Online Appendix to

"The Intensive Margin in Trade: How Big and How Important?"

Ana M. Fernandes
World Bank

Peter J. Klenow
Stanford and NBER

Sergii Meleshchuk
IMF

Martha Denisse Pierola
World Bank

Andrés Rodríguez-Clare *
UC Berkeley and NBER

February 5, 2022

*We are grateful to Arnaud Costinot, Caroline Freund, Cecile Gaubert, Keith Head, Sam Kortum, Thierry Mayer, Eduardo Morales and Jesse Perla for useful discussions, to seminar participants at various institutions, and to Matthias Hoelzlein and Nick Sander for outstanding research assistance. The World Bank provided access to the *Exporter Dynamics Database*. Research for this paper has in part been supported by the World Bank's Multidonor Trust Fund for Trade and Development and the Strategic Research Program on Economic Development, as well as the Stanford Institute for Economic Policy Research. The findings expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank or its member countries. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

Contents

A Extended Sample Details

Table A1: Additional EDD countries and years in the extended sample (only EDD statistics available)

ISO3	Country name	1st year	Last year
BRA	Brazil	2003	2013
DNK	Denmark	2003	2012
ESP	Spain	2005	2013
EST	Estonia	2003	2011
KWT	Kuwait	2009	2010
LKA	Sri Lanka	2013	2013
NOR	Norway	2003	2013
PRT	Portugal	2003	2012
SWZ	Swaziland	2012	2012
TUR	Turkey	2003	2013

B Intensive Margin: Robustness Checks

Intensive Margin Elasticity

Here we present details of various alternative approaches to estimating the IME discussed in Section 2..

In Table B1 we reproduce the regressions in Table 2 but for the extended sample of countries. In the preferred specification with origin-year and destination-year fixed effects, the IME is 0.38 for origin-destination pairs with at least 100 exporting firms and 0.52 for all origin-destination pairs. Tables B2 and B3 show the results of estimating the IME after excluding firms whose annual exports fell below $1,000 in any year, for the core and extended samples, respectively. The IME estimates change only slightly. Figure B1 shows IME estimates obtained separately for each year which range from 0.55 to 0.6.

When we allow the IME to differ across origin countries depending on their GDP per capita or on continents, we find that IME estimates are close to or larger than 0.4 for any group of countries, as shown in Table B4.

In Figure B2 we plot the (demeaned) intensive and extensive margins against total exports at the origin-industry-destination-year level using HS 2-digit industries. The pattern here is similar to that in Figure 1. Table B5 shows that the IME actually increases when moving to industry-level data. At the lowest level of aggregation available (HS 6-digit), for the core sample of countries the IME is 0.51 with origin-year-industry and destination-year-industry fixed effects. The results also hold in the extended sample, for which we calculated IME disaggregated at HS2 product level. As reported in Table B6, this IME is also close to 0.52. Estimates of the IME based on a sample including only HS 2-digit industries with low shares of firms exporting via intermediaries, as defined in Chan (2019), are shown in Table B7. The results show an almost unchanged IME at 0.53.

If measurement error in exports per firm and thus in total exports is serially uncorrelated, then instrumenting total exports with its leads and/or lags should yield an unbiased estimate of the IME. Table B8 shows that instrumented IMEs are very close to the OLS IME, both economically and statistically.

As an alternative to the use of cross-sectional variation in bilateral trade flows to estimate the IME, we exploit only time-series variation in bilateral export flows in Table B9. The results from regressions that include origin-destination fixed effects or regressions in first-differences (where the IME is identified only off the panel dimension) for the core and the extended sample show significant and larger IMEs in magnitude than those obtained exploiting cross-sectional variation in Table 2. The results for the extended sample (available upon request) are quantitatively almost identical. The evidence shows very clearly that the intensive margin is an important determinant of changes in bilateral export flows.

Table B1: IME regressions, extended sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
Panel a: country pairs with $N_{ij} \geq 100$			
IM elasticity	0.437***	0.450***	0.381***
Standard error	[0.0037]	[0.0028]	[0.0039]
R^2	0.55	0.75	0.83
Variation in $\ln X_{ij}$ explained by FE,%	0.01	0.14	0.55
Observations	14,318	14,300	13,964
Panel b: all country pairs			
IM elasticity	0.434***	0.477***	0.516***
Standard error	[0.0016]	[0.0016]	[0.0023]
R^2	0.69	0.77	0.80
Variation in $\ln X_{ij}$ explained by FE, %	0.00	0.22	0.56
Observations	52,775	52,775	52,658
Year FE	Yes		
Origin × year FE		Yes	Yes
Destination × year FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1 and Table A1 in the Online Appendix. Panel a) presents the regression on the sample of country-pairs with at least 100 exporters. Panel b) presents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B2: IME regressions, small firms excluded, core sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
Panel a: country pairs with $N_{ij} \geq 100$			
IM elasticity	0.437***	0.459***	0.398***
Standard error	[0.0058]	[0.0042]	[0.0055]
R^2	0.54	0.74	0.85
Variation in $\ln X_{ij}$ explained by FE,%	0.01	0.19	0.59
Observations	7,698	7,684	7,234
Panel b: all country pairs			
IM elasticity	0.497***	0.525***	0.573***
Standard error	[0.0018]	[0.0017]	[0.0022]
R^2	0.77	0.81	0.84
Variation in $\ln X_{ij}$ explained by FE, %	0.00	0.19	0.50
Observations	46,925	46,925	46,832
Year FE	Yes		
Origin × year FE		Yes	Yes
Destination × year FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports on log total exports per firm. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Average and total exports per destination are calculated using the sales of firms with at least $1000 to that destination. Panel a) represents the regression on the sample of country-pairs with at least 100 exporters. Panel b) represents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B3: IME regressions, small firms excluded, extended sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
Panel a: country pairs with $N_{ij} \geq 100$			
IM elasticity	0.437***	0.450***	0.379***
Standard error	[0.0037]	[0.0027]	[0.0039]
R^2	0.55	0.75	0.83
Variation in $\ln X_{ij}$ explained by FE,%	0.01	0.13	0.55
Observations	14,216	14,196	13,858
Panel b: all country pairs			
IM elasticity	0.431***	0.475***	0.512***
Standard error	[0.0015]	[0.0015]	[0.0021]
R^2	0.69	0.77	0.80
Variation in $\ln X_{ij}$ explained by FE, %	0.00	0.21	0.56
Observations	52,593	52,593	52,447
Year FE	Yes		
Origin × year FE		Yes	Yes
Destination × year FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1 and Table A1. Average and total exports per destination are calculated using the sales of firms with at least $1000 to that destination. Panel a) presents the regression on the sample of country-pairs with at least 100 exporters. Panel b) presents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B4: IME regressions and distance elasticity by income group and continent, core sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$ or ln distance					
Panel a: interactions with origin income group						
$\ln X_{ij}$ × richer	0.438***	0.444***	0.380***			
	[0.00576]	[0.00471]	[0.00587]			
$\ln X_{ij}$ × poorer	0.438***	0.527***	0.477***			
	[0.00584]	[0.00747]	[0.00914]			
ln distance × richer				0.0839***	0.136***	-0.263***
				[0.0109]	[0.0174]	[0.0167]
ln distance × poorer				0.0401***	0.0638*	-0.344***
				[0.0114]	[0.0320]	[0.0328]
Observations	7,736	7,723	7,310	7,736	7,723	7,310
R^2	0.543	0.744	0.851	0.076	0.307	0.691
Panel b: interactions with origin continent						
$\ln X_{ij}$ × Europe	0.467***	0.507***	0.450***			
	[0.00458]	[0.00861]	[0.00967]			
$\ln X_{ij}$ × America	0.476***	0.555***	0.470***			
	[0.0046]	[0.00861]	[0.00847]			
$\ln X_{ij}$ × Africa	0.461***	0.490***	0.422***			
	[0.00473]	[0.0131]	[0.0133]			
$\ln X_{ij}$ × Asia	0.431***	0.381***	0.319***			
	[0.00458]	[0.00557]	[0.00790]			
ln distance × Europe				0.157***	-0.0845***	-0.527***
				[0.0140]	[0.0253]	[0.0261]
ln distance × America				0.199***	0.373***	-0.273***
				[0.0123]	[0.0265]	[0.0318]
Distance × Africa				0.137***	0.529***	-0.0387
				[0.0123]	[0.0305]	[0.0374]
Distance × Asia				0.0978***	-0.280***	-0.0648
				[0.0114]	[0.0333]	[0.0431]
Observations	7,746	7,733	7,320	7,746	7,733	7,320
R^2	0.639	0.753	0.855	0.147	0.352	0.702
Year FE	Yes			Yes		
Origin-year FE		Yes	Yes		Yes	Yes
Destination-year FE			Yes			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log $\ln X_{ij}$ or on log distance interacted with origin income level group or continent. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B5: IME regressions, disaggregated within manufacturing, core sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
Panel a: HS 2-digit			
IM elasticity	0.569***	0.510***	0.467***
Standard error	[0.0022]	[0.0019]	[0.0060]
Observations	37,321	35,621	10,732
Panel b: HS 4-digit			
IM elasticity	0.651***	0.569***	0.515***
Standard error	[0.0018]	[0.0017]	[0.0085]
Observations	62,776	58,516	4,640
Panel c: HS 6-digit			
IM elasticity	0.664***	0.593***	0.508***
Standard error	[0.0019]	[0.0018]	[0.0013]
Observations	67,967	61,501	2,972
Year × HS FE	Yes		
Origin × Year × HS FE		Yes	Yes
Destination × Year × HS FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports. The data are aggregated at the origin-destination-year-HS industry level for a set of origin-years listed in Table 1. Panels a), b), and c) represent the regressions for industries defined at the HS 2-digit, 4-digit, and 6-digit levels respectively. The sample is restricted to the origin-destination-product cells with at least 100 exporters. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

Table B6: IME regressions, disaggregated by HS2 product, extended sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
IM elasticity	0.646***	0.598***	0.518***
Standard error	[0.0020]	[0.0019]	[0.0034]
Observations	58,609	56,560	29,906
Year × HS FE	Yes		
Origin × Year × HS FE		Yes	Yes
Destination × Year × HS FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports. The data are aggregated at the origin-destination-year-HS2 industry level for a set of origin-years listed in Table 1 and Table A1. The sample is restricted to country pairs with $N_{ij} \geq 100$. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B7: IME regression for industries with low share of exporting via intermediaries, core sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
Panel a: country pairs with $N_{ij} \geq 100$			
IM elasticity	0.437***	0.450***	0.381***
Standard error	[0.0037]	[0.0028]	[0.0034]
R^2	0.55	0.75	0.83
Variation in $\ln X_{ij}$ explained by FE,%	0.01	0.14	0.55
Observations	14,318	14,300	13,964
Panel b: all country pairs			
IM elasticity	0.434***	0.477***	0.516***
Standard error	[0.0016]	[0.0016]	[0.0016]
R^2	0.69	0.77	0.80
Variation in $\ln X_{ij}$ explained by FE, %	0.00	0.22	0.56
Observations	52,775	52,775	52,658
Year FE	Yes		
Origin × year FE		Yes	Yes
Destination × year FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports dropping industries with a large share of intermediaries as defined in Chan (2019). The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Panel a) presents the regression on the sample of country-pairs with at least 100 exporters. Panel b) presents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B8: IME instrumental variables regressions, core sample

	Coefficient from $\ln x_{ij}$ on $\ln X_{ij}$		
	IV lag	IV lead	IV lag and lead
Panel a: country pairs with $N_{ij} \geq 100$			
IM elasticity	0.392***	0.392***	0.399***
Standard error	[0.0061]	[0.0063]	[0.0068]
Observations	6,372	6,181	5,224
Panel b: all country pairs			
IM elasticity	0.476***	0.479***	0.468***
Standard error	[0.0028]	[0.0028]	[0.0030]
Observations	36,065	36,065	28,672
Origin × year FE	Yes	Yes	Yes
Destination × year FE	Yes	Yes	Yes

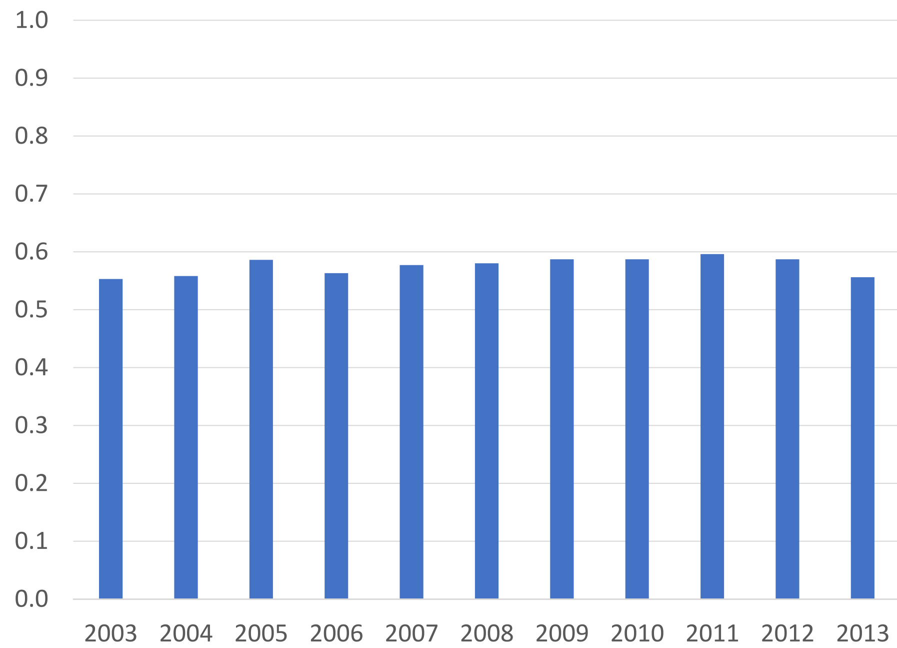
Note: the table presents the estimated coefficients of the regression of log average exports per firm on log total exports using lags and/or leads of the independent variable as instruments. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Panel a) presents the regression on the sample of country-pairs with at least 100 exporters. Panel b) presents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Table B9: IME regressions, exploiting time-series variation

	Core sample		Extended sample	
Panel a: country pairs with $N_{ij} \geq 100$				
IM elasticity (levels)	0.715***		0.734***	
Standard error	[0.0095]		[0.0071]	
IM elasticity (first differences)		0.848***		0.865***
Standard error		[0.0079]		[0.0056]
R^2	0.98	0.82	0.98	0.85
Observations	7,701	6,373	14,176	12,069
Panel b: all country pairs				
IM elasticity (levels)	0.861***		0.846***	
Standard error	[0.0022]		[0.0025]	
IM elasticity (first differences)		0.871***		0.876***
Standard error		[0.0025]		[0.0028]
R^2	0.97	0.91	0.97	0.89
Observations	45,972	36,213	51,661	42,236
Year FE	Yes		Yes	
Origin × destination FE	Yes		Yes	
First differences		Yes		Yes

Note: columns 1 and 3 of the table presents the estimated coefficients of the regression of log average exports on log total exports with year and origin-destination fixed effects. Columns 2 and 4 report the results of the same regression in first differences with no fixed effects. The data are aggregated at the year-origin-destination level for a set of origin-years listed in Table 1 and Table A1 . Panel a) represents the regression on the sample of country-pairs with at least 100 exporters. Panel b) represents the regression on the full sample. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

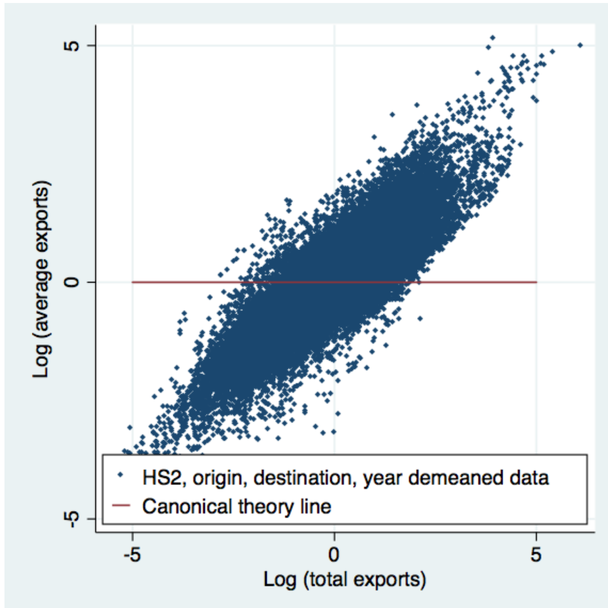
Figure B1: IME by year, data



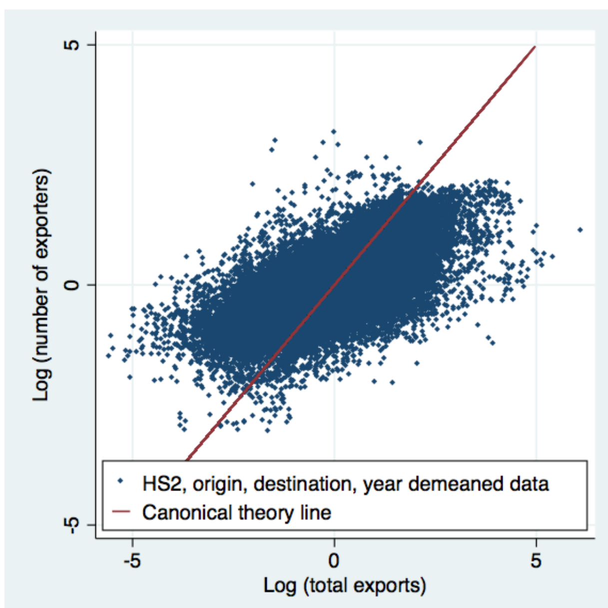
Source: Exporter Dynamics Database. The bars are intensive margin elasticities from yearly regressions that include origin and destination fixed effects. The core sample of countries, i.e., the set of origin-years listed in Table 1 is used.

Figure B2: Intensive and extensive margins of exporting, by industry

Panel a: Average size of exporters (intensive margin) and total exports



Panel b: Number of exporters (extensive margin) and total exports

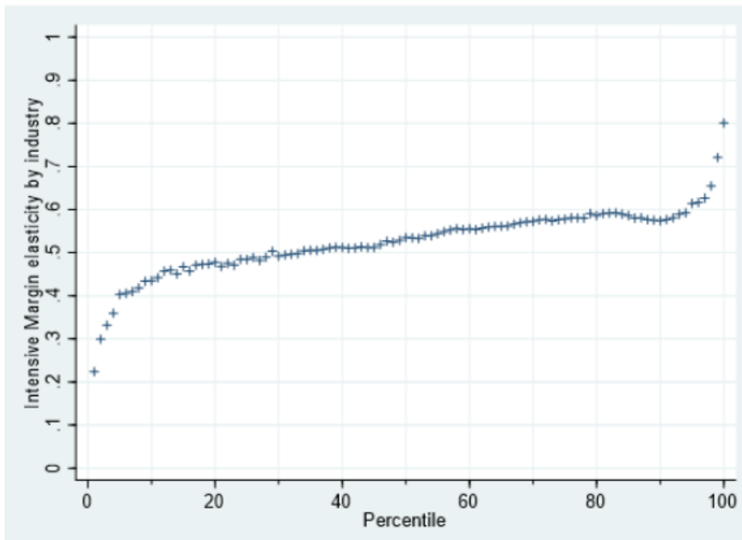


Source: Exporter Dynamics Database, extended sample of countries. The x-axis represents log total exports at the origin-HS 2-digit-destination-year level demeaned by origin-HS 2-digit-year and destination-HS 2-digit-year fixed effects. Only origin-HS 2-digit-destination triplets with more than 100 exporting firms are considered. The line is the slope predicted by the Melitz-Pareto model.

Intensive Margin Elasticity by Percentile

Figure B3 presents the estimated intensive margin across size percentiles disaggregated by industry relying on data at the origin-destination-year-HS 2-digit industry level. The IME by percentile within industries is qualitatively consistent with the original IME by percentile in Figure 2 (which was aggregated at the origin-destination-year level).

Figure B3: IME for each percentile, disaggregated by industry



Source: Exporter Dynamics Database, core sample of countries. The x-axis represents percentiles of the average exporter size distribution. Each plus represents the coefficient from the regression of log average exports per firm in an exporter size percentile on log total exports. The data used is at the is origin-destination-year-HS 2-digit industry level and it demeaned by origin-year-HS 2-digit industry and destination-year-HS 2-digit industry fixed effects.

Product Intensive Margin Elasticity

The product intensive margin regressions for the core sample in Table B10 show an estimate of about 0.29 when origin-year and destination-year fixed effects are controlled for.

Table B10: Product-level IME regressions, core sample

	Coefficient from $\ln x_{ij}^{p}$ on $\ln X_{ij}$		
IM elasticity	0.375***	0.393***	0.287***
Standard error	[0.0069]	[0.0055]	[0.0076]
R^2	0.36	0.59	0.77
Variation in $\ln X_{ij}$ explained by FE,%	0.01	0.20	0.59
Observations	7,485	7,472	7,023
Year FE	Yes		
Origin × year FE		Yes	Yes
Destination × year FE			Yes

Note: the table presents the estimated coefficients of the regression of log average exports per product (total exports divided by the number of HS6 products exported by all firms from origin i to destination j in a given year) on log total exports. The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Robust standard errors are reported in brackets. Egypt is not included in the sample since its data does not include HS 6-digit product level disaggregation. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

Intensive Margin and Distance

The elasticity of average exports per firm with respect to distance remains negative when disaggregated at the industry level, as reported in Table B11. In addition, the distance elasticity remains positive even when controlling for the destination country GDP as seen in Table B12.

