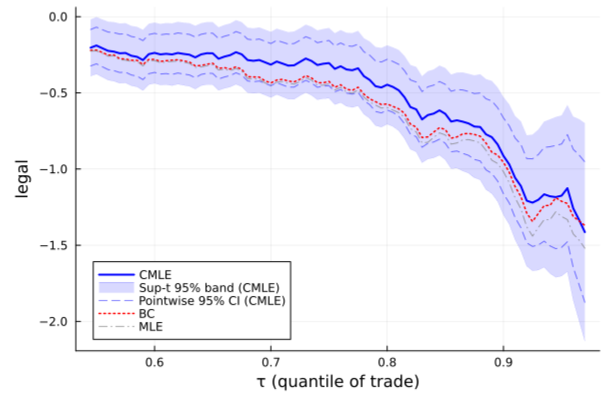
(a) Border (b) Language (c) Religion

Figure S.10: Distribution regression estimates of the effect of remaining covariates on bilateral trade. Solid blue: CMLE; dotted red: BC; dash-dotted gray: MLE. Shaded region and dashed lines show the 95% simultaneous sup-t confidence band and pointwise confidence intervals for the CMLE, respectively. Thresholds correspond to empirical quantiles in $[0.545, 0.990]$ at intervals of 0.005.

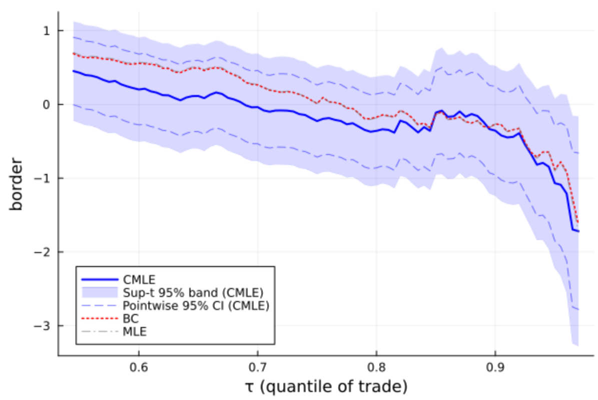
Figure S.12 displays pointwise 95% confidence intervals for all three estimators. The CMLE confidence intervals are wider than those of the BC estimator throughout the distribution, reflecting the efficiency cost of conditioning out the fixed effects rather than estimating and correcting for them. This efficiency loss is the expected trade-off for eliminating the incidental parameter problem entirely. Note that the uncorrected MLE confidence intervals have similar width to those of the BC, since both are based on the same logit variance estimate. All three sets of confidence intervals widen in the upper tail. For the



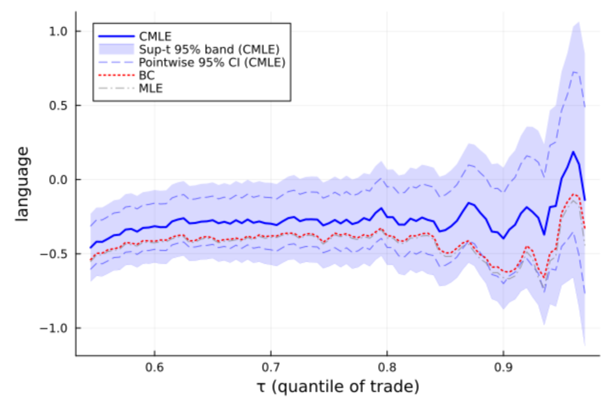
(a) Log distance



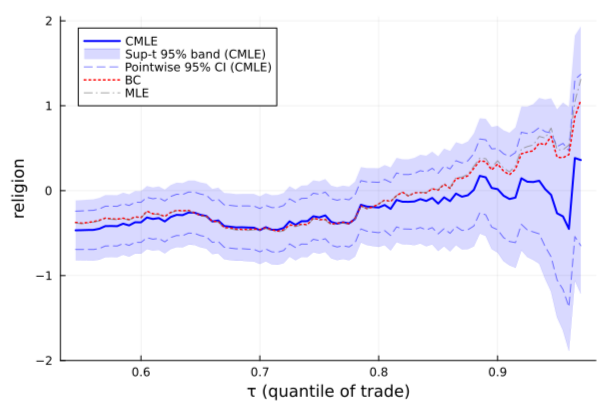
(b) Legal system



(c) Border



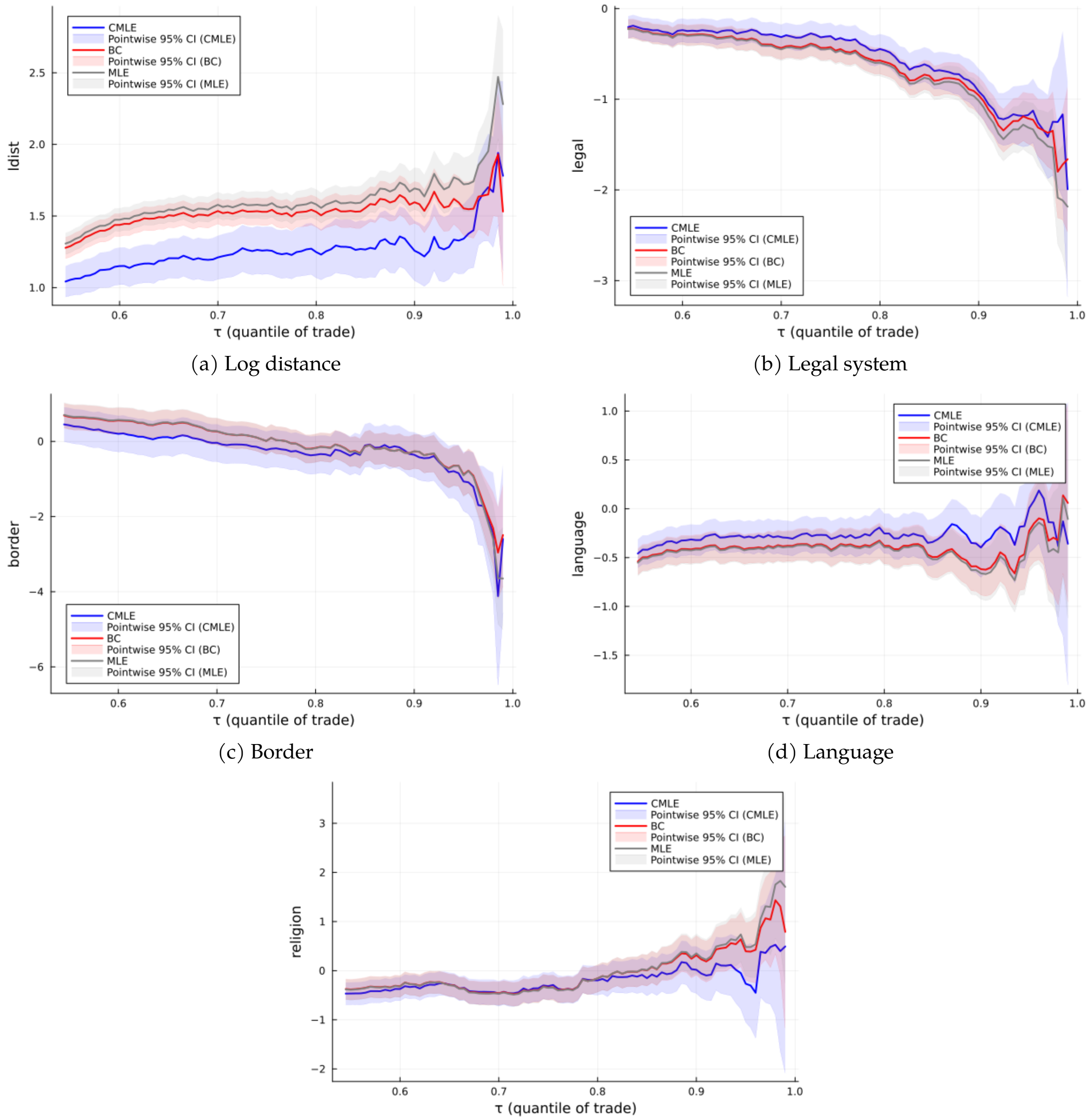
(d) Language



(e) Religion

Figure S.11: Distribution regression estimates of the effect of all covariates on bilateral trade. Solid blue: CMLE; dotted red: BC; dash-dotted gray: MLE. Shaded region and dashed lines show the 95% simultaneous sup-t confidence band and pointwise confidence intervals for the CMLE, respectively. Thresholds correspond to empirical quantiles in $[0.545, 0.970]$ at intervals of 0.005.