Table B11: Intensive margin and distance, disaggregated within manufacturing, core sample

	Elasticity with respect to distance					
	HS 2-digit		HS 4-digit		HS 6-digit	
x_{ij}	0.0125	-0.561***	0.065	-0.701***	0.152***	-0.600***
Standard error	[0.0085]	[0.0244]	[0.0063]	[0.0415]	[0.0056]	[0.056]
Observations	35,505	10,615	58,470	4,586	61,501	2,972
Origin × year × HS FE	Yes	Yes	Yes	Yes	Yes	Yes
Destination × year × HS FE		Yes		Yes		Yes

Note: the table presents the estimated coefficients of the regression of log average exports per firm on log distance between origins and destinations. The data are aggregated at the origin-destination-year-HS industry level for a set of origin-years listed in Table 1. Population-weighted distance between origins and destinations is taken from Mayer and Zignago (2011). The sample is restricted to the origin-destination pairs with at least 100 exporters. Egypt is not included in the sample since its data does not include HS 6-digit product level disaggregation. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

Table B12: Distance elasticity, controlling for destination GDP, core sample

	$\ln x_{ij}$		$\ln x_{ij}^{p}$	
ln distance	0.128***	-0.239***	0.302***	-0.0825***
Standard error	[0.0158]	[0.0151]	[0.00165]	[0.00152]
R^2	0.31	0.53	0.34	0.55
Observations	7,450	7,114	7,450	7,114
ln destination GDP		Yes		Yes
Origin × year FE	Yes	Yes	Yes	Yes

Note: the table presents the estimated coefficients of the regression of log average exports and log average exports per product (total exports divided by the number of HS6 products exported by all firms from origin i to destination j in a given year) on log distance between origins and destinations and log destination GDP (in columns 2 and 4). The data are aggregated at the origin-destination-year level for a set of origin-years listed in Table 1. Population-weighted distance between origins and destinations is taken from Mayer and Zignago (2011). The sample is restricted to the origin-destination pairs with at least 100 exporters. Egypt is not included in the sample since its data does not include HS 6-digit product level disaggregation. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

C The Intensive Margin in the Melitz Model: Additional Results

To show that a positive IME requires $cov(\ln\widetilde{\tau}_{ij}, \ln\widetilde{F}_{ij}) < 0$ under Pareto-distributed productivity, note that if $\Omega(n)$ is a constant then equations (6) and (8) combined with $G_i(\varphi) = 1 - (\varphi/b_0)^{-\theta}$ imply

$$\ln x_{ij} = \mu_i^{x,o} + \mu_j^{x,d} + \ln\widetilde{F}_{ij} \tag{OA.1}$$

and

$$\ln N_{ij} = \mu_i^{N,o} + \mu_j^{N,d} - \theta\ln\widetilde{\tau}_{ij} - \bar{\theta}\ln\widetilde{F}_{ij} \tag{OA.2}$$

where $\mu_i^{x,o}$, $\mu_i^{x,d}$, $\mu_i^{N,o}$ and $\mu_j^{N,d}$ are functions of parameters as well as origin and destination variables. Combining the definition of the intensive margin elasticity given in the previous section (i.e., IME =

$\frac{\text{cov}(\ln \tilde{x}_{ij}, \ln \tilde{X}_{ij})}{\text{var}(\ln \tilde{X}_{ij})}$) with equations (OA.1) and (OA.2), the model implies that

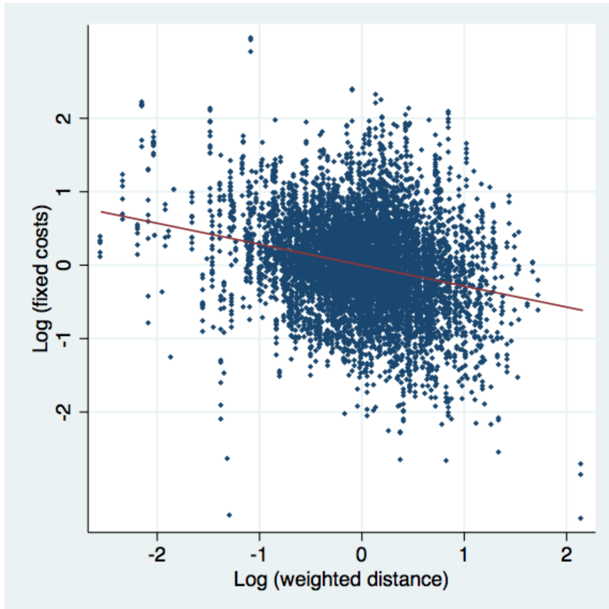
$$\text{IME} = \frac{-\left(\bar{\theta}-1\right)\text{var}(\ln \widetilde{F}_{ij}) - \theta \text{cov}(\ln \widetilde{\tau}_{ij}, \ln \widetilde{F}_{ij})}{\text{var}\left(-\theta \ln \widetilde{\tau}_{ij} - \left(\bar{\theta}-1\right)\ln \widetilde{F}_{ij}\right)}. \tag{OA.3}$$

Combined with the assumption that $\bar{\theta} > 1$, this result implies that IME> 0 if and only if $\text{cov}(\ln \widetilde{\tau}_{ij}, \ln \widetilde{F}_{ij}) < 0$.

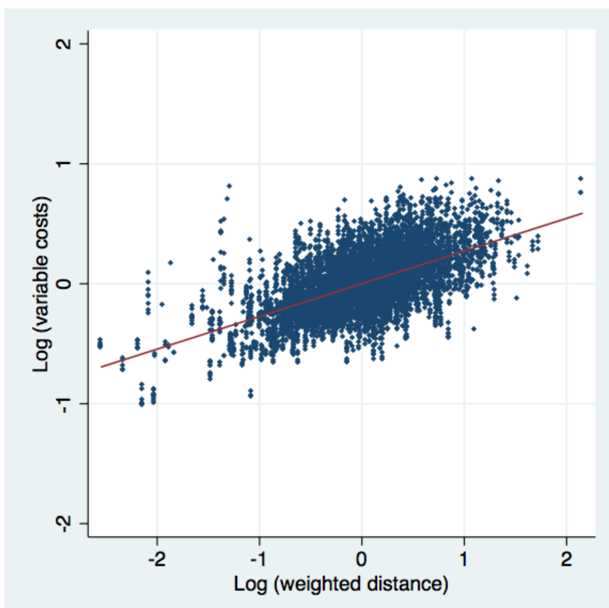
Figure C1 shows the plots between distance and either model-implied fixed trade costs or model-implied variable trade costs that correspond to the elasticities shown in Table 4. Model-implied fixed trade costs are decreasing with distance, while model-implied variable trade costs are increasing with distance.

Figure C1: Model-implied fixed and variable trade costs and distance

Panel a: fixed trade costs and distance



Panel b: variable trade costs and distance



Source: Exporter Dynamics Database. The x-axis represents log distance demeaned by origin and destination fixed effects taken from Mayer and Zignago (2011). The y-axis represents the fixed or variable trade costs implied by the basic Melitz-Pareto model demeaned by origin-year and destination-year fixed effects. To calculate the model-implied fixed and variable trade costs we use $\theta = 5$ from Head and Mayer (2014) and $\sigma = 5$ from Bas et al. (2017)

.

D Multi-Product Extension of Melitz-Pareto model

Here we study how allowing for multi-product firms affects the implications of the Melitz-Pareto model for the IME. As in Bernard et al. (2011), each firm can produce a differentiated variety of each of a continuum of products in the interval [0,1] with productivity $\varphi\lambda$, where φ is common across products and λ is product-specific. The firm component φ is drawn from a Pareto distribution $G^f(\varphi)$ with shape parameter θ^f, while the firm-product component λ is drawn from a Pareto distribution $G^p(\lambda)$ with shape parameter θ^p. To have well-defined terms given a continuum of firms, we impose $\theta^f > \theta^p > \sigma - 1$. To sell any products in market j, firms from country i have to pay a fixed cost F_{ij}, and to sell each individual product requires an additional fixed cost of f_{ij}. Variable trade costs are still τ_{ij}.

The cutoff λ for a firm from country i with productivity φ that wants to export to market j, $\lambda^*_{ij}(\varphi)$, is given implicitly by

$$A_j \left(\frac{w_i \tau_{ij}}{\varphi \lambda^*_{ij}(\varphi)}\right)^{1-\sigma} = \sigma f_{ij}. \tag{OA.4}$$

We can then write the profits in market j for a firm from country i with productivity φ as

$$\pi_{ij}(\varphi) \equiv \int_{\lambda^*_{ij}(\varphi)}^{\infty} \left[\left(\frac{\lambda}{\lambda^*_{ij}(\varphi)}\right)^{\sigma-1} - 1\right] f_{ij} \, dG^p(\lambda). \tag{OA.5}$$

The cutoff productivity for firms from i to sell in j is implicitly $\pi_{ij}(\varphi^*_{ij}) = F_{ij}$. As in the canonical model, the number of firms from country i that export to market j is $N_{ij} = \left[1 - G^f(\varphi^*_{ij})\right] N_i$, while the number of products sold by firms from i in j is $M_{ij} = N_i \int_{\varphi^*_{ij}}^{\infty} \left[1 - G^p\left(\lambda^*_{ij}(\varphi)\right)\right] dG^f(\varphi)$. Combining the previous expressions, using the fact that $G^p(\lambda)$ and $G^f(\varphi)$ are Pareto, writing $f_{ij} = f^o_i f^d_j \tilde{f}_{ij}$, $F_{ij} = F^o_i F^d_j \tilde{F}_{ij}$, and $\tau_{ij} = \tau^o_i \tau^d_j \tilde{\tau}_{ij}$, and defining variables appropriately we get

$$\ln X_{ij} = \mu^{X,o}_i + \mu^{X,d}_j - \theta^f \ln \tilde{\tau}_{ij} - \left(\frac{\theta^f}{\sigma-1} - \frac{\theta^f}{\theta^p}\right) \ln \tilde{f}_{ij} - \left(\frac{\theta^f}{\theta^p} - 1\right) \ln \tilde{F}_{ij}, \tag{OA.6}$$

$$\ln x^p_{ij} \equiv \ln X_{ij} - \ln M_{ij} = \mu^{x^p,o}_i + \mu^{x^p,d}_j + \ln \tilde{f}_{ij}, \tag{OA.7}$$

$$\ln x_{ij} \equiv \ln X_{ij} - \ln N_{ij} = \mu^{x^f,d}_i + \mu^{x^f,d}_j + \ln \tilde{F}_{ij}. \tag{OA.8}$$

It is easy to verify that if $f_{ij} = 0$ for all i, j then this model collapses to the canonical model with single-product firms.

Recalling our definition of the intensive margin elasticity at the firm and product level introduced in Section 2 and letting $\bar{\theta} \equiv \theta^f / (\sigma - 1)$ and $\chi \equiv \theta^f / \theta^p$, then from equations (OA.6) to (OA.8) we have

$$\text{IME} = -\frac{(\chi - 1)\,\text{var}\left(\ln \tilde{F}_{ij}\right) + \left(\bar{\theta} - \chi\right)\text{cov}(\ln \tilde{f}_{ij}, \ln \tilde{F}_{ij}) + \theta^f \,\text{cov}(\ln \tilde{F}_{ij}, \ln \tilde{\tau}_{ij})}{\text{var}(\ln \tilde{X}_{ij})} \tag{OA.9}$$

and

$$\mathrm{IME}^p = -\frac{\left(\bar{\theta}-\chi\right)\mathrm{var}\left(\ln\widetilde{f}_{ij}\right)+\left(\chi-1\right)\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{F}_{ij})+\theta^f+\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{\tau}_{ij})}{\mathrm{var}(\ln\widetilde{X}_{ij})}. \tag{OA.10}$$

As in the single-product firm model, if $\mathrm{var}\left(\ln\widetilde{F}_{ij}\right)=0$ then IME = 0, so this basic implication is not affected. We now have an analogous observation for the product-level intensive margin elasticity, namely that if $\mathrm{var}\left(\ln\widetilde{f}_{ij}\right)=0$ then $\mathrm{IME}^p=0$.

The assumption $\theta^f>\theta^p>\sigma-1$ implies that $\chi>1$ and $\bar{\theta}>\chi>1$ and in turn implies that for IME > 0 then either $\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{F}_{ij})<0$ or $\mathrm{cov}(\ln\widetilde{F}_{ij},\ln\widetilde{\tau}_{ij})<0$ (or both). In addition, if $\mathrm{IME}^p>0$ then either $\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{F}_{ij})<0$ or $\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{\tau}_{ij})<0$ (or both).

The result in the single-product model that, to match the data, we would need firm-level fixed trade costs to fall with distance remains valid in the multi-product firm extension. This comes directly from equation (OA.8) combined with the fact that $\mathrm{cov}(\ln\widetilde{x}_{ij},\ln\widetilde{dist}_{ij})<0$, which imply that $\mathrm{cov}(\ln\widetilde{F}_{ij},\ln\widetilde{dist}_{ij})$ must be negative in the model. We now have an analogous observation for product-level fixed trade costs: equation (OA.7) combined with the fact that $\mathrm{cov}(\ln\widetilde{x}^p_{ij},\ln\widetilde{dist}_{ij})<0$ (see Table 3) implies that $\mathrm{cov}(\ln\widetilde{f}_{ij},\ln\widetilde{dist}_{ij})<0$, so product-level fixed trade costs must also fall with distance. The results of the IME regression were reported above in Table B10.

E Granularity

Theory

With a discrete and finite number of firms it is possible to generate a positive covariance between the intensive margin and total exports even with $\mathrm{cov}(\tilde{F}_{ij})=0$. To state this formally, we rely on the extension of the Melitz-Pareto model to allow for granularity in Eaton et al. (2012). Equations (OA.1) and (OA.2) now become

$$\ln x_{ij}=\mu_i^{x,o}+\mu_j^{x,d}+\ln\widetilde{F}_{ij}+\varepsilon_{ij}, \tag{OA.11}$$

and

$$\ln N_{ij}=\mu_i^{N,o}+\mu_j^{N,d}-\theta\ln\widetilde{\tau}_{ij}-\bar{\theta}\ln\widetilde{F}_{ij}+\xi_{ij} \tag{OA.12}$$

where ξ_{ij} and ε_{ij} are error terms arising from the fact that now the number of firms is discrete and random. The equation analogous to (OA.3) now becomes

$$\mathrm{IME}=\frac{-\left(\bar{\theta}-1\right)\mathrm{var}(\ln\widetilde{F}_{ij})-\theta\mathrm{cov}\left(\ln\widetilde{\tau}_{ij},\ln\widetilde{F}_{ij}\right)+\mathrm{var}(\varepsilon_{ij})+\mathrm{COV}}{\mathrm{var}\left(-\theta\ln\widetilde{\tau}_{ij}-\left(\bar{\theta}-1\right)\ln\widetilde{F}_{ij}+\varepsilon_{ij}+\xi_{ij}\right)}, \tag{OA.13}$$

where $\mathrm{COV}\equiv\mathrm{cov}(\ln\widetilde{F}_{ij}+\varepsilon_{ij},\xi_{ij})-\mathrm{cov}(\bar{\theta}\ln\widetilde{F}_{ij}+\theta\ln\widetilde{\tau}_{ij},\varepsilon_{ij})$. If $\mathrm{var}(\varepsilon_{ij})$ is large relative to the other terms in the numerator then this could explain IME > 0 even with $\mathrm{cov}\left(\ln\widetilde{F}_{ij},\ln\widetilde{\tau}_{ij}\right)>0$. Thus, in theory, granularity could explain the positive intensive margin elasticity that we find in the data without relying on implausible patterns for fixed trade costs.

To check whether granularity is a plausible explanation for the positive IME in the data we will conduct two tests. First, we will estimate the fixed trade cost elasticity with respect to distance taking into account granularity and the possible biases it may induce. Second, we will simulate firm-level exports under granularity and the assumption of fixed trade costs that vary by origin and destination only and estimate the implied IME. We describe each of these tests in turn.

Fixed Trade Costs and Distance with Granularity

In the Melitz-Pareto model with a continuum of firms, average exports per firm can be expressed as $x_{ij} = \kappa F_{ij}$, where $\kappa \equiv \frac{\sigma\tilde{\theta}}{\tilde{\theta}-1}$. If we relax the continuum assumption to allow for granularity, then average exports per firm can be expressed as $x_{ij} = \kappa F_{ij} + \varepsilon_{ij}$, where ε_{ij} is an error term that arises from random realizations of productivity draws, the first moment of which is independent of any variables that determine bilateral fixed trade costs. If we further assume that $F_{ij} = F_i^o F_j^d e^{\zeta \ln dist_{ij}} + v_{ij}/\kappa$, where v_{ij} satisfies $\mathbb{E}(v_{ij}|dist_{ij}) = 0$, we can then write

$$x_{ij} = \kappa F_i^o F_j^d e^{\zeta \ln dist_{ij}} + u_{ij}, \tag{OA.14}$$

where $u_{ij} \equiv v_{ij} + \varepsilon_{ij}$ is an error term that captures both the deviation of F_{ij} from its mean as well as the granularity error term ε_{ij}. Since both $\mathbb{E}(v_{ij}|\ln dist_{ij})$ and $\mathbb{E}(\varepsilon_{ij}|\ln dist_{ij})$ are equal to zero, it follows that $\mathbb{E}(u_{ij}|dist_{ij}) = 0$. The challenge in estimating the fixed trade costs elasticity with respect to distance, ζ, from this equation is that we cannot simply take logs to obtain a log-linear equation to be estimated by OLS, because the error term that comes from granularity is not log-additive.

To take advantage of the time dimension of our data, we extend equation (OA.14) to allow for origin-time and destination-time specific components in the expression of fixed trade costs,

$$x_{ijt} = \kappa F_{it}^o F_{jt}^d e^{\zeta \ln dist_{ij}} + u_{ijt}, \tag{OA.15}$$

where again $\mathbb{E}(u_{ijt}|dist_{ij}) = 0$. We estimate equation (OA.15) using Poisson pseudo-maximum likelihood method as in Silva and Tenreyro (2011).

The IME under Granularity: Simulation

To assess how well granularity can explain a positive IME, we simulate exports of N_{ij} firms for each of the country pairs in the sample. We add demand shocks to allow for a less than perfect correlation between exports of different firms across different destinations. In the standard Melitz model with demand shocks, exports from i to j of a firm with productivity φ and destination-specific demand shock α_j can be calculated as

$$x_{ij}(\varphi, \alpha_j) = \sigma F_{ij} \left(\frac{\alpha_j \varphi}{\alpha^*_{ij} \varphi^*_{ij}} \right)^{\sigma-1}, \tag{OA.16}$$

where $\alpha^*_{ij}\varphi^*_{ij}$ is a combination of productivity and demand shocks of the smallest exporter from i selling to j. To estimate the IME in simulations we perform the following steps:

1. Draw φ and α_j from some distribution. The number of draws is equal to N_{ij}, the number of exporters in the EDD dataset for each origin-destination pair in 2009. To be more precise, we draw the product $\alpha_j\varphi$ for each firm-destination pair assuming either that, as in the standard Melitz model, there are no demand shocks and hence the product $\alpha_j\varphi$ is perfectly correlated across destinations or that, at the other extreme, there is no correlation in the product $\alpha_j\varphi$ across destinations (pure demand shocks case). In both cases, we draw $\alpha_j\varphi$ from a Pareto distribution with a shape parameter to be specified below.

2. Assume that $var\left(\widetilde{F}_{ij}\right) = 0$, so that $F_{ij} = F^o_i F^d_j$. This will allow us to study the IME generated by granularity by itself.

3. Use equation (OA.16) to simulate the exports for each firm and to calculate average exports per firm (in total and in each percentile) for each origin-destination pair.

4. Run the IME regression 1 on the simulated export data, with $\ln x_{ij}$ being either the intensive margin for all firms exporting from i to j, or for each percentile in the size distribution of exporters from i to j.

Evidence

We now discuss the evidence obtained first for the fixed trade costs elasticity with respect to distance and second for the IME with simulated data.

The results shown in Table E1 imply that although the distance elasticities are significantly lower than those estimated ignoring granularity in Table 4, they remain negative, indicating that model-implied fixed trade costs are decreasing with distance. Hence granularity does not help to eliminate one of the puzzles emerging from the comparison between the Melitz-Pareto model and the data.

Table E1: Fixed trade costs distance elasticity and granularity

	Fixed trade costs elasticity	
	Firm level	Product level
ζ	-0.022***	-0.007**
Standard error	[0.0029]	[0.0026]
Observations	7,320	7,320

Note: the table presents the estimated coefficients of the regression of the implied log fixed firm-level trade costs (column 1) and log fixed product-level trade costs (column 2) on log distance between origins and destinations using Poisson pseudo maximum likelihood procedure discussed in the Online Appendix E. Population-weighted distance between origins and destinations is taken from Mayer and Zignago (2011). The sample is restricted to the origin-destination pairs with at least 100 exporters. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

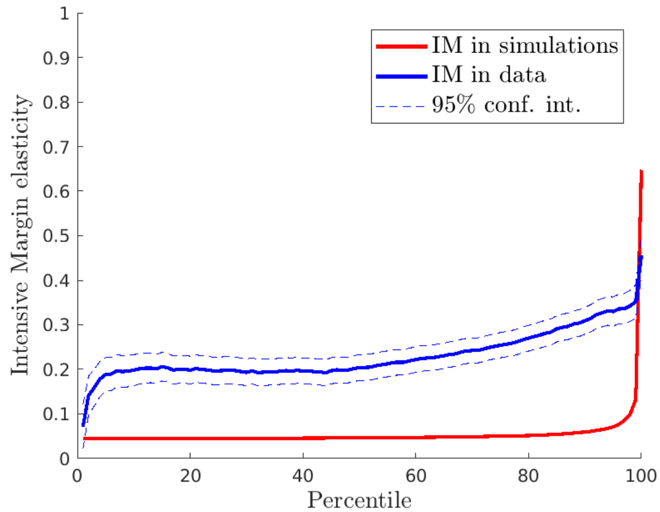
Table E2 reports the results of the simulation exercise. We simulate productivities for 3 values of $\bar{\theta}$: an estimate of θ using the procedure in EKK, as outlined in Appendix F, which yields $\bar{\theta} = 2.4$; the value that can be inferred from standard estimates of θ and σ in the literature (i.e., $\theta = 5$, the central estimate of the trade elasticity in Head and Mayer (2014), and $\sigma = 5$ from Bas et al. (2017), so $\bar{\theta} = 1.25$); and finally and $\bar{\theta} = 1$ (as in Zipf's Law). Two broad patterns emerge from the table. First, the simulated IME decreases with $\bar{\theta}$. This is because the effect of granularity on the IME is stronger when there is more dispersion in productivity levels. Second, the simulated IME is highest when productivity is less correlated across destinations, again because this gives granularity more room to generate a covariance between average exports per firm and total exports. For our estimate of $\bar{\theta}$ ($\bar{\theta} = 2.4$) and with no demand shocks (so there is perfect correlation in firm-level productivity across destinations), the simulated IME of 0.001 is quite low. The highest simulated IME occurs for the case in which $\bar{\theta} = 1$ and there is no correlation between the product of demand shocks and productivity across destinations. In this case the simulated IME is 0.33, not too far from our preferred estimate based on the data of 0.4. But we think of this as an extreme case because $\bar{\theta} = 1$ is far from the estimates that come out of trade data, and because of the implausible assumption that firm-level exports are completely uncorrelated across destinations. A low $\bar{\theta}$ also implies that, in contrast to the data, virtually all of the action behind a positive IME comes from the superstar firms. To see that, we calculate average simulated exports per firm in each percentile and use those to estimate an IME per percentile. We plot the resulting 100 IME estimates in Figure E1 along with the corresponding IME estimates based on the actual data. The IME based on the actual data is increasing with a spike at the top percentile. Granularity and the Pareto distribution fail to reproduce this pattern in the simulated data, since the corresponding IME is much smaller than in the data for most percentiles. The IME in the simulated data is almost zero for small percentiles and is relatively high for a small number of top percentiles. We conclude that granularity does not offer a plausible explanation for the positive estimated IME in the data.

Table E2: IME under granularity

	$corr(\alpha_j\varphi, \alpha_k\varphi)$	
	0	1
$\widetilde{\theta} = 2.4$	0.005	0.001
$\widetilde{\theta} = 1.25$	0.133	0.036
$\widetilde{\theta} = 1$	0.333	0.103

Note: the table presents the estimated coefficients of the regression of the implied log average exports on log total exports using numerical simulations. The sample is restricted to the year 2007 and to the origin-destination pairs with at least 100 exporters (867 observations). The first column reports the results from the model with zero correlation between the product of demand and productivity shocks across destinations. The second column reports the results for the model with perfect correlation. Three different values of $\widetilde{\theta}$ are used.

Figure E1: IME for each percentile, Pareto and granularity



Source: Exporter Dynamics Database. The darker solid line corresponds to IME for each percentile estimated using EDD and four main destinations: France, Germany, Japan and the U.S.. Dashed lines indicate 95% confidence intervals. The lighter solid line is IME for each percentile implied by the model with Pareto distribution of productivity and granularity, $\widetilde{\theta} = 1$. The level of bilateral fixed trade costs was chosen to match overall IME in the data. The number of draws for each origin-destination pair is equal to the number of exporters from origin to destination in EDD as of 2007.

F Estimation of $\widetilde{\theta}$

In this Appendix we estimate $\widetilde{\theta}$ following the same approach as in Eaton et al. (2011). First, we derive the following expression from the Melitz-Pareto model:

$$\frac{x_{il|j}}{x_{il|l}} = \left(\frac{N_{ij}}{N_{il}}\right)^{-1/\tilde{\theta}}, \tag{OA.17}$$

where $x_{il|j}$ are average exports per firm for firms from i that sell in market l but restricted to those firms that sell in markets l and j. EKK have information on domestic sales for each firm, so they use $l = i$. We do not have such information, so we use $l^*(i) \equiv \arg\max_k N_{ik}$, that is, the largest destination market for each origin country i (e.g., the United States for Mexico). Letting

$$z_{ij} \equiv \frac{x_{il^*(i)|j}}{x_{il^*(i)|l^*(i)}} \tag{OA.18}$$

and

$$m_{ij} \equiv \frac{N_{ij}}{N_{il^*(i)}} \tag{OA.19}$$

then we have

$$\ln z_{ij} = -\frac{1}{\bar{\theta}} \ln m_{ij}. \tag{OA.20}$$

This suggests an OLS regression to recover an estimate for $\bar{\theta}$.

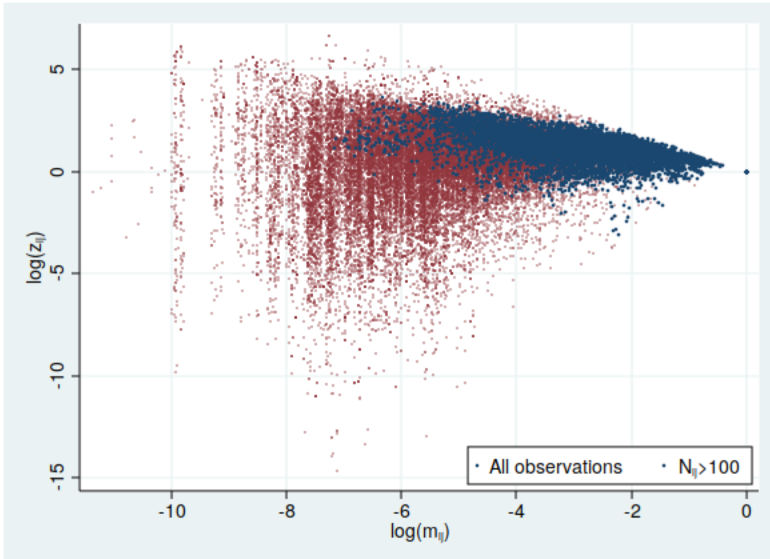
Eaton et al. (2011) estimate this regression for French firm-level data (including information on sales in France) and obtain a coefficient of −0.57, which implies $\bar{\theta} = 1.75$. In their case, they keep in their estimating sample only firms with positive sales in France, so the variables $x_{FF|j}$ and N_{Fj} are calculated based on the same set of firms. To implement an approach comparable to theirs, we drop all firms from country i that do not sell to $l^*(i)$, so the sample includes only $N_{il^*(i)}$ firms for country i. This implies that all firms that make up N_{ij} are also selling to $l^*(i)$. Figure F1 reproduces Figure 3 from Eaton et al. (2011) by plotting the variables in equation (OA.20). The slope in the graph is equal to $1/\bar{\theta}$, and the corresponding estimated values are reported in Table F1. Based on all observations in the core sample of countries and using no weighting, the estimated $\bar{\theta}$ is over 19. But in Figure F1 for small values of m_{ij}, which correspond to small values of N_{ij}, there is a lot of dispersion in z_{ij}. To minimize the effect of that noise we weight observations by $\sqrt{N_{ij}}$ and this lowers the estimate of $\bar{\theta}$ to 4.8. Finally, when we drop all observations with $N_{ij} < 100$ (remember that here N_{ij} measures the number of firms from country i that sell to country j *and* also to $l^*(i)$) we obtain $\bar{\theta} = 2.4$, which is still higher than in Eaton et al. (2011). We will use this estimate in our simulations of the intensive margin elasticity.

Table F1: Estimates of $\bar{\theta}$

	$\bar{\theta}$	s. e.	Observations
All observations, no weights	18.61***	[0.787]	39,712
Weights $\sqrt{N_{ij}}$	4.481***	[0.0360]	39,712
Dropping $N_{ij} < 100$	2.657***	[0.0175]	7,781
Dropping $M_{ij} < 100$	2.360***	[0.0147]	5,267

Note: the table presents estimates of $\widetilde{\theta}$ as discussed in Section F of this Online Appendix. N_{ij} denotes the number of exporters from i to j and M_{ij} denotes the number of exporters from i to j that also export to i's largest destination. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Figure F1: Exports to largest destination and market entry



Source: Exporter Dynamics Database, core sample of countries. The x-axis represents for each country i the log of the ratio of average exports per exporter to destination j to average exports per exporter to i's most popular destination market, $\log m_{ij}$. The y-axis represents for each country i the log of the ratio of the number of exporters to destination j to the number of exporters to i's most popular destination market, $\log z_{ij}$. A more formal definition of the variables can be found in the Online Appendix F. For the calculation of both average exports per exporter and number of exporters we focus only on firms from i that sell both in j and in the most popular destination.

G EKK extension

Eaton et al. (2011) (henceforth EKK) extend the basic Melitz-Pareto model to allow for lognormally distributed firm-level destination-specific demand and fixed-cost shocks. Except for constants that capture the net effects of these shocks, our equations (OA.1) and (OA.2) remain valid in the EKK environment, and hence the behaviour of the intensive margin in this environment is exactly the same as that of the simple Melitz-Pareto model as described in Section 3.1.[1]

It is important to note, however, that if productivity is distributed Pareto then the presence of log-normally distributed demand or fixed-cost shocks would imply that equations (OA.1) and (OA.2) no longer hold. The critical assumption in EKK that allows their model to be consistent with our equations (OA.1) and (OA.2) is that, loosely speaking, they consider the limit as the scale parameter of the Pareto distribution converges to zero. That is, EKK specify a function for the measure of firms with productivity above some level, with that measure going to infinity as productivity goes to zero. This is equivalent to taking a limit with the (exogenous) measure of firms going to infinity and the scale parameter of the

[1] This can be confirmed by simple manipulation of equations (20) and (28) in EKK.

Pareto distribution going to zero. Although equations (OA.1) and (OA.2) do not hold anywhere in this sequence, they do hold in the limit.

To formally establish this result, recall that to get equations (OA.1) and (OA.2) we assumed that $\varphi^*_{ij} > b_i$. If instead $\varphi^*_{ij} \le b_i$ then $N_{ij} = N_i$ and $x_{ij} = \left(\frac{\theta}{\theta-(\sigma-1)}\right) A_j \left(\frac{w_i \tau_{ij}}{b_i}\right)^{1-\sigma}$. In the extreme, if $\varphi^*_{ij} \le b_i$ holds for all i, j pairs, then we would have IME = 1 rather than IME = 0. Now think about the case with firm-specific demand and fixed-cost shocks. Specifically, assume that each firm is characterized by a productivity level φ as well as a demand shock α_j and a fixed cost shock f_j in each destination j, with φ drawn from a Pareto distribution (with scale parameter b_i and shape parameter θ) and α_j and f_j drawn iid from some distribution. Let $x_{ij}(\varphi, \alpha_j) \equiv A_j \alpha_j (\bar{\sigma} \frac{w_i \tau_{ij}}{\varphi})^{(1-\sigma)}$ and let $\varphi^*_{ij}(\alpha_j, f_j)$ be implicitly defined by $x_{ij}(\varphi^*_{ij}, \alpha_j) = \sigma f_j$. By the same argument we used in Section 3.1, if for all i, j and all possible (α_j, f_j) we have $\varphi^*_{ij}(\alpha_j, f_j) > b_i$, we can easily show that we still have IME = 0.[2] However, if α_j and f_j are lognormally distributed, then for $b_i > 0$ for all i there must be a positive mass of firms for which $\varphi^*_{ij}(\alpha_j, f_j) < b_i$, and for those firms there would be a positive intensive margin elasticity. EKK essentially avoid this by taking the limit with $b_i \to 0$ for all i.

In principle, one could use this result to argue that a Melitz model with Pareto distributed productivity but extended to allow for log-normally distributed demand and fixed-cost shocks could match the positive IME that we see in the data. However, such a model would not exhibit any of the convenient features of the canonical Melitz-Pareto model: the sales distribution is not distributed Pareto, the trade elasticity is not common across country pairs and fixed, and the gains from trade are not given by the ACR formula. Given that, our approach in this paper is to move all the way to a model where productivity as well as destination-specific demand and fixed-cost shocks are lognormally distributed. Such a model has at least the advantage that it is computationally tractable, and amenable to likelihood estimation methods, as we show in Section 4.

H QQ-Estimation of σ_φ

Exports from country i to country j of a firm with productivity φ in the model with CES preferences and monopolistic competition is given by $x_{ij}(\varphi) = \sigma F_{ij} \left(\varphi / \varphi^*_{ij}\right)^{\sigma-1}$. Since $\ln \varphi \sim N\left(\mu_{\varphi,i}, \sigma_\varphi\right)$ then $\ln x_{ij}(\varphi) \sim N_{trunc}\left(\bar{\mu}_{\varphi,ij}, \bar{\sigma}_\varphi; \ln\left(\sigma F_{ij}\right)\right)$, where $\bar{\sigma}_\varphi = \sigma_\varphi(\sigma-1)$, $\bar{\mu}_{\varphi,ij} = \mu_{\varphi,i}(\sigma-1) + \ln\left(\sigma F_{ij}\right) + (1-\sigma)\ln\left(\varphi^*_{ij}\right)$, and the truncation point is $\ln\left(\sigma F_{ij}\right)$.

As in Head et al. (2014), we estimate $\bar{\sigma}_\varphi$ using a quantile-quantile regression, which minimizes the distance between the theoretical and empirical quantiles of log exports. Empirical quantiles are given by

[2] Consider the group of firms from country i that have some given draw $\{(\alpha_j, f_j), j = 1, ..., n\}$. The exact same argument used in Section 3.1 can be used to show that the sample of firms obtained by combining such firms across all origins i satisfies IME = 0. One can then simply integrate across all possible draws $\{(\alpha_j, f_j), j = 1, ..., n\}$ to show that IME = 0 for the whole set of firms.

$\ln x_{ij,n}$, where n is the rank of the firm among exporters from i to j. We calculate theoretical quantiles of exports from i to j as $\bar{\mu}_{\varphi,ij} + \bar{\sigma}_{\varphi}\Phi^{-1}\left(\hat{\Phi}_{ij,n}\right)$, where $\hat{\Phi}_{ij,n} = \frac{N_i-(n-1)}{N_i}$ is the empirical CDF and N_i is the imputed number of firms from the BR data.[3] Following Head et al. (2014) we adjust the empirical CDF so that $\hat{\Phi}_{ij,n} = \frac{N_i-(n-1)-0.3}{N_i+0.4}$ since otherwise we would get $\Phi^{-1}\left(\hat{\Phi}_{ij,1}\right) = \infty$ when $n = 1$. The QQ-estimator of $\bar{\sigma}_{\varphi}$ is the coefficient β obtained from the regression

$$\ln x_{ij,n} = \alpha_{ij} + \beta\Phi^{-1}\left(\hat{\Phi}_{ij,n}\right) + \varepsilon_{ij,n}. \tag{OA.21}$$

Table H2 reports the QQ-estimate of $\bar{\sigma}_{\varphi}$. We report three sets of estimates: for the full sample, the largest 50% of firms and the largest 25% of firms for each origin-destination pair in each year. According to the model, the estimates of the slope should not change when we consider different sub-samples, but this is not the case in Table H2. This comes from a not very surprising empirical failure of the simple Melitz-lognormal model outlined in the first part of Section 3: whereas this model implies that the sales distribution for any country pair should be distributed as a truncated lognormal (with the truncation at sales of σF_{ij}), no such truncation exists in the data (i.e., we observe exporters with very small sales).

A related issue is that our estimates for either of the sub-samples are significantly larger than the estimate of 2.4 in Head et al. (2014). The difference comes from the fact that Head et al. (2014) assume that the sales distribution for any ij pair is lognormal, whereas we stick close to the simple model and assume that it is a truncated lognormal, and then use data for N_{ij} and our estimated values N_i to derive implicit truncation points. These truncation points tend to be on the right tail of the distribution, since N_{ij}/N_i tends to be quite low, hence the small $\bar{\sigma}_{\varphi}$ estimated by Head et al. (2014) would not be able to match the observed dispersion in the sales of exporters. In general, the higher the N_i one takes as an input in the QQ regression, the higher the estimate of the shape parameter one obtains.

In private correspondence, the authors of Head et al. (2014) pointed out that their approach would be consistent with the Melitz-lognormal model if one allows for heterogeneous fixed costs and lets the variance of these costs go to infinity, whereas our approach would be right if the variance goes to zero. This is part of our motivation in allowing for heterogeneous fixed costs and then in using MLE to estimate the full Melitz-lognormal model.

[3] The elasticity of the number of firms with respect to the population is close to 1, as reported in Table H1 and charted on Figure H1.

Table H1: Elasticity of number of firms to population

	log number of firms	
log population	0.945***	0.944***
Standard error	[0.0136]	[0.0139]
Observations	468	468
Year FE		Yes

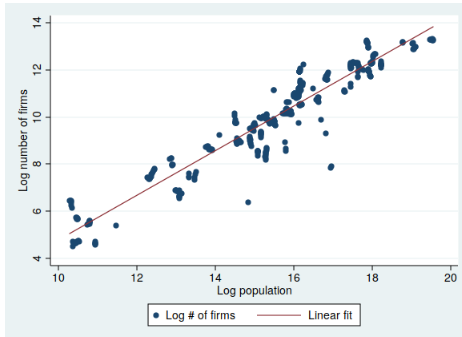
Note: the table presents the estimated coefficients of the regression of log number of firms, taken from Bento and Restuccia (2017) on log population, taken from World Development Indicators. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels, respectively.

Table H2: QQ estimates of $\bar{\sigma}_{\varphi}$

	All firms	Top 50%	Top 25%
$\bar{\sigma}_{\varphi}$	6.829***	4.676***	4.020***
	[0.0010]	[0.0006]	[0.0008]
Observations	11,902,823	5,917,685	2,949,514
R^2	0.81	0.93	0.94
Bilateral FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Note: the table presents estimates of $\widetilde{\sigma}_{\varphi}$ as discussed in Section H of this Online Appendix. Robust standard errors are reported in brackets. *, **, and *** represent the 5%, 1%, and 0.1% significance levels respectively.

Figure H1: Number of firms and population



Source: the x-axis represents log of population taken from the World Development Indicators and the y-axis represents the number of firms as computed by Bento and Restuccia (2017). The sample includes all country-years for which the EDD and the data from Bento and Restuccia (2017) overlap.

I Full Melitz-Lognormal model

Join distribution

Joint distribution of productivity, demand, and fixed trade costs in the full Melitz-lognormal model is given by:

$$\begin{bmatrix} \ln\varphi \\ \ln\alpha_1 \\ \vdots \\ \ln\alpha_J \\ \ln f_1 \\ \vdots \\ \ln f_J \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_{\varphi,i} \\ \mu_\alpha \\ \vdots \\ \mu_\alpha \\ \mu_{f,i1} \\ \vdots \\ \mu_{f,iJ} \end{bmatrix}, \begin{bmatrix} \sigma^2_{\varphi,i} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma^2_{\alpha,i} & \dots & 0 & \sigma_{\alpha f,i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_{\alpha,i} & 0 & \dots & \sigma_{\alpha f,i} \\ 0 & \sigma_{\alpha f,i} & \dots & 0 & \sigma^2_{f,i} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\alpha f,i} & 0 & \dots & \sigma^2_{f,i} \end{bmatrix}\right). \tag{OA.22}$$

Free entry condition

To show equation 21, note that profits gross of fixed costs by firms from i are equal to $\frac{1}{\sigma}\sum_j \lambda_{ij} X_j$ and total fixed costs of exporting by firms from i in market j are

$$N_i w_i e^{\mu_{f,ij}} \int_{-\infty}^{\infty} \int_{-\infty}^{h_{ij}+\tilde{\varphi}} e^{\tilde{f}} g(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi},$$

hence the free entry condition for firms in i is

$$F^e w_i N_i = \sum_j \left(\frac{1}{\sigma} \lambda_{ij} X_j - N_i w_i e^{\mu_{f,ij}} \int_{-\infty}^{\infty} \int_{-\infty}^{h_{ij}+\tilde{\varphi}} e^{\tilde{f}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi} \right).$$

Using (17), (18) and (20) we get

$$e^{\mu_{f,ij}} = \frac{\lambda_{ij} X_j}{e^{h_{ij}} N_i \sigma w_i \int_{-\infty}^{\infty} e^{\tilde{\varphi}} \int_{-\infty}^{h_{ij}+\tilde{\varphi}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}},$$

and so we can rewrite the free entry condition as equation (21).