CMLE, this reflects the smaller number of informative quadruples at extreme thresholds. For the MLE and BC, which use all observations regardless of the threshold, the widening reflects the reduced information content at extreme thresholds, where predicted proba-



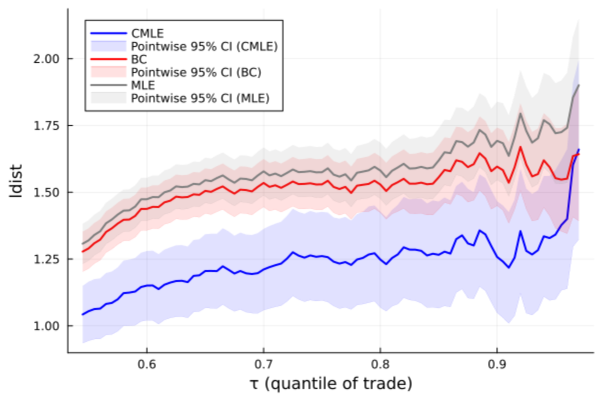
(a) Log distance (b) Legal system

(c) Border (d) Language

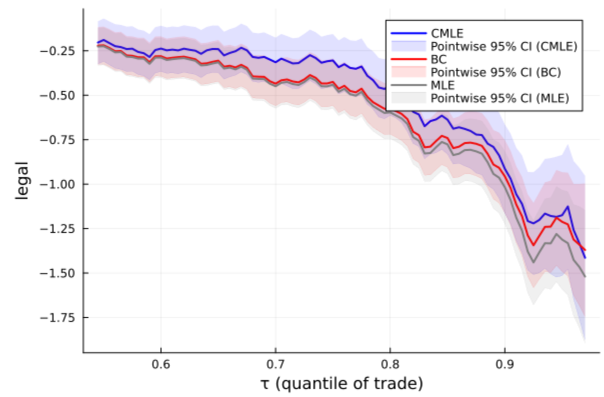
(e) Religion

Figure S.12: Pointwise comparison of CMLE, BC, and MLE estimates with 95% pointwise confidence intervals for all covariates. Solid blue: CMLE; solid red: BC; solid gray: MLE. Shaded regions show pointwise 95% confidence intervals for each estimator. Thresholds correspond to empirical quantiles in $[0.545, 0.990]$ at intervals of 0.005.

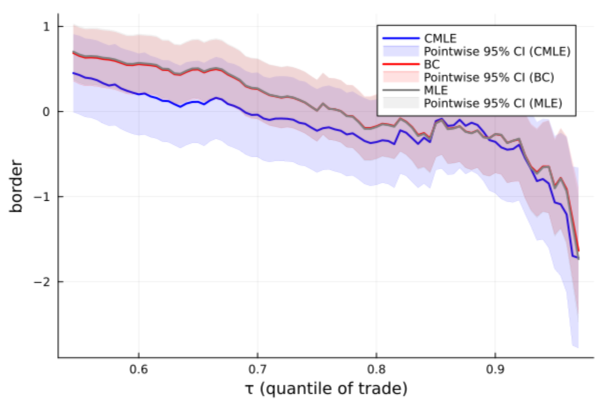
bilities approach zero or one.