Counterfactuals

A system of equations to compute counterfactual changes in the endogenous variables is given by:

$$h_{ij}(\hat{h}_{ij} - 1) = \ln \left(\frac{(\hat{w}_i \hat{\tau}_{ij})^{1-\sigma} \hat{P}_j^{\sigma-1} \hat{X}_j}{\hat{w}_i} \right) \tag{OA.23}$$

$$\hat{P}_j^{1-\sigma} = \sum_k \lambda_{kj} \hat{P}_{kj}^{1-\sigma} \tag{OA.24}$$

$$\hat{P}_{ij}^{1-\sigma} = \hat{N}_i (\hat{w}_i \hat{\tau}_{ij})^{1-\sigma} \frac{\int_{-\infty}^{\infty} e^{\tilde{\varphi}} \int_{-\infty}^{\hat{h}_{ij} h_{ij}+\tilde{\varphi}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}}{\int_{-\infty}^{\infty} e^{\tilde{\varphi}} \int_{-\infty}^{h_{ij}+\tilde{\varphi}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}} \tag{OA.25}$$

$$\hat{\lambda}_{ij} = \frac{\hat{P}_{ij}^{1-\sigma}}{\hat{P}_j^{1-\sigma}} \tag{OA.26}$$

$$\begin{aligned} \hat{w}_i \hat{N}_i \sum_j \lambda_{ij} X_j & \left(1 - \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{h_{ij}+\tilde{\varphi}} e^{\tilde{f}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}}{e^{h_{ij}} \int_{-\infty}^{\infty} e^{\tilde{\varphi}} \int_{-\infty}^{h(n_{ij})+\tilde{\varphi}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}} \right) \\ = \sum_j \lambda_{ij} X_j \hat{\lambda}_{ij} \hat{X}_j & \left(1 - \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\hat{h}_{ij} h_{ij}+\tilde{\varphi}} e^{\tilde{f}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}}{e^{\hat{h}_{ij}} \int_{-\infty}^{\infty} e^{\tilde{\varphi}} \int_{-\infty}^{\hat{h}_{ij} h_{ij}+\tilde{\varphi}} g_{ij}(\tilde{\varphi}, \tilde{f}) d\tilde{f} d\tilde{\varphi}} \right) \end{aligned} \tag{OA.27}$$

$$\hat{w}_i Y_i = \sum_j \lambda_{ij} X_j \hat{\lambda}_{ij} \hat{X}_j \tag{OA.28}$$

$$\hat{X}_j X_j = \hat{w}_j Y_j + \hat{\Delta}_j (X_j - Y_j) \tag{OA.29}$$

J Quasi-Bayesian Estimation for the full Melitz-lognormal model

The likelihood function is a product of density functions of individual firms that sell or do not sell to multiple destinations. In this section we will use the notation from Section 4 of the paper. Let $\tilde{\varphi}_i \equiv (\sigma - 1)[\ln \varphi - \mu_{\varphi,i}]$ be a random variable that denotes deviations from mean productivity for country i (adjusted by $\sigma - 1$). Individual firm density of export sales $(x_{i1}, ..., x_{iJ})$ can be written as

$$f_{X_{i1},..,X_{iJ}}(x_{i1},...,x_{iJ}) = \int_{\omega} f_{X_{i1},...,X_{iJ}|\bar{\varphi}_i}(x_{i1},...,x_{iJ}|\omega) f_{\bar{\varphi}_i}(\omega) d\omega \tag{OA.30}$$

$$= \int_{\omega} \prod_j f_{X_{ij}|\bar{\varphi}_i}(x_{ij}|\omega) f_{\bar{\varphi}_i}(\omega) d\omega, \tag{OA.31}$$

where the second equality comes from the fact that, conditional on productivity, sales are independent across markets (as well as the probability of selling to those markets). We now need to characterize $f_{X_{ij}|\bar{\varphi}_i}(f_{ij}|\omega)$ to calculate the likelihood function. In general we have

$$f_{X_{ij}|\bar{\varphi}_i}(x_{ij}|\omega) = \left[f_{Z_{ij}|\bar{\varphi}_i}(x_{ij}|\omega) \Pr\{Z_{ij} \geq \ln\sigma + \ln f_{ij}|\bar{\varphi}_i = \omega, Z_{ij} = x_{ij}\}\right]^{\mathbb{I}(x_{ij} \neq \emptyset)} \times$$
$$\times \left[\Pr\{\ln\sigma + \ln f_{ij} \geq Z_{ij}|\bar{\varphi}_i = \omega\}\right]^{\mathbb{I}(x_{ij} = \emptyset)}. \tag{OA.32}$$

The term in the first line of equation (OA.32) corresponds to the density function for the cases when we observe exports, while the second line corresponds to the mass at the point $x_{ij} = \emptyset$.

For the case when sales are not zero $X_{ij} = Z_{ij}$ and

$$Z_{ij}|\left[\bar{\varphi}_i = \omega\right] = \omega + d_{ij} + \ln\alpha - \mu_\alpha, \tag{OA.33}$$

$$Z_{ij}|\left[\bar{\varphi}_i = \omega\right] \sim N(d_{ij} + \omega, \sigma^2_{\alpha,i}). \tag{OA.34}$$

In addition

$$\Pr[Z_{ij} \geq \ln\sigma + \ln f_{ij}|\bar{\varphi}_i = \omega, Z_{ij} = x_{ij}] = \Pr[\ln\sigma + \ln f_{ij} \leq x_{ij}|\ln\alpha - \mu_\alpha = x_{ij} - d_{ij} - \omega], \tag{OA.35}$$
$$\ln\sigma + \ln f_{ij}|\left[\ln\alpha - \mu_\alpha = x_{ij} - d_{ij} - \omega\right] \sim N\left(\mu_1, \sigma^2_{1,i}\right),$$

where

$$\mu_1 \equiv \bar{\mu}_{f,ij} + \frac{\sigma_{\alpha f,i}}{\sigma^2_{\alpha,i}}(x_{ij} - d_{ij} - \omega),$$
$$\sigma^2_{1,i} \equiv \sigma^2_{f,i}(1 - \rho^2_i).$$

Finally we have

$$\Pr[Z_{ij} \leq \ln\sigma + \ln f_{ij}|\bar{\varphi}_i = \omega] = \Pr[-\ln\sigma - \ln f_{ij} + (\ln\alpha - \mu_\alpha) + d_{ij} \leq -\omega], \tag{OA.36}$$

$$-\ln\sigma - \ln f_{ij} + (\ln\alpha - \mu_\alpha) + d_{ij} \sim N(-\bar{\mu}_{f,ij} + d_{ij}, \sigma^2_{2,i}), \tag{OA.37}$$

where

$$\sigma_{2,i}^2 \equiv \sigma_{f,i}^2 + \sigma_{\alpha,i}^2 - 2\sigma_{\alpha f,i}.$$

Let ϕ and Φ denote the PDF and CDF of the standard normal, respectively. Plugging functional forms into equation (OA.32) we can get the object of interest,

$$f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ}) = \int_\omega \prod_j \left\{ \left[\frac{1}{\sigma_{\alpha,i}} \phi\left(\frac{x_{ij}-d_{ij}-\omega}{\sigma_{\alpha,i}}\right) \Phi\left(\frac{x_{ij} - \left[\bar{\mu}_{f,ij} + \frac{\sigma_{\alpha f,i}}{\sigma_{\alpha,i}^2}(x_{ij}-d_{ij}-\omega)\right]}{\sqrt{\sigma_{f,i}^2(1-\rho_i^2)}} \right) \right]^{\mathbb{I}(x_{ij}\neq\emptyset)} \times \right.$$
$$\left. \times \left[\Phi\left(\frac{-\omega + \bar{\mu}_{f,ij} - d_{ij}}{\sqrt{\sigma_{f,i}^2 + \sigma_{\alpha,i}^2 - 2\sigma_{\alpha f,i}}} \right) \right]^{\mathbb{I}(x_{ij}=\emptyset)} \right\} \frac{1}{\sigma_{\bar{\varphi},i}} \phi\left(\frac{\omega}{\sigma_{\bar{\varphi},i}}\right) d\omega \tag{OA.38}$$

However, since we only have a truncated sample of $X'_{ij}s$ (as we do not observe sales of firms that do not export), we need to normalize the density by the inverse of probability that a firm is selling to at least one destination, and so we are interested in the object

$$g_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ}) = f_{X_{i1},..,X_{iJ} \cap \text{Is an exporter}}(x_{1i},...,x_{iJ} \cap \text{Is an exporter}), \tag{OA.39}$$

and hence

$$\begin{aligned} g_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ}) &= \\ \frac{f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ})}{\Pr_i[\text{observe sales to at least 1 destination}]} &= \\ \frac{f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ})}{1-\Pr_i[\text{observe sales to no destinations}]} &= \\ \frac{f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ})}{1-\int_\omega \prod_j \left[\Phi\left(\frac{-\omega+\bar{\mu}_{f,ij}-d_{ij}}{\sqrt{\sigma_{f,i}^2+\sigma_{\alpha,i}^2-2\sigma_{\alpha f,i}}} \right) \right] \frac{1}{\sigma_{\bar{\varphi},i}} \phi\left(\frac{\omega}{\sigma_{\bar{\varphi},i}}\right) d\omega}&. \end{aligned} \tag{OA.40}$$

Let C_i denote the probability that a firm from origin i sells to at least one of the destinations we consider.[4] We know the number of firms, N_i^e, that sell to those destinations, and we can thus infer the number of draws $N_i = N_i^e / C_i$.

The likelihood function is a product of density functions as in equation (OA.40). Parameters to estimate are

$$\Theta_i = \left\{ \left\{ d_{ij}, \bar{\mu}_{f,ij} \right\}_{i,j}, \bar{\sigma}_{\varphi,i}, \sigma_{\alpha,i}, \sigma_{f,i}, \rho_i \right\}.$$

[4] C_i is given by the denominator in OA.40.

We next compute the density in the numerator of equation (OA.40). We can write this density in the following general form:

$$
\begin{aligned}
f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ}) &= \int_{\omega} G(\omega)\phi\left(\frac{\omega}{\sigma_{\bar{\varphi},i}}\right)d\omega \\
&= \int_{\omega} G(\omega)\frac{1}{\sqrt{2\pi}}exp\left(-\left[\frac{\omega}{\sqrt{2}\sigma_{\bar{\varphi},i}}\right]^2\right)d\omega,
\end{aligned} \tag{OA.41}
$$

where $G(\omega)$ is a known function of ω. Using change of variables $\widetilde{\omega} = \frac{\omega}{\sqrt{2}\sigma_{\bar{\varphi},i}}$ and $d\omega = \sqrt{2}\sigma_{\bar{\varphi},i}d\widetilde{\omega}$ we can write:

$$
f_{X_{i1},..,X_{iJ}}(x_{1i},...,x_{iJ}) = \int_{\widetilde{\omega}} G(\sqrt{2}\sigma_{\bar{\varphi},i}\widetilde{\omega})\frac{\sigma_{\bar{\varphi},i}}{\sqrt{\pi}}exp\left(-\widetilde{\omega}^2\right)d\widetilde{\omega}. \tag{OA.42}
$$

We can speed up the process to calculate the object in equation (OA.42) by applying a Gauss-Hermite quadrature. In general:

$$
\int_x g(x)exp(-x^2)dx \approx \sum_k g(r_k)w_k, \tag{OA.43}
$$

where r_k are the roots of the Hermite polynomial and ω_k are associated weights. We calculate 33 values of r_k and corresponding weights w_k using the Gauss-Hermite method. The number of points was chosen to ensure that the quadrature approximation is accurate and that calculations take a reasonable amount of time.

K MCMC algorithm

Since the likelihood function in equation (14) in the paper is highly nonlinear there may exist multiple local maxima. We thus estimate a vector of parameters Θ (defined in equation (15) in the paper) using the methodology developed in Chernozhukov and Hong (2003). This procedure not only gives us point estimates, but also yields confidence intervals for the estimated parameters, intensive marginal elasticity implied by the model, and elasticity of trade costs with respect to distance. We implement Chernozhukov and Hong (2003) procedure using the Metropolis-Hastings Monte-Carlo Markov chain algorithm. This algorithm yields a chain of parameter draws $\left\{\Theta_i^{(n)}\right\}_{n=1}^N$ for each origin i such that $\bar{\Theta}_i \equiv \frac{1}{N}\sum_n \Theta_i^{(n)}$ is a consistent estimate of Θ_i. Moreover, using the values of the parameters in the chain we can construct confidence intervals for some functions $f(\Theta_i)$. The chain of parameters $\left\{\Theta_i^{(n)}\right\}_{n=1}^N$ for each origin is constructed in the following way:

Step 1. Randomly choose a starting guess $\Theta_i^{(0)}$.

Step 2. Draw a candidate vector of parameters for the chain's $n+1$ value as $\widetilde{\Theta}_i^{n+1} = \Theta_i^{(n)} + \psi^{(n)}$, where $\psi^{(n)}$ is a vector of *iid* shocks taken from the mean-zero normal distribution. The variance-covariance matrix of this distribution is diagonal. The initial values of the diagonal elements are set at $0.2\Theta_i^{(0)}$. At each step all but 1 elements of $\psi^{(n)}$ are zero. In other words, we only add an *iid* shock to one parameter at each step of the chain. Since the vector Θ_i has 34 elements for each i, we try a new value for each parameter every 34 steps.

Step 3. Calculate $\Theta_i^{(n+1)}$ in the following way:

$$\Theta_i^{(n+1)} = \begin{cases} \widetilde{\Theta}_i^{(n+1)} & \text{with probability } \min\left[1; L\left(\widetilde{\Theta}_i^{(n+1)}\right) - L\left(\Theta_i^{(n)}\right)\right] \\ \Theta_i^{(n)} & \text{otherwise,} \end{cases}$$

where $L(\theta)$ is defined in equation 14 in the paper. Every 3,400 iterations (100 iterations per parameter) during the first 100,000 iterations we update diagonal elements of the variance-covariance matrix of the shocks so that the acceptance rate for each parameter is in the interval 0.25–0.35, as recommended by Chernozhukov and Hong (2003). We calculate the acceptance rate as a share of draws for which $\Theta_i^{(n+1)} = \widetilde{\Theta}_i^{(n+1)}$.

We repeat the procedure until we have at least 34 million draws (1 million draws per parameter) in the chain after we discard the first 100,000 draws (known as "burn-in period"). For each origin we construct 5 different chains with different starting guesses to check that our estimates are robust with respect to the starting values (we discuss convergence of the chains in Appendix L).

Having estimated the chains, we take 1,000 random draws from the chains for each origin with replacement. We use those draws to calculate point estimates (averages) as well as 95% confidence intervals. Using those draws we simulate the model 1,000 times and run IME regressions the way we run them in the data. Finally, we run 1,000 regressions of the estimates d_{ij} and $\mu_{f,ij}$ on log distance with origin and destination fixed effects for the 4 destinations (USA, France, Germany, and Japan) and interpret the results as elasticity of variable and fixed trade costs with respect to distance.

L Convergence of the Monte Carlo Markov chains

We ran our estimation algorithm 5 times per origin country which gave us 5 different chains that started at different random initial guesses. This lets us compare underlying distributions of parameter estimates across different chains. We checked the convergence of the chains using the following criteria:

1) comparing the means of parameter estimates in the first and second half of the chain;

2) comparing parameter means across different chains for the same origin.

If the means calculated for the two parts of a chain are statistically indistinguishable, we conclude that the chain converged. If the means across different chains are statistically indistinguishable, we conclude that multiple chains converged to the same region. It turns out that in some cases the chains did not converge.[5] In those cases we disregarded the chains that didn't converge. Out of remaining chains, we randomly choose one chain per origin to conduct our numerical analysis.

Figures L1-L34 present the moments from the parameter draws that we obtained using the procedure to estimate the full Melitz-lognormal model parameters. Each graph corresponds to one parameter and consists of multiple panels for different origin countries in the EDD. Each panel consists of five horizontal areas corresponding to 5 different MCMC chains that we ran for each origin. The black, red, and blue stars denote the mean of the parameter draws in the full chain, the first half of the chain, and the second half of the chain, respectively. The black lines correspond to the 2.5 – 97.5 percentiles range for the draws. The plots show that the means and their confidence intervals are virtually indistinguishable across all the chains and subchains that converged for all countries and all parameters.

[5]It happens, for example, when some of the values for $\sigma_{\varphi,i}$ and $\sigma_{\alpha,i}$ exploded.

Figure L1: Summary plots for QBE chains: d_{i1}

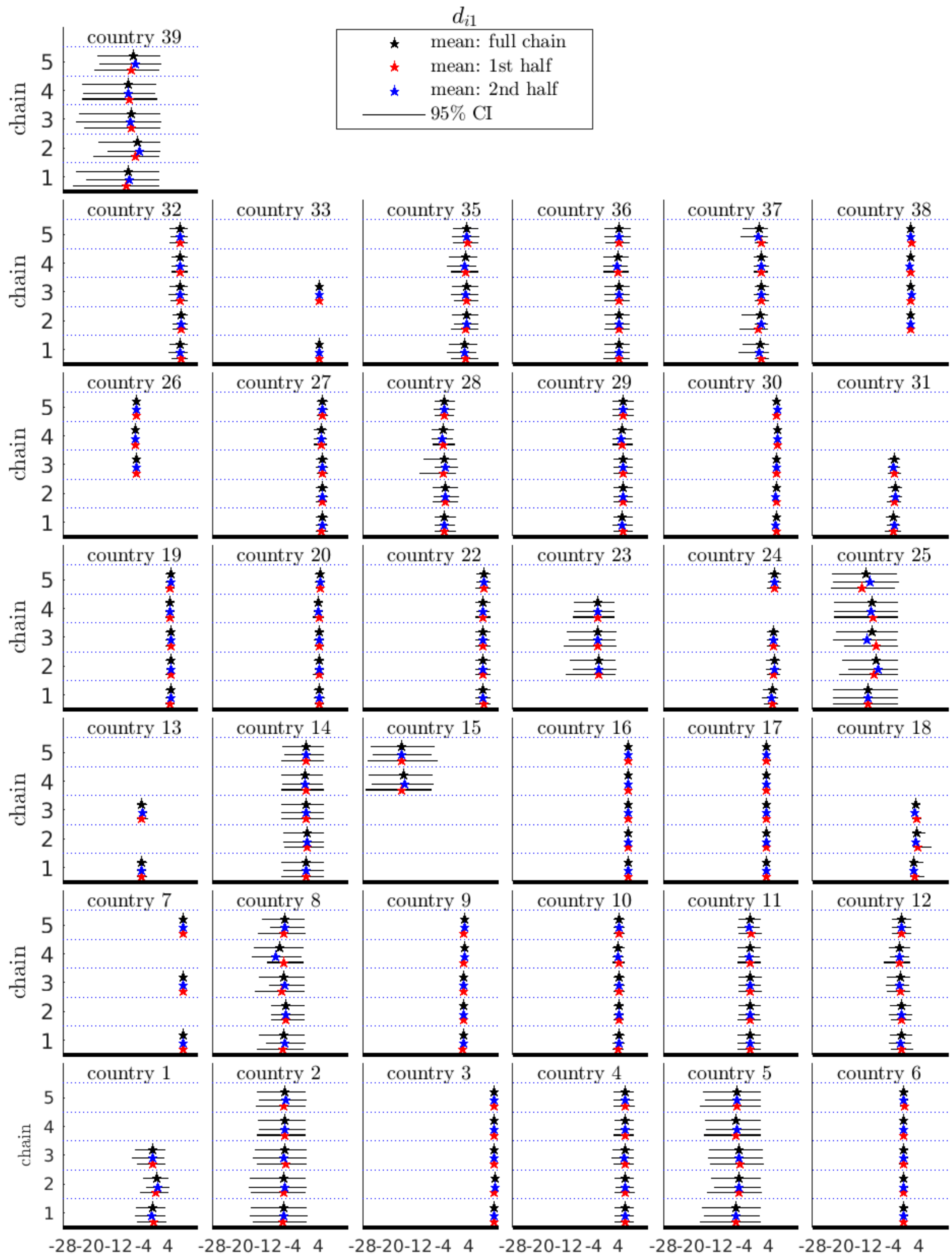


Figure L2: Summary plots for QBE chains: d_{i2}

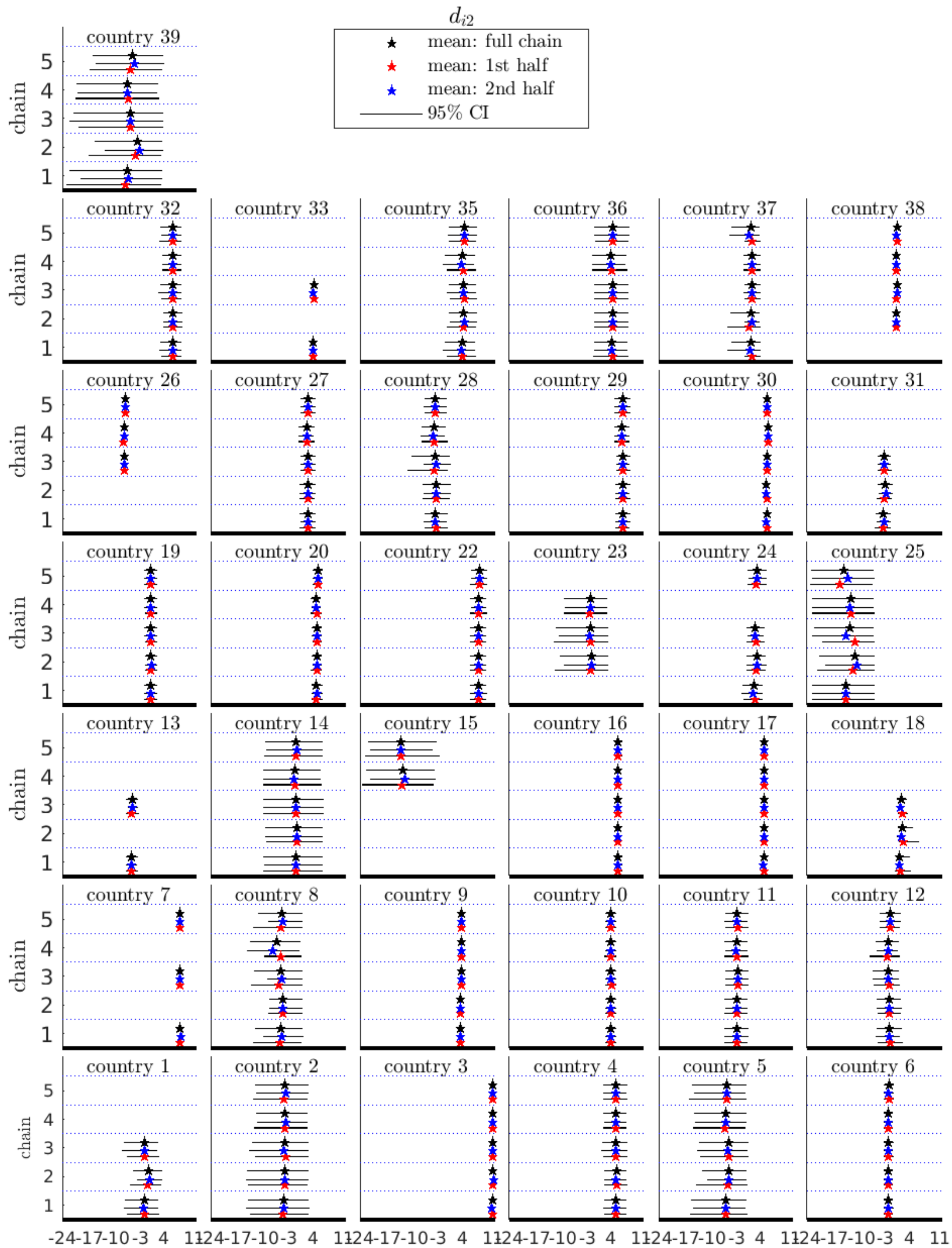


Figure L3: Summary plots for QBE chains: d_{i3}

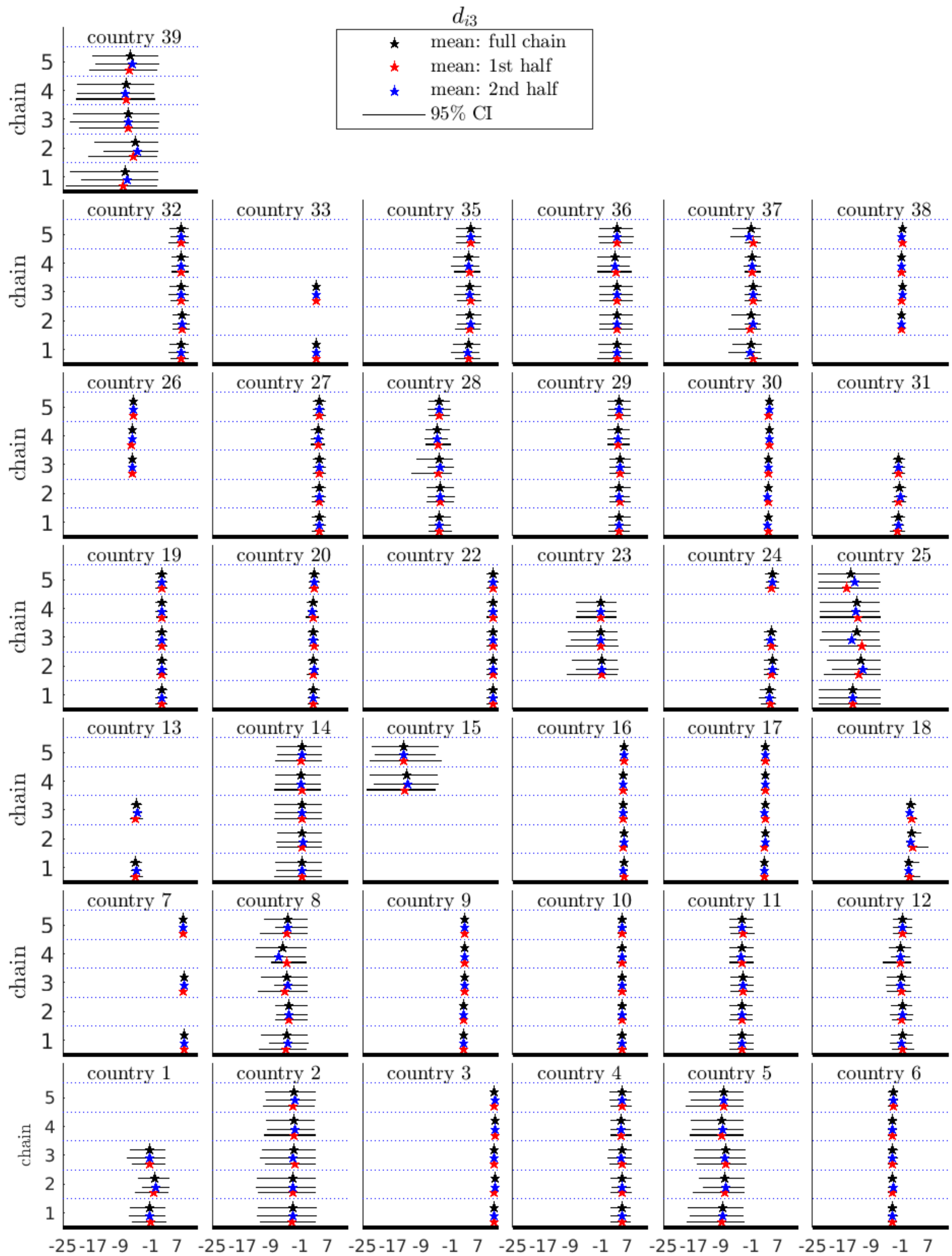


Figure L4: Summary plots for QBE chains: d_{i4}

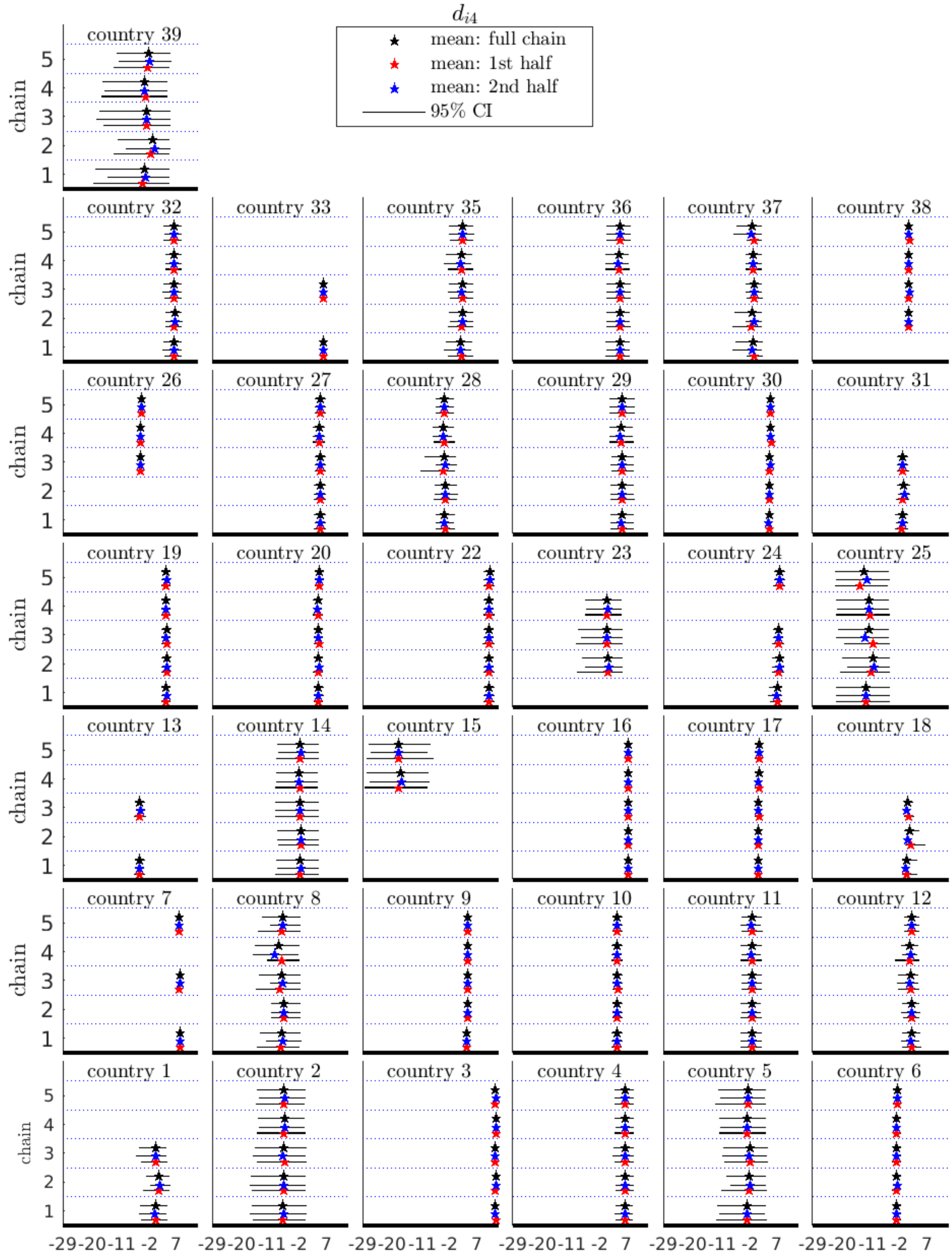


Figure L5: Summary plots for QBE chains: d_{i5}

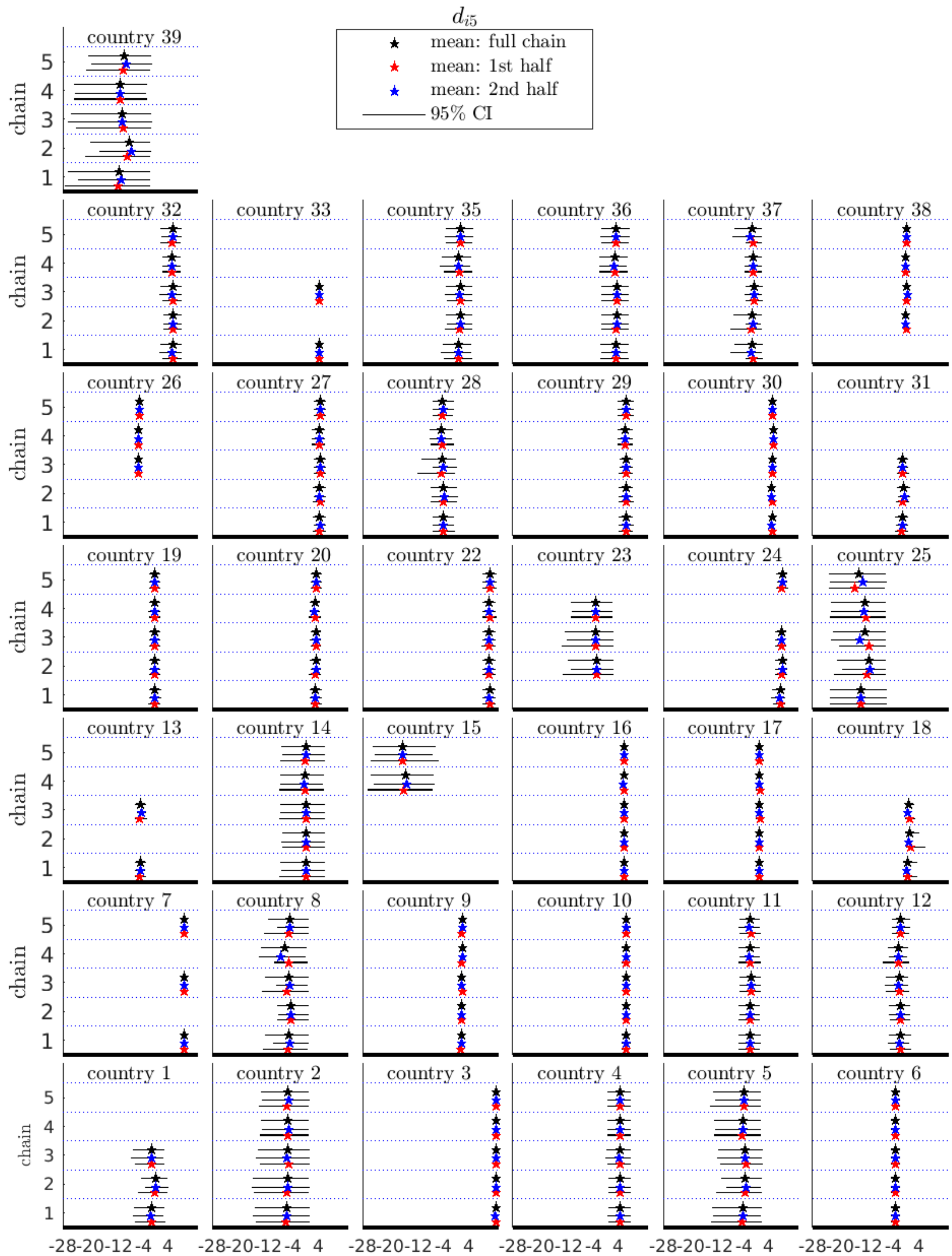


Figure L6: Summary plots for QBE chains: d_{i6}

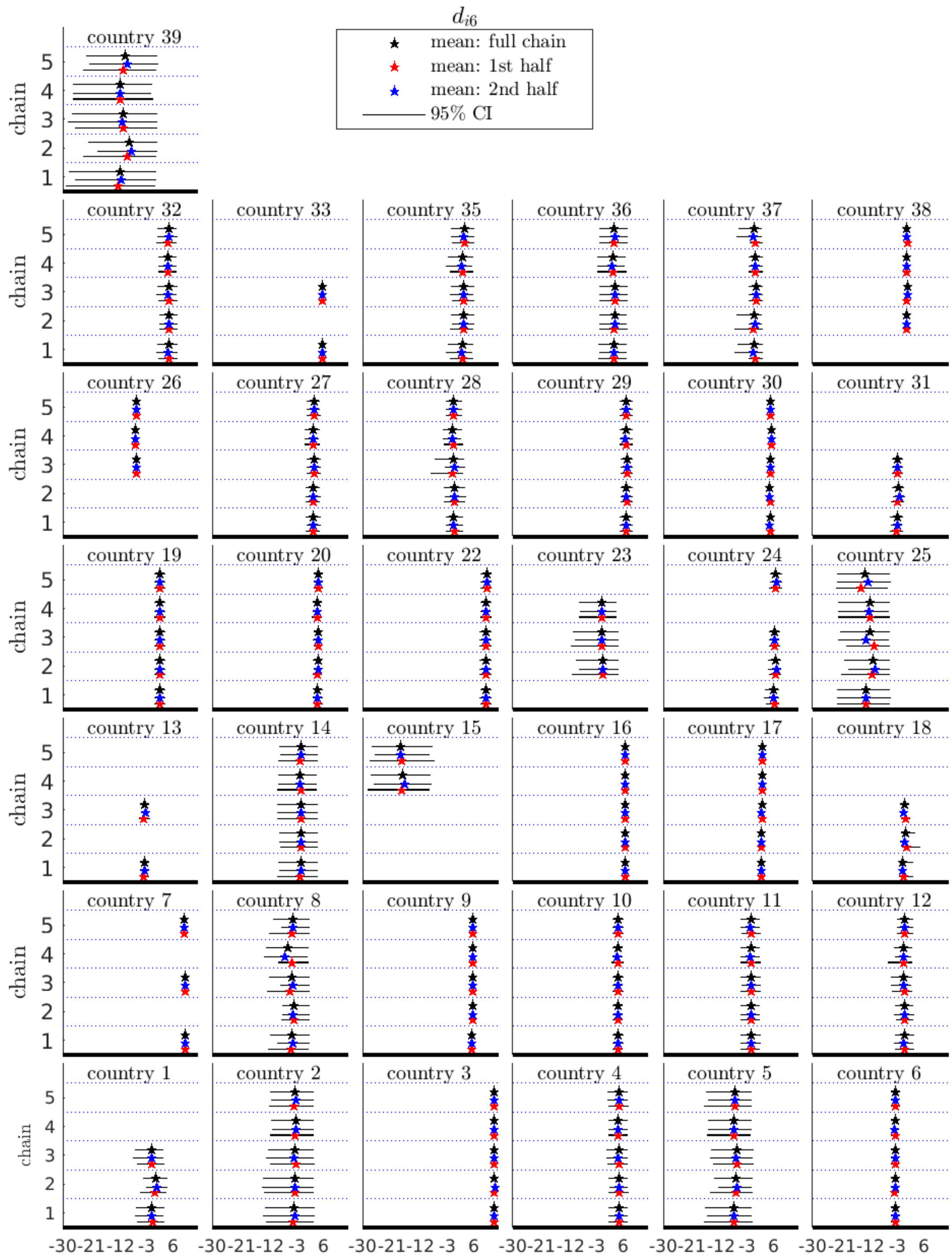


Figure L7: Summary plots for QBE chains: d_{i7}