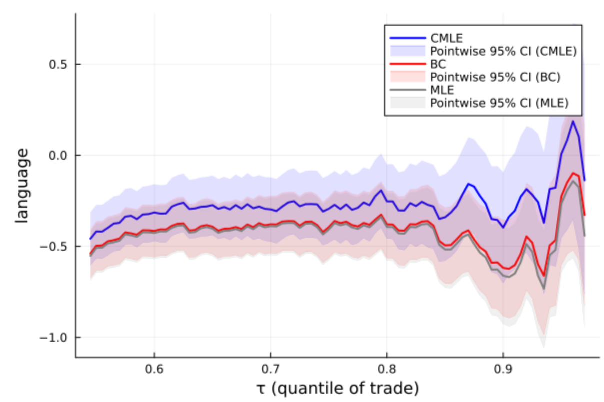
(a) Log distance



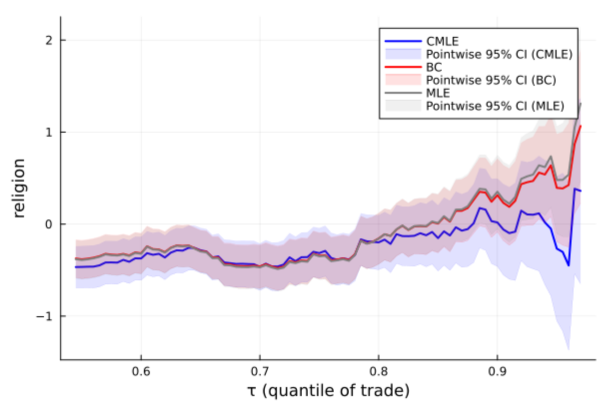
(b) Legal system



(c) Border

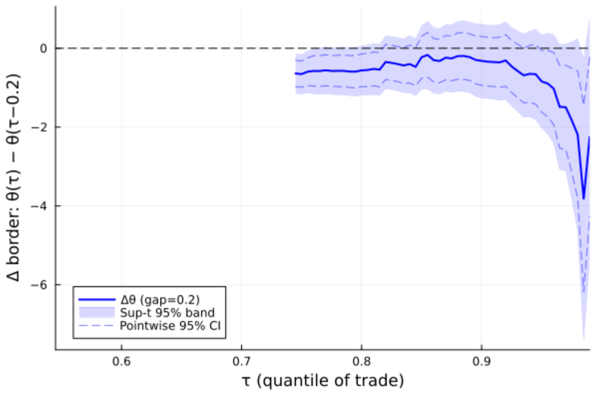


(d) Language

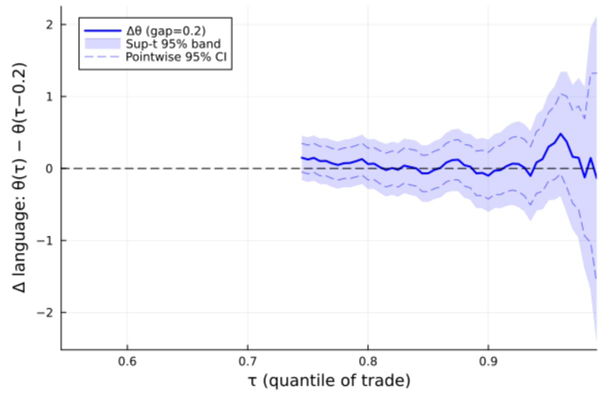


(e) Religion

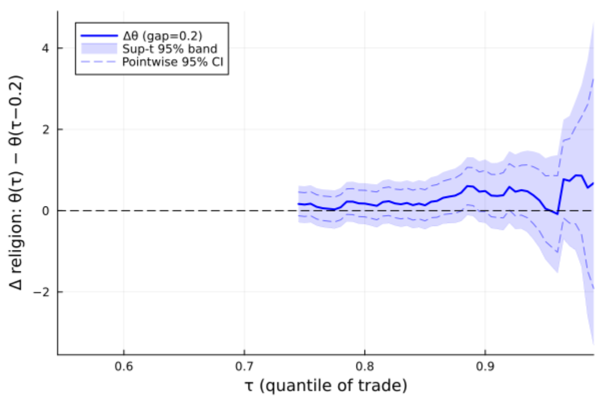
Figure S.13: Pointwise comparison of CMLE, BC, and MLE estimates with 95% pointwise confidence intervals for all covariates. Solid blue: CMLE; solid red: BC; solid gray: MLE. Shaded regions show pointwise 95% confidence intervals for each estimator. Thresholds correspond to empirical quantiles in $[0.545, 0.970]$ at intervals of 0.005.