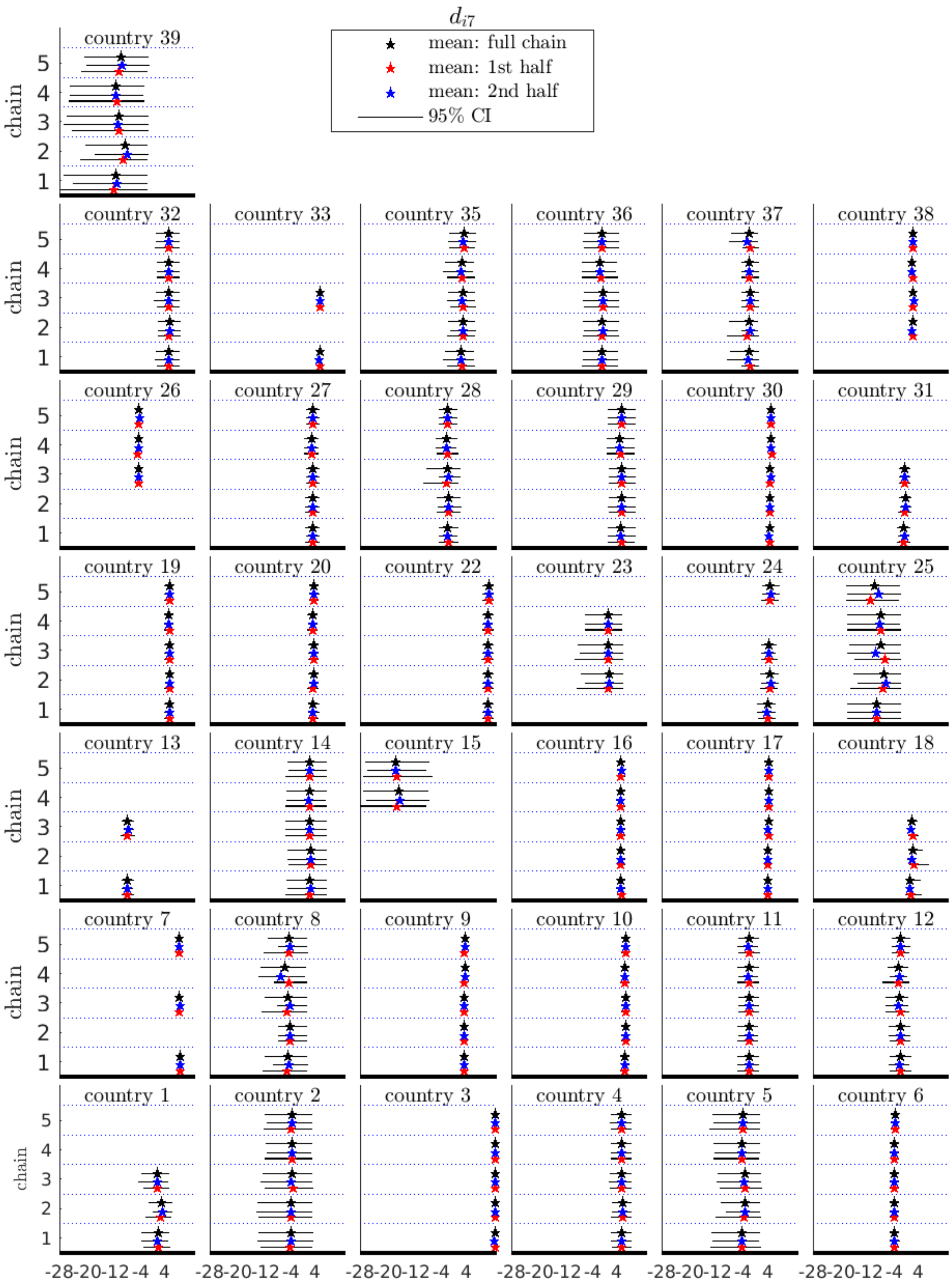


Figure L8: Summary plots for QBE chains: d_{i8}

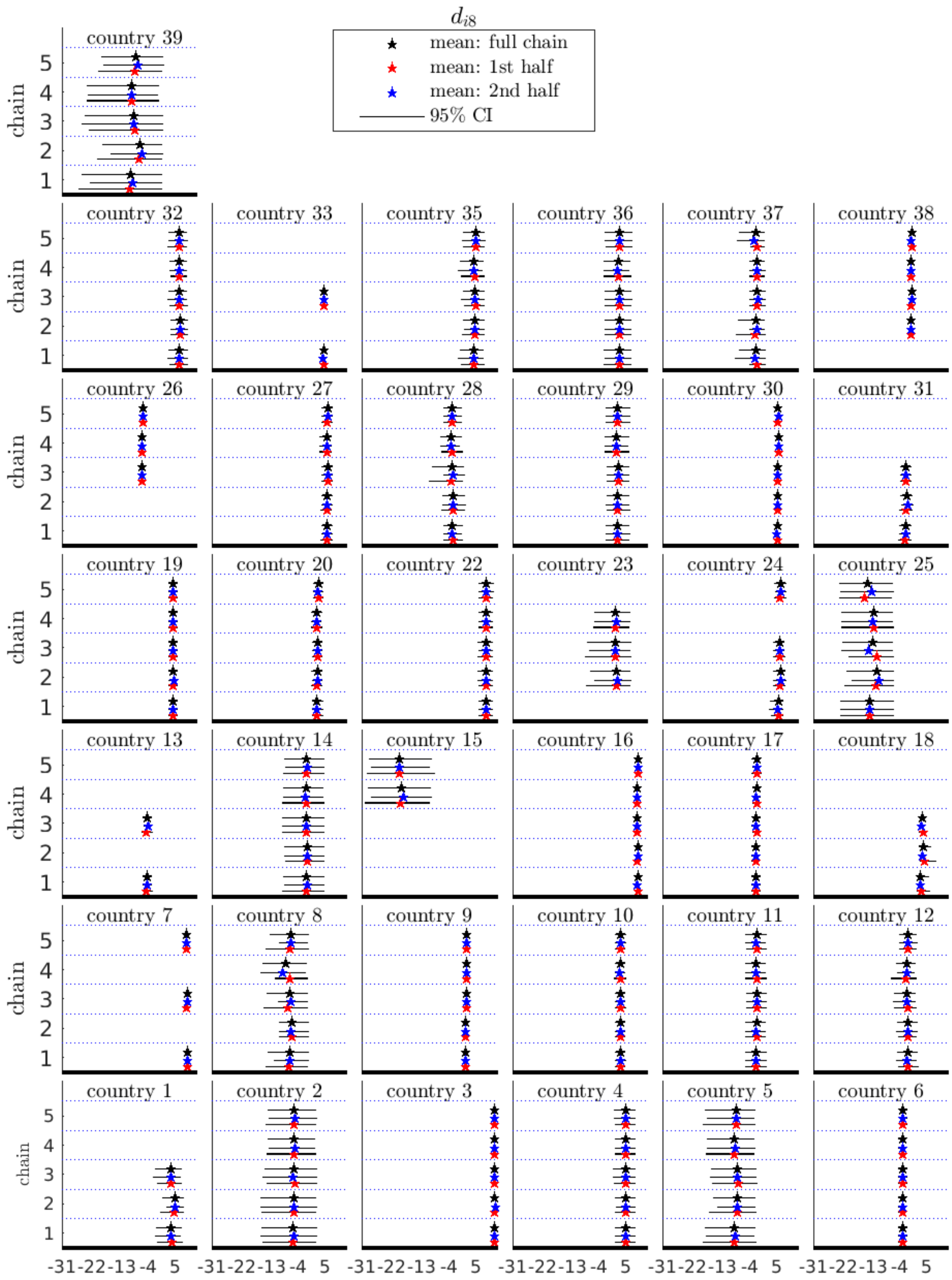


Figure L9: Summary plots for QBE chains: d_{i9}

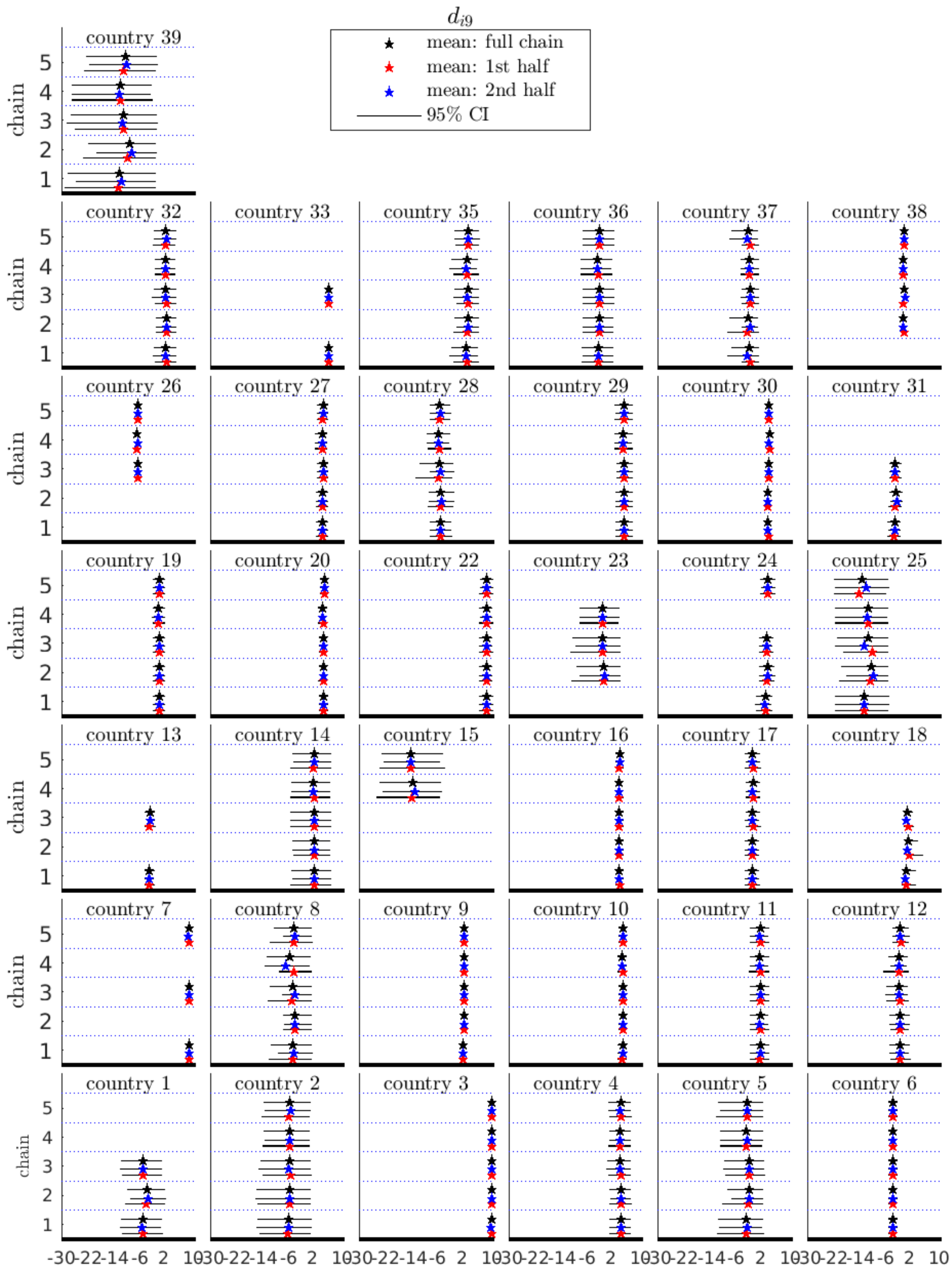


Figure L10: Summary plots for QBE chains: d_{i10}

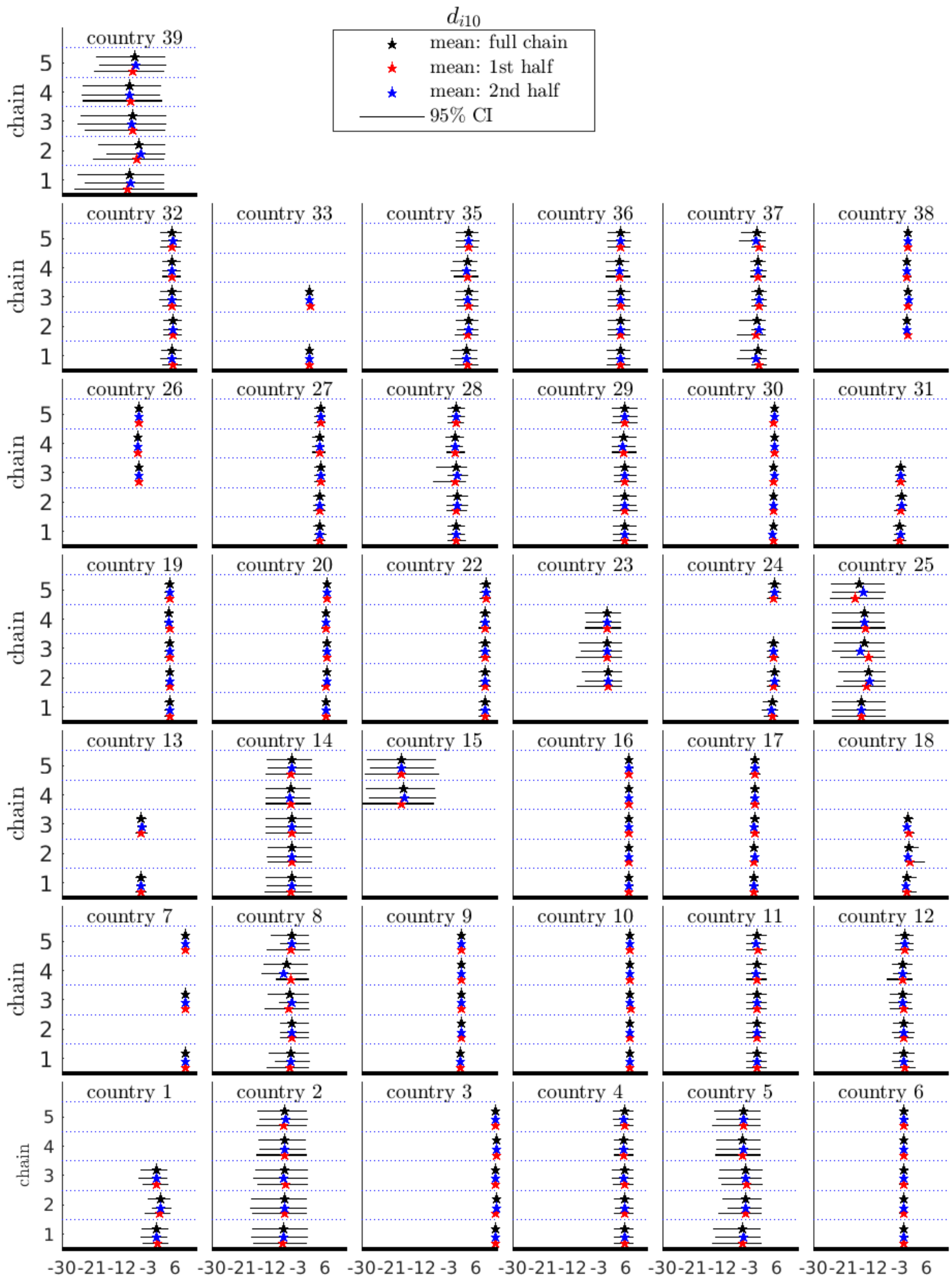


Figure L11: Summary plots for QBE chains: d_{i11}

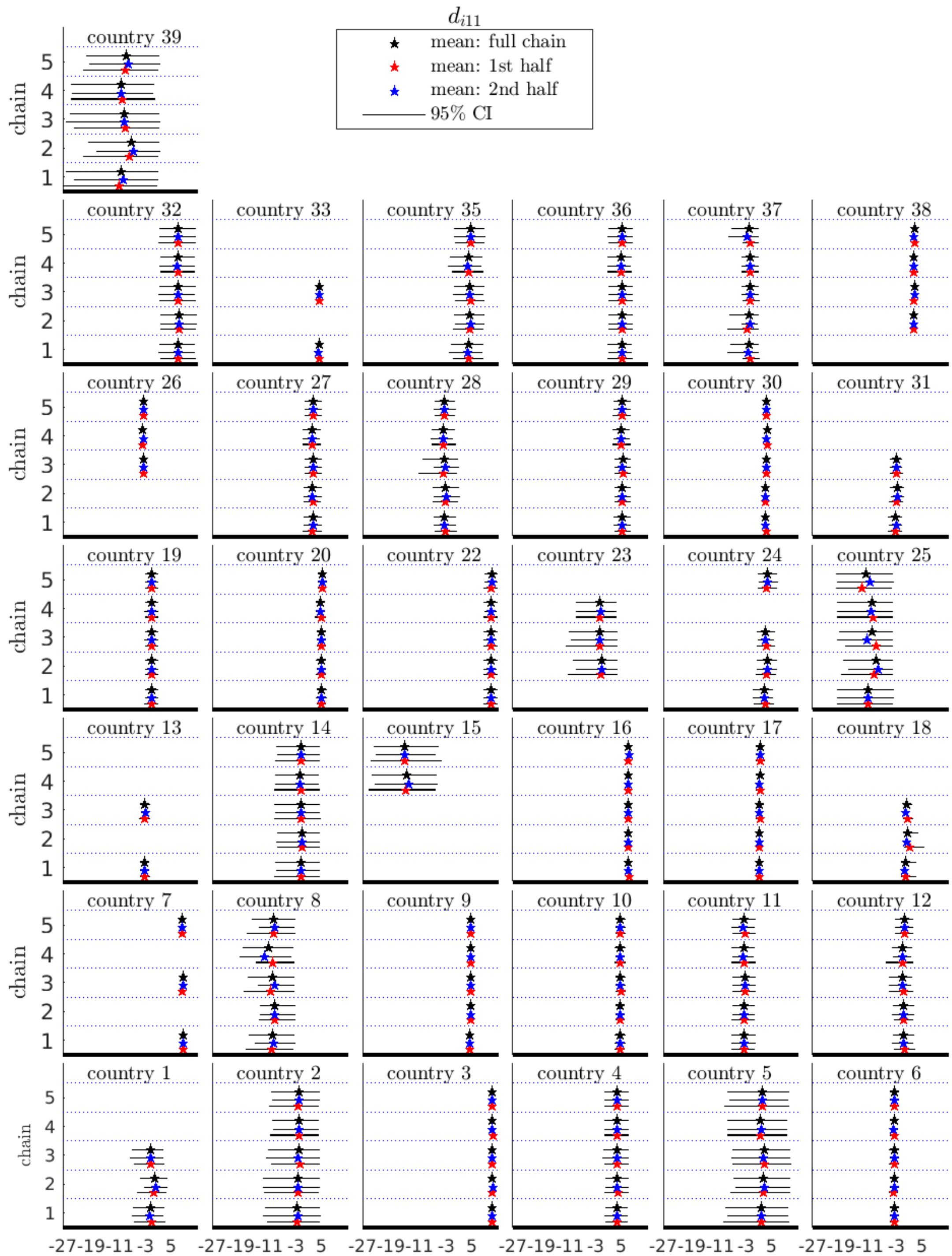


Figure L12: Summary plots for QBE chains: d_{i12}

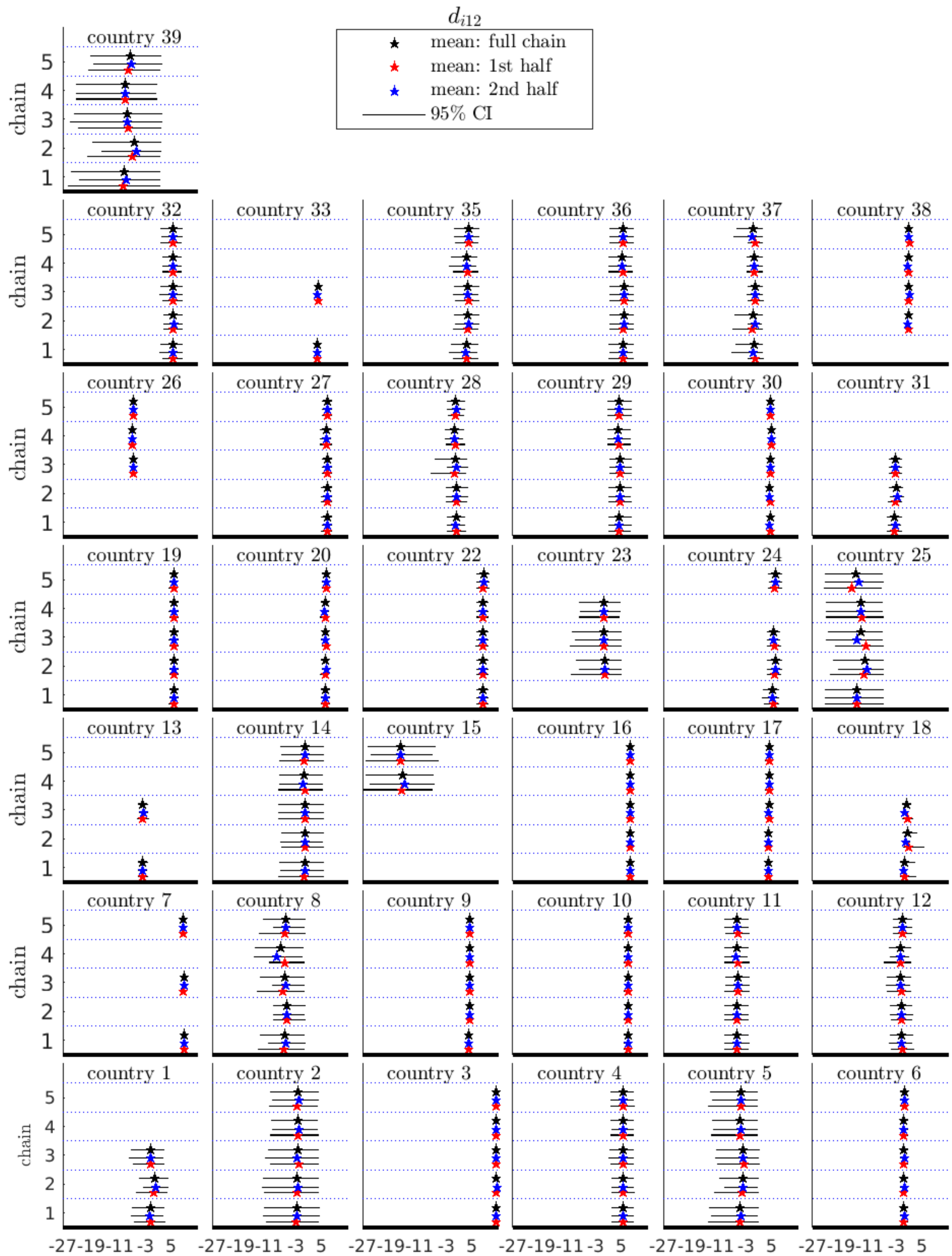


Figure L13: Summary plots for QBE chains: d_{i13}

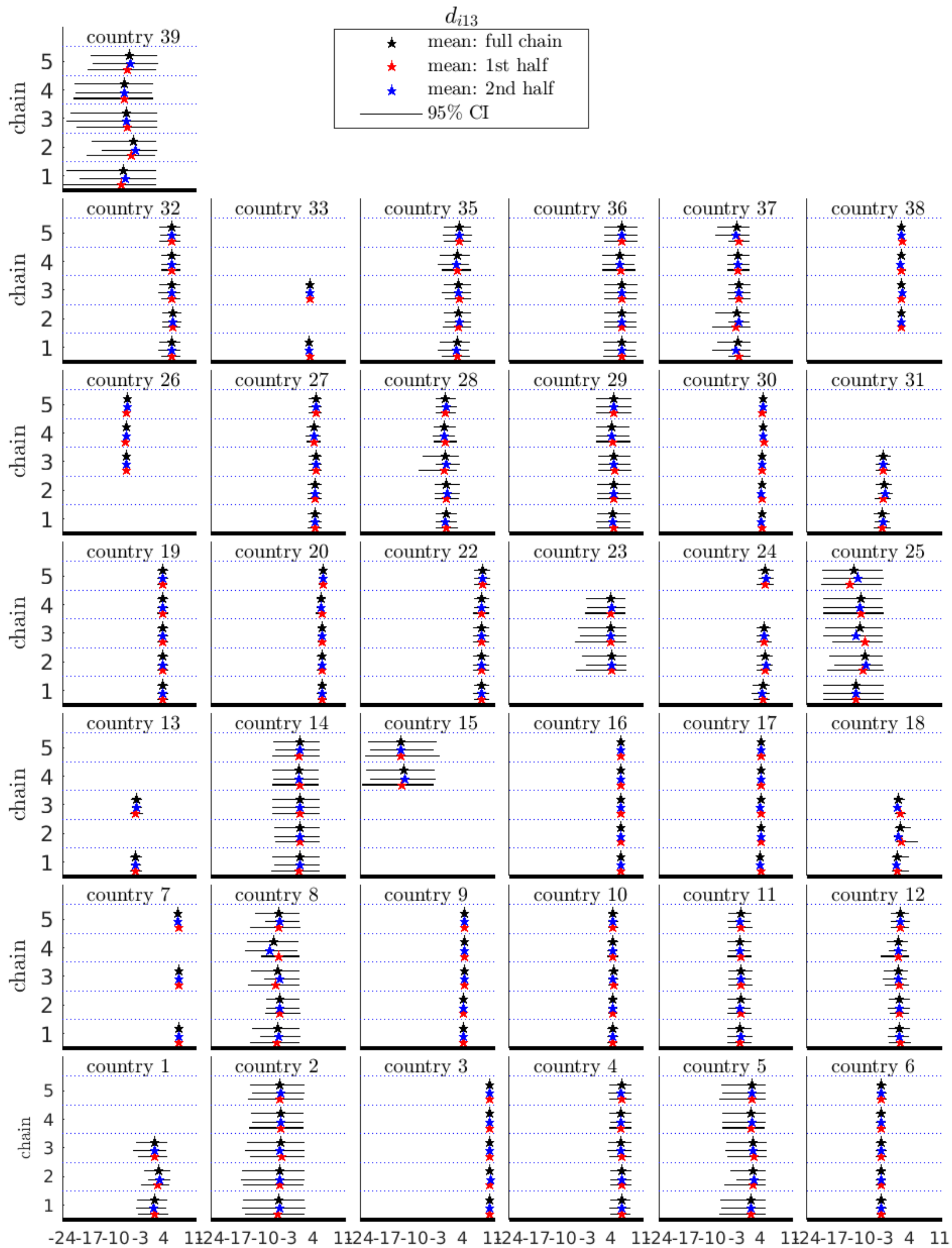


Figure L14: Summary plots for QBE chains: d_{i14}

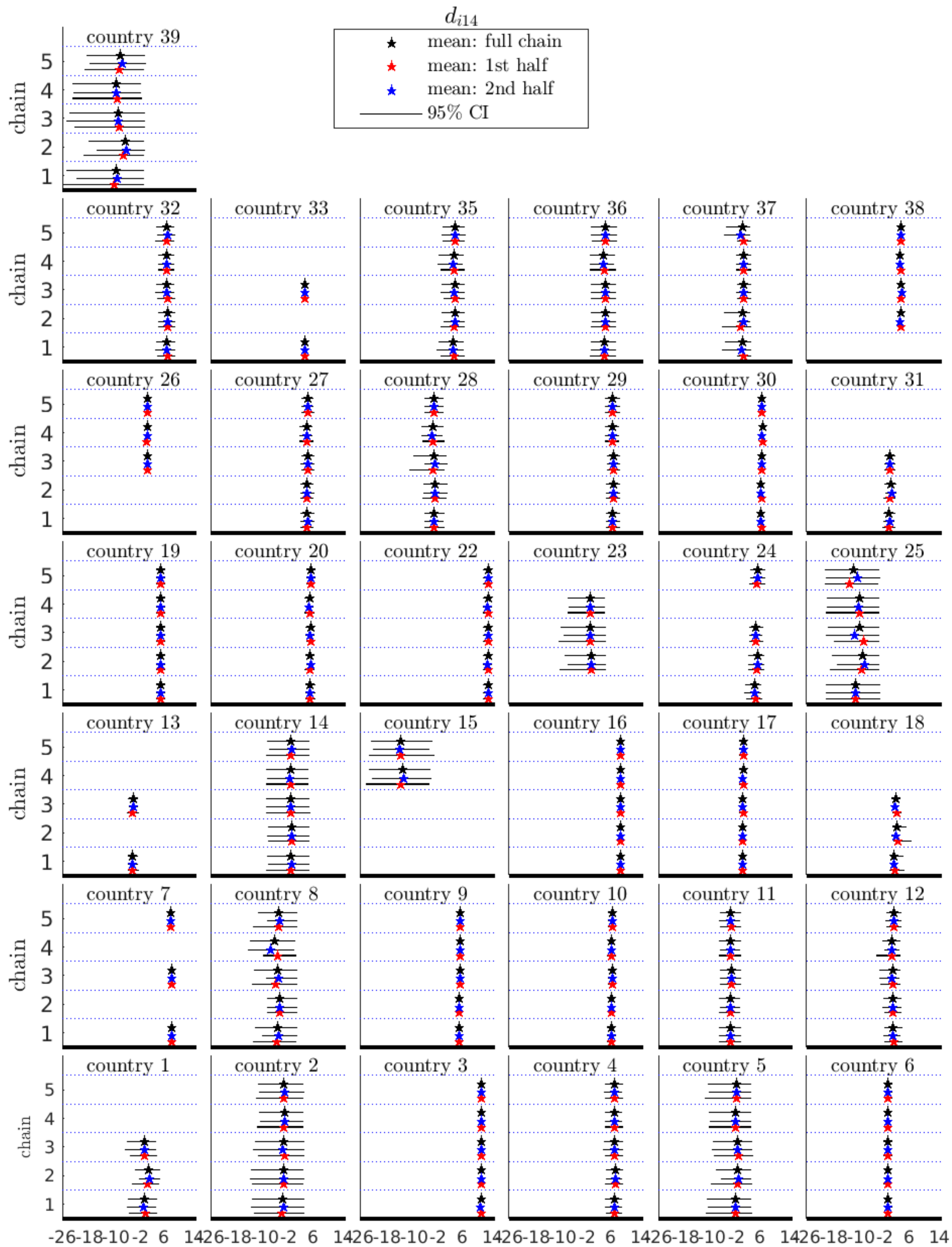


Figure L15: Summary plots for QBE chains: d_{i15}

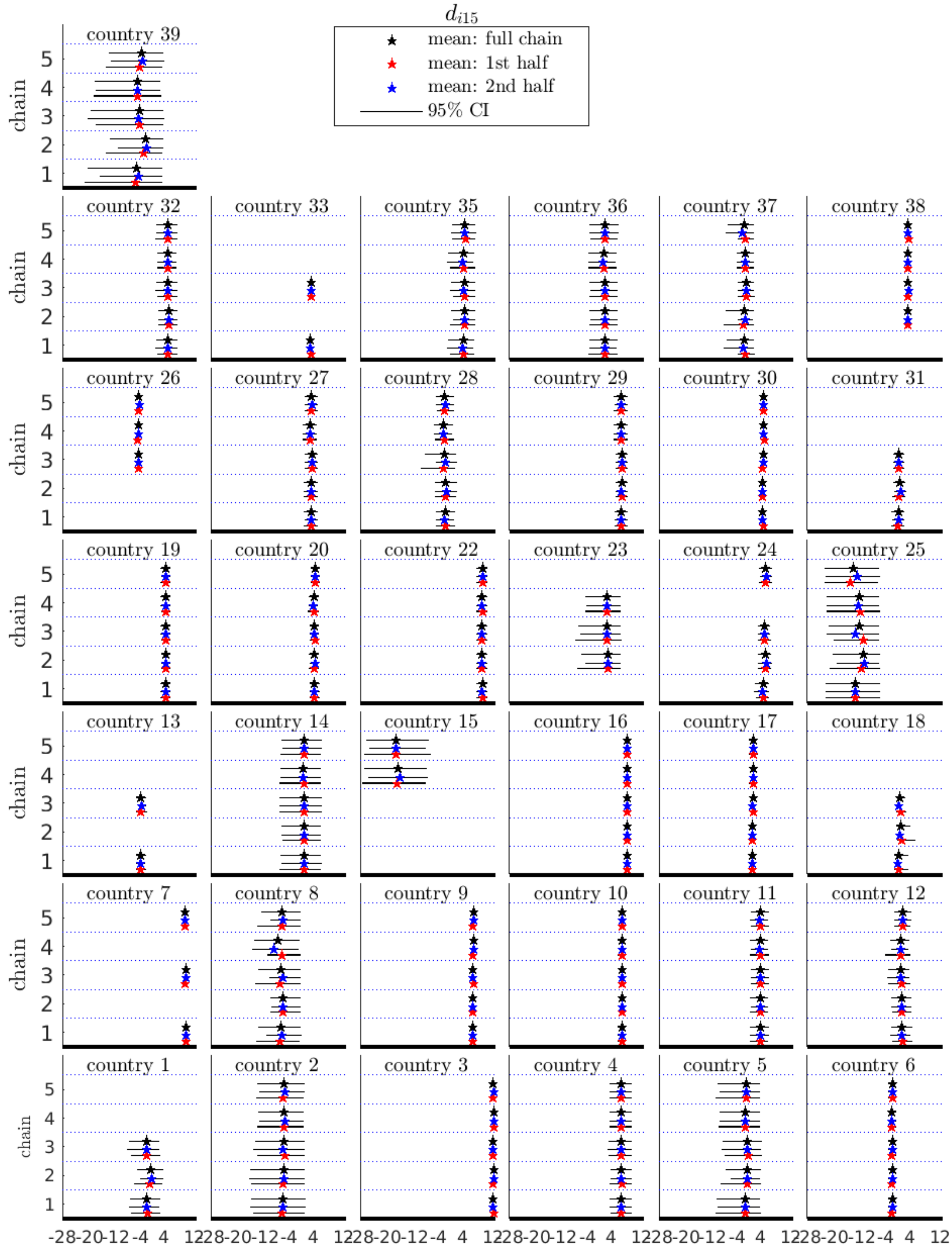


Figure L16: Summary plots for QBE chains: $\mu_{f,i1}$

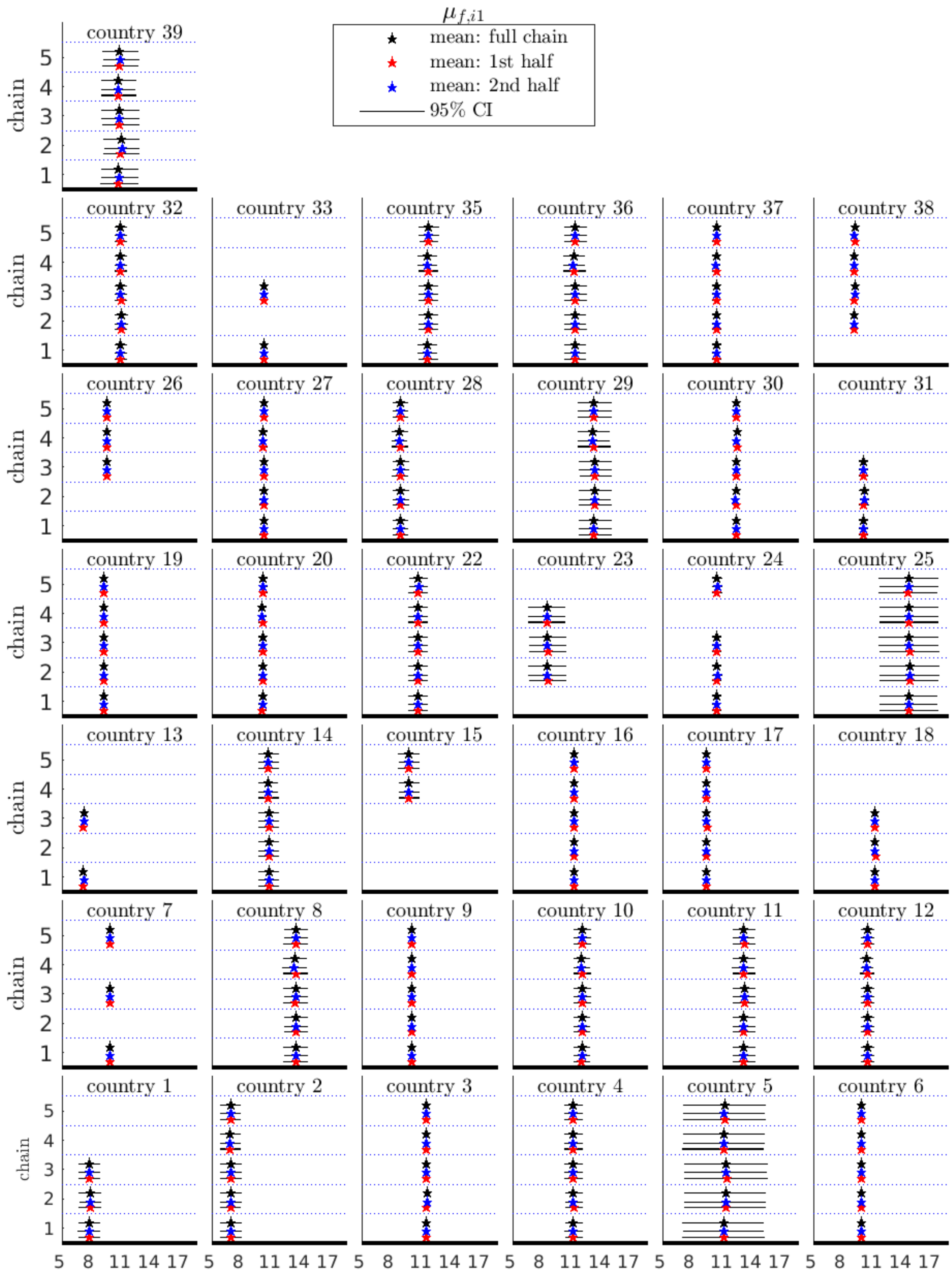


Figure L17: Summary plots for QBE chains: $\mu_{f,i2}$

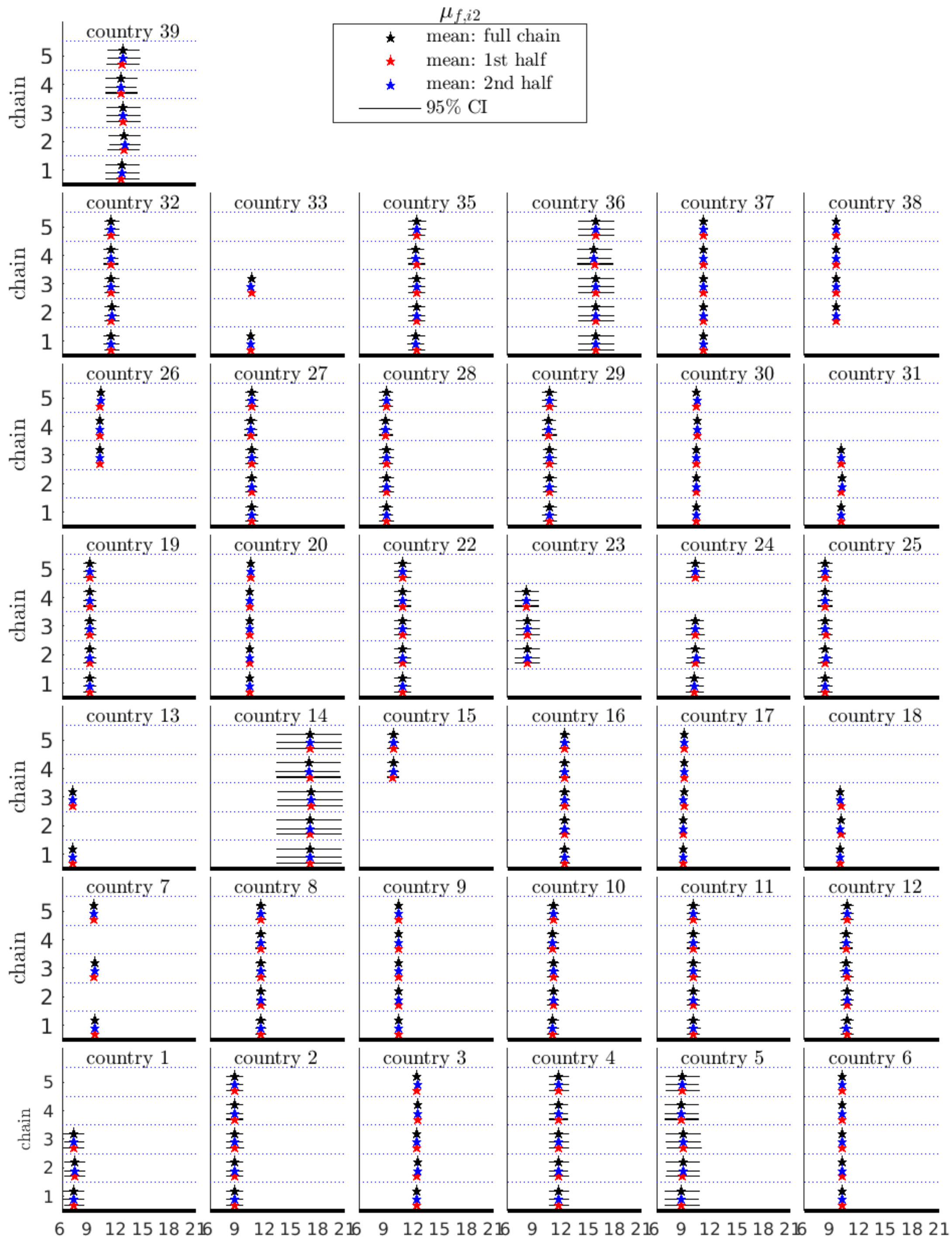


Figure L18: Summary plots for QBE chains: $\mu_{f,i3}$

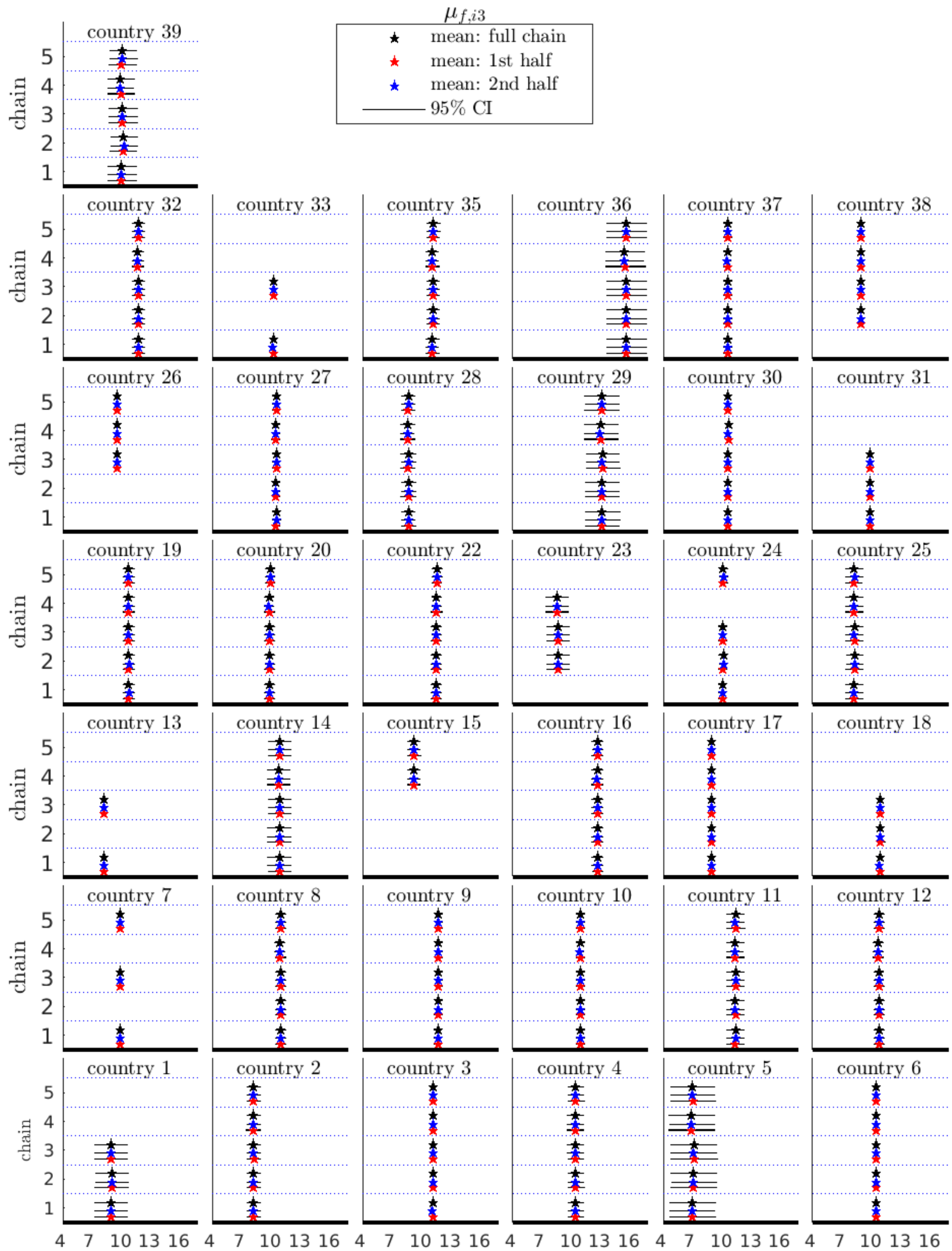


Figure L19: Summary plots for QBE chains: $\mu_{f,i4}$

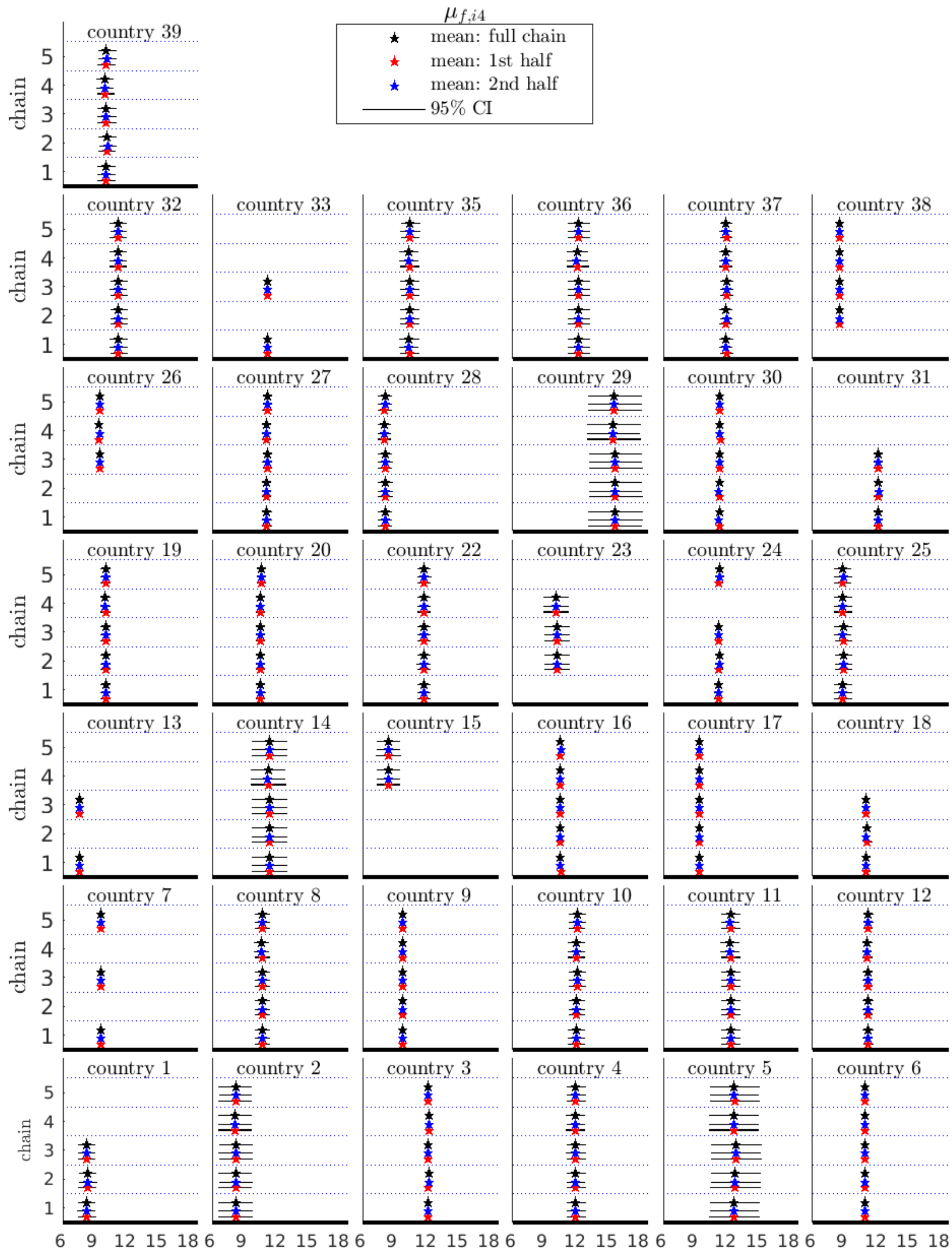


Figure L20: Summary plots for QBE chains: $\mu_{f,i5}$