(a) Border

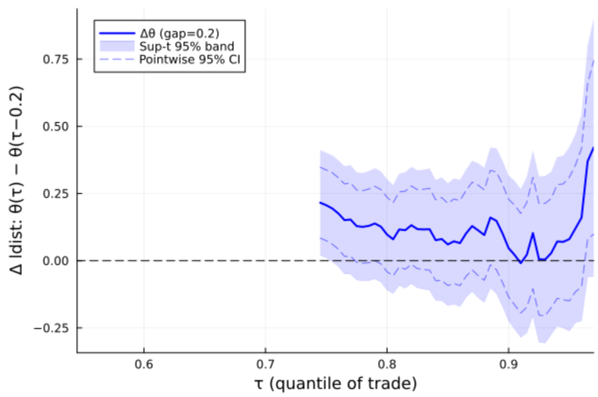


(b) Language

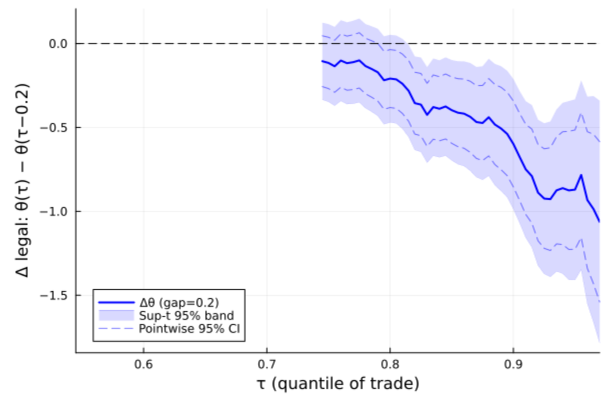


(c) Religion

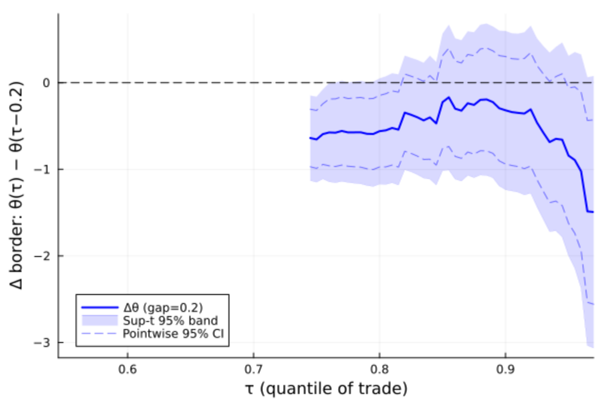
Figure S.14: Differences $\theta_{n,d}(\tau) - \theta_{n,d}(\tau - 0.20)$ from the CMLE estimator for log distance and legal system. Shaded region and dashed lines show the 95% simultaneous sup-t confidence band and pointwise confidence intervals for the differences, respectively. The horizontal dashed line marks zero. Maximal quantile considered is 0.990.



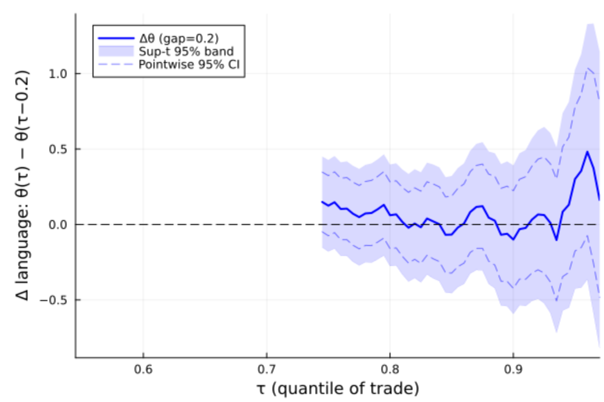
(a) Log distance



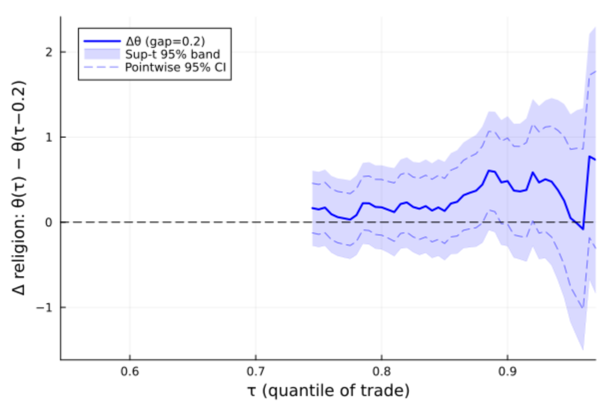
(b) Legal system



(c) Border



(d) Language



(e) Religion

Figure S.15: Differences $\theta_{n,d}(\tau) - \theta_{n,d}(\tau - 0.20)$ from the CMLE estimator for log distance and legal system. Shaded region and dashed lines show the 95% simultaneous sup-t confidence band and pointwise confidence intervals for the differences, respectively. The horizontal dashed line marks zero. Maximal quantile considered is 0.970.

References

Chernozhukov, V., I. Fernandez-Val, and M. Weidner (2024): "Network and panel quantile effects via distribution regression," *Journal of Econometrics*, 240, 105009.

Helpman, E., M. Melitz, and Y. Rubinstein (2008): "Estimating trade flows: Trading partners and trading volumes," *The Quarterly Journal of Economics*, 123, 441–487.

Jochmans, K. (2018): "Semiparametric analysis of network formation," *Journal of Business & Economic Statistics*, 36, 705–713.