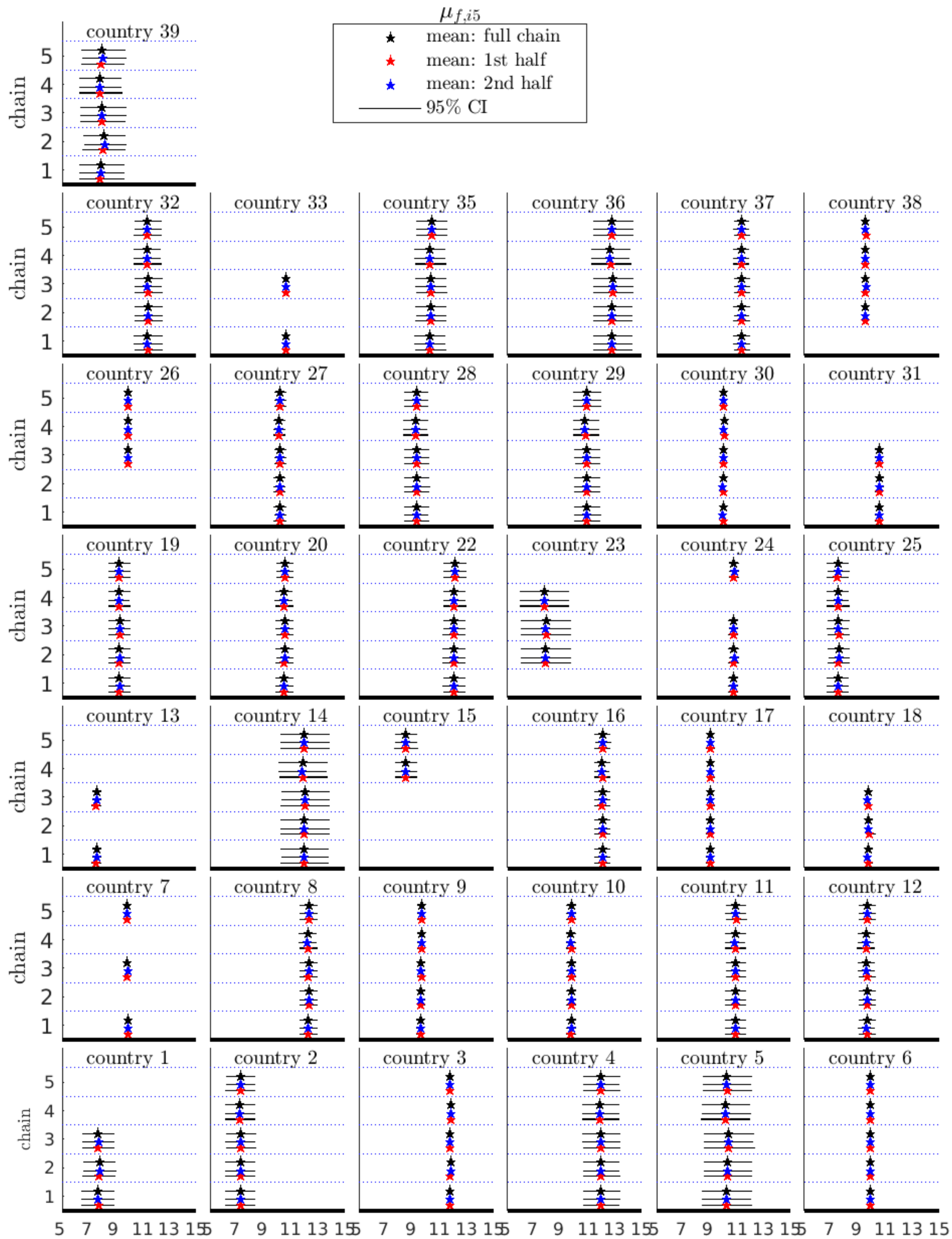


Figure L21: Summary plots for QBE chains: $\mu_{f,i6}$

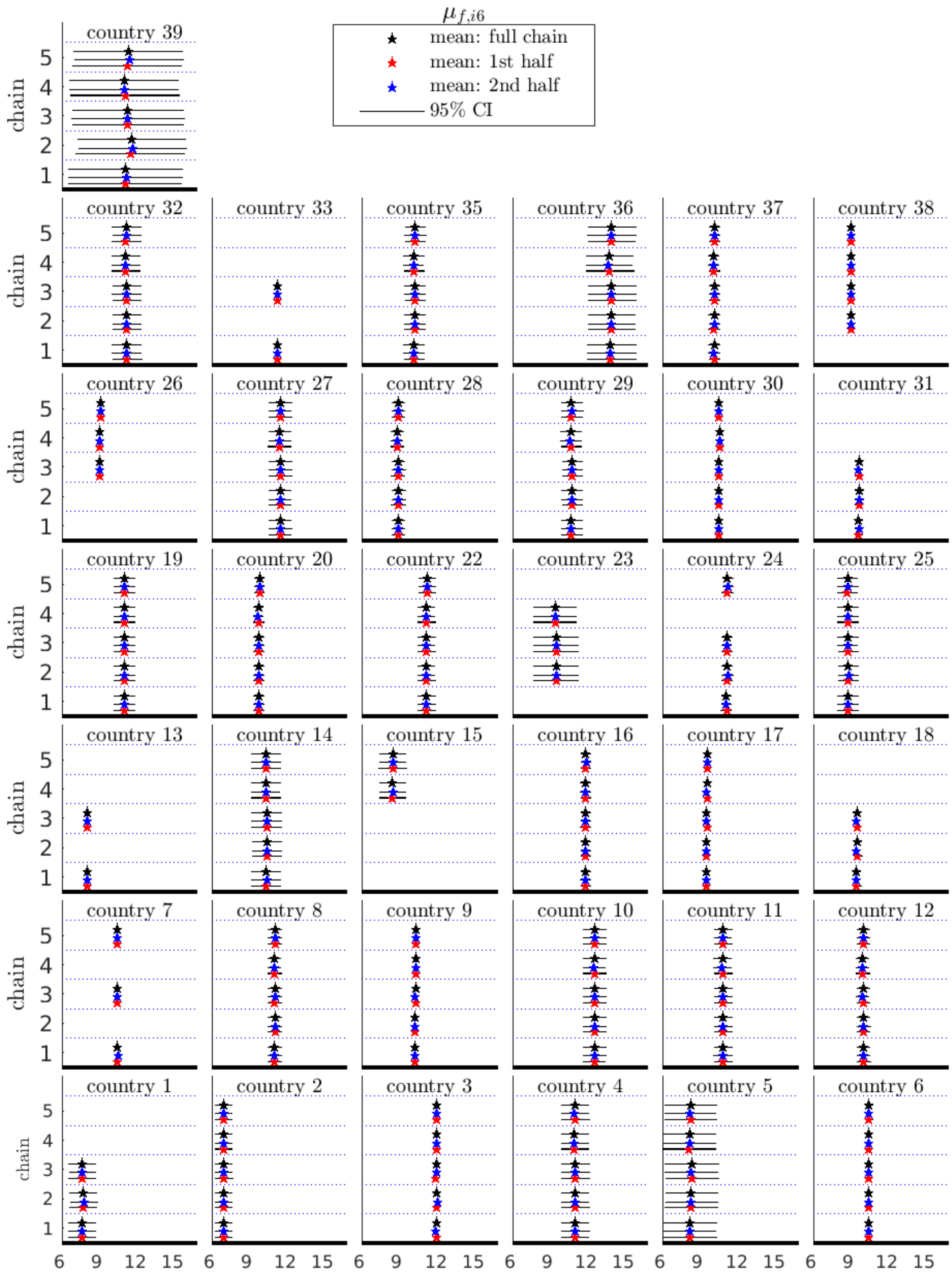


Figure L22: Summary plots for QBE chains: $\mu_{f,i7}$

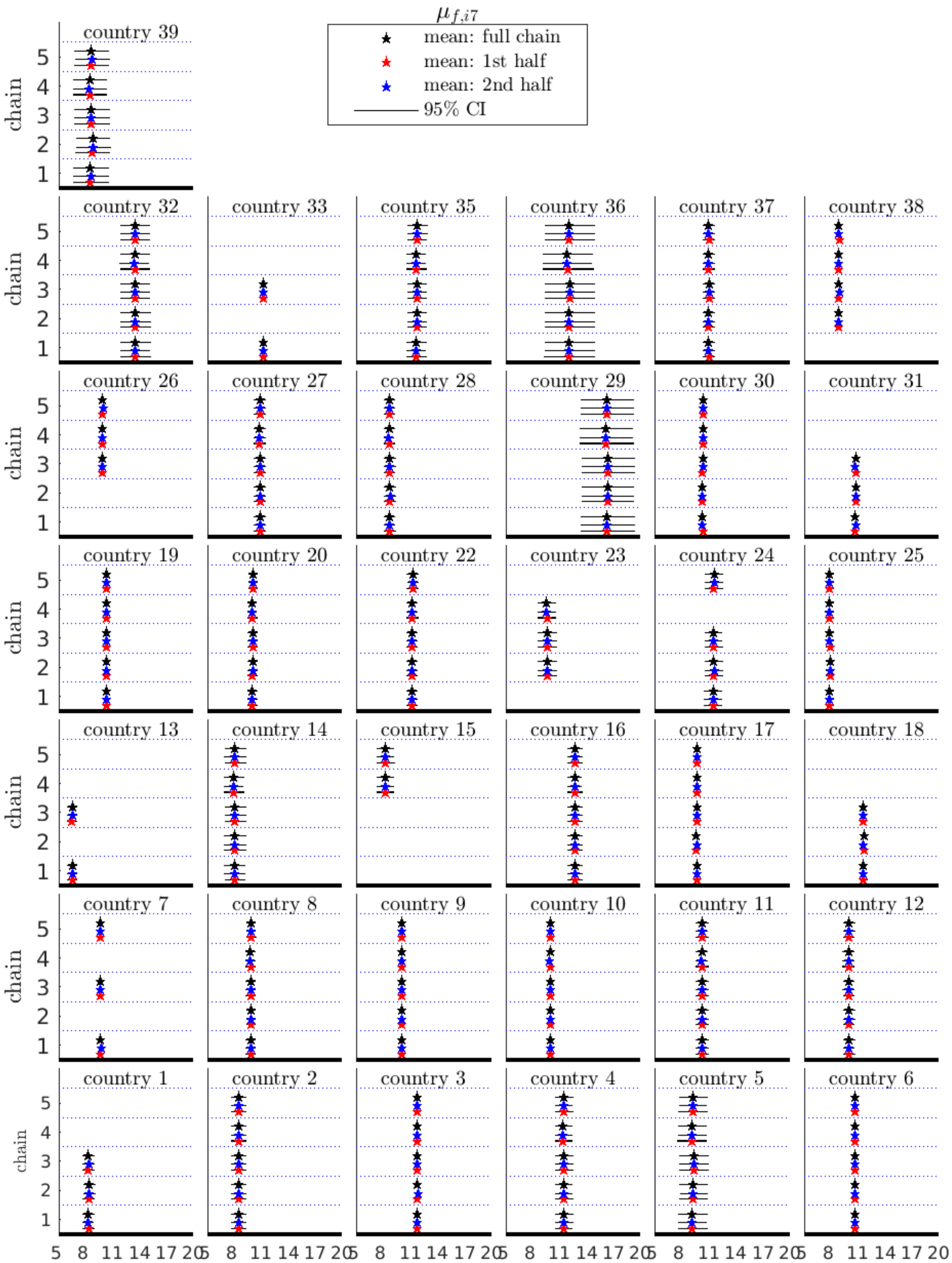


Figure L23: Summary plots for QBE chains: $\mu_{f,i8}$

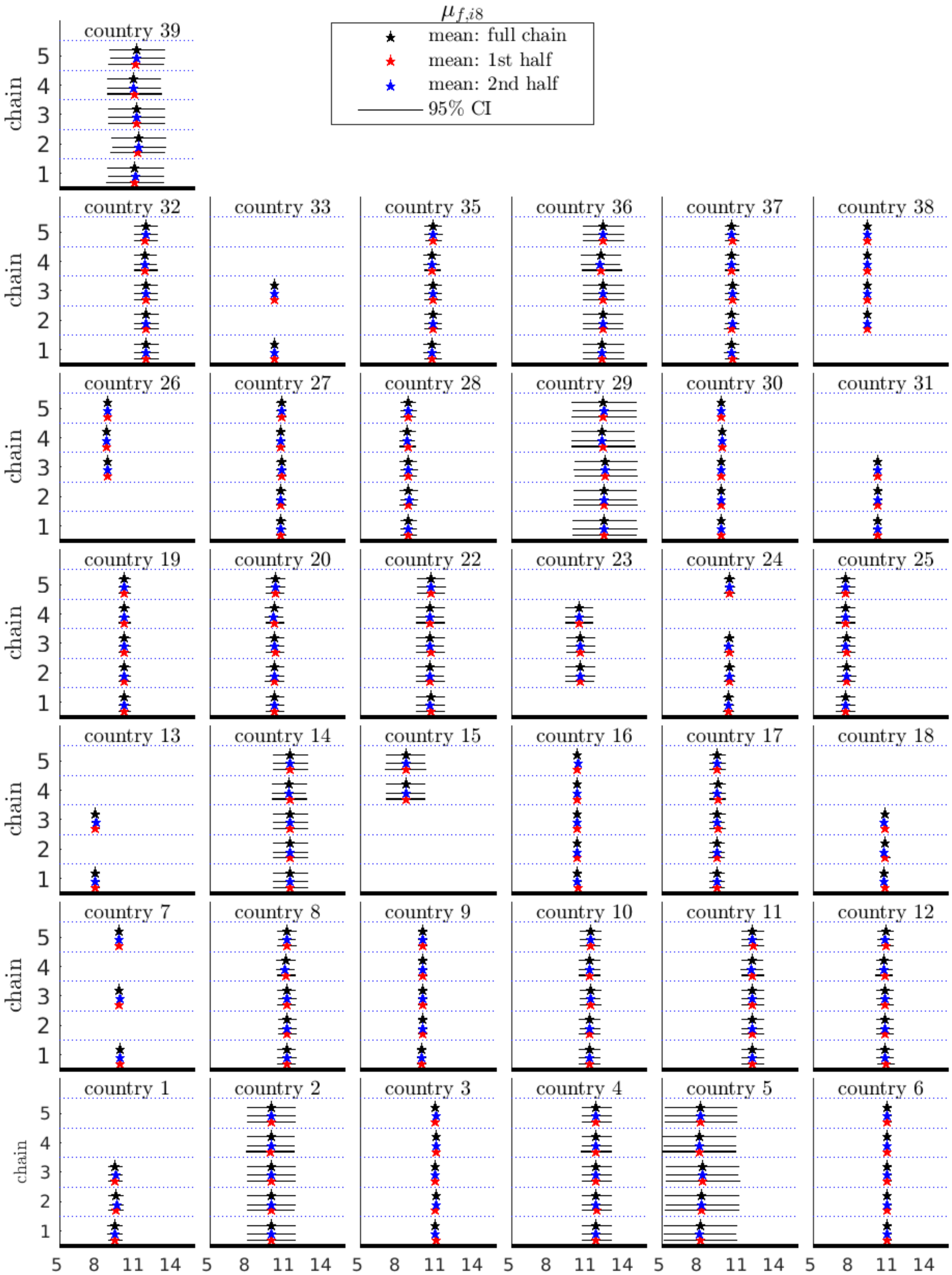


Figure L24: Summary plots for QBE chains: $\mu_{f,i9}$

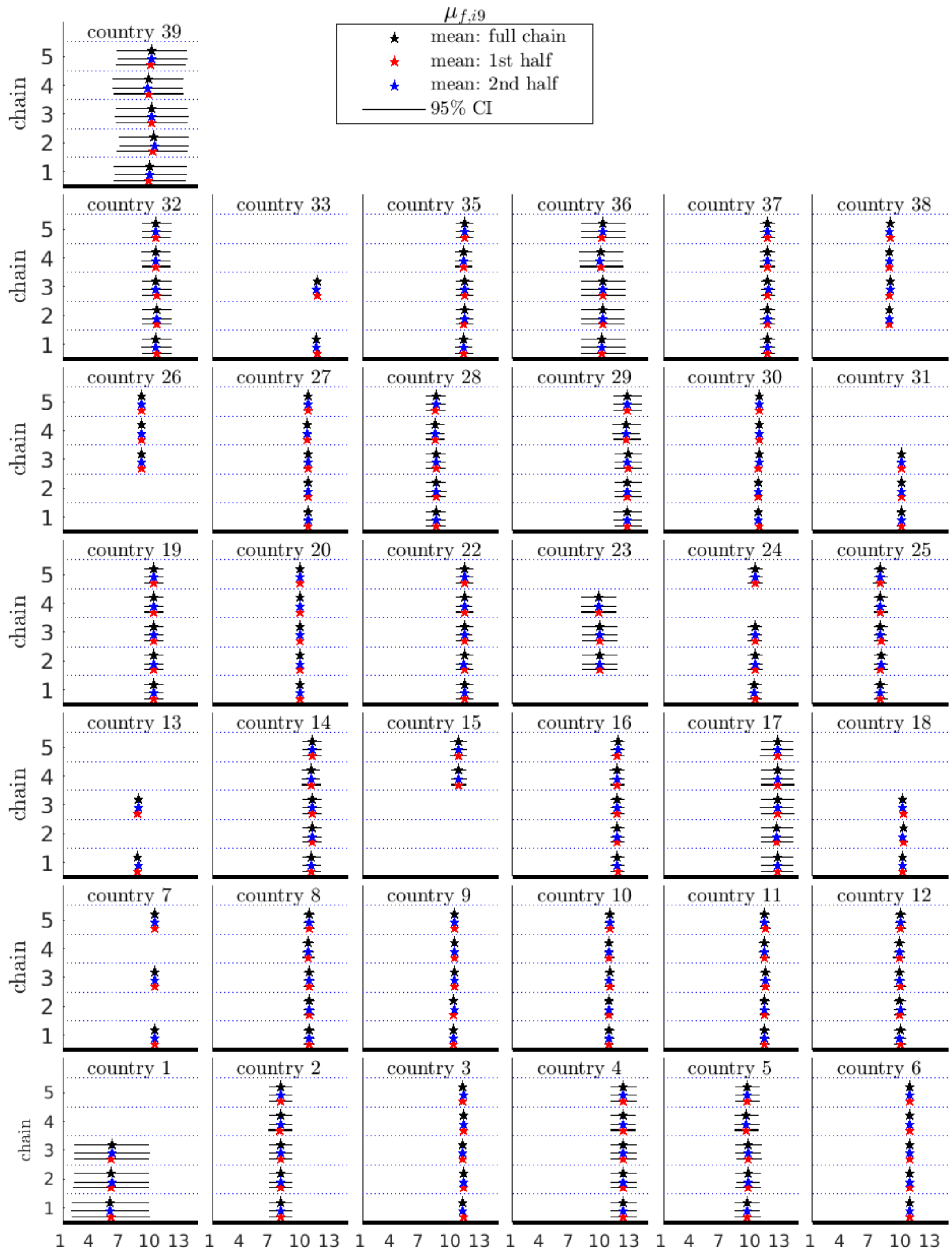


Figure L25: Summary plots for QBE chains: $\mu_{f,i10}$

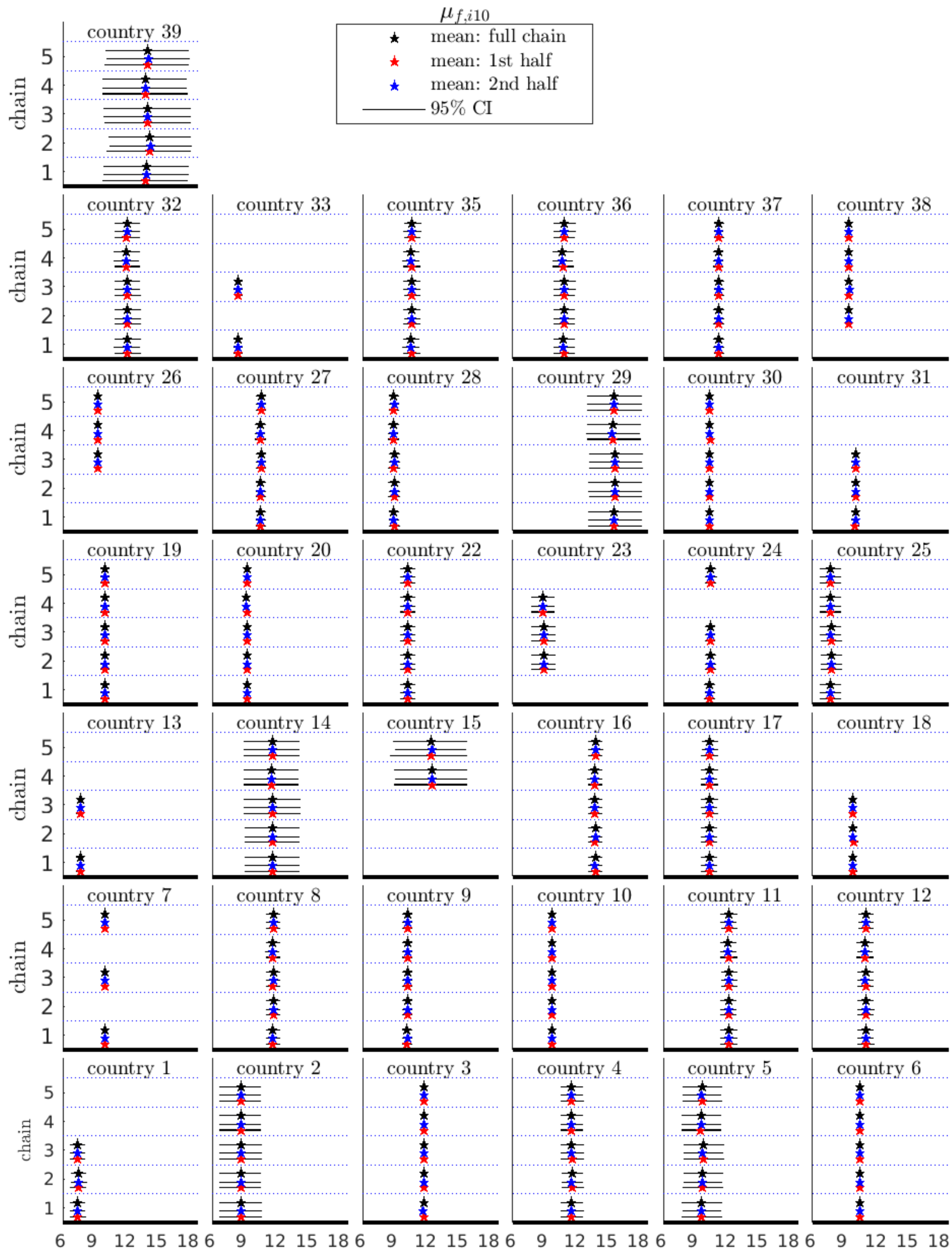


Figure L26: Summary plots for QBE chains: $\mu_{f,i11}$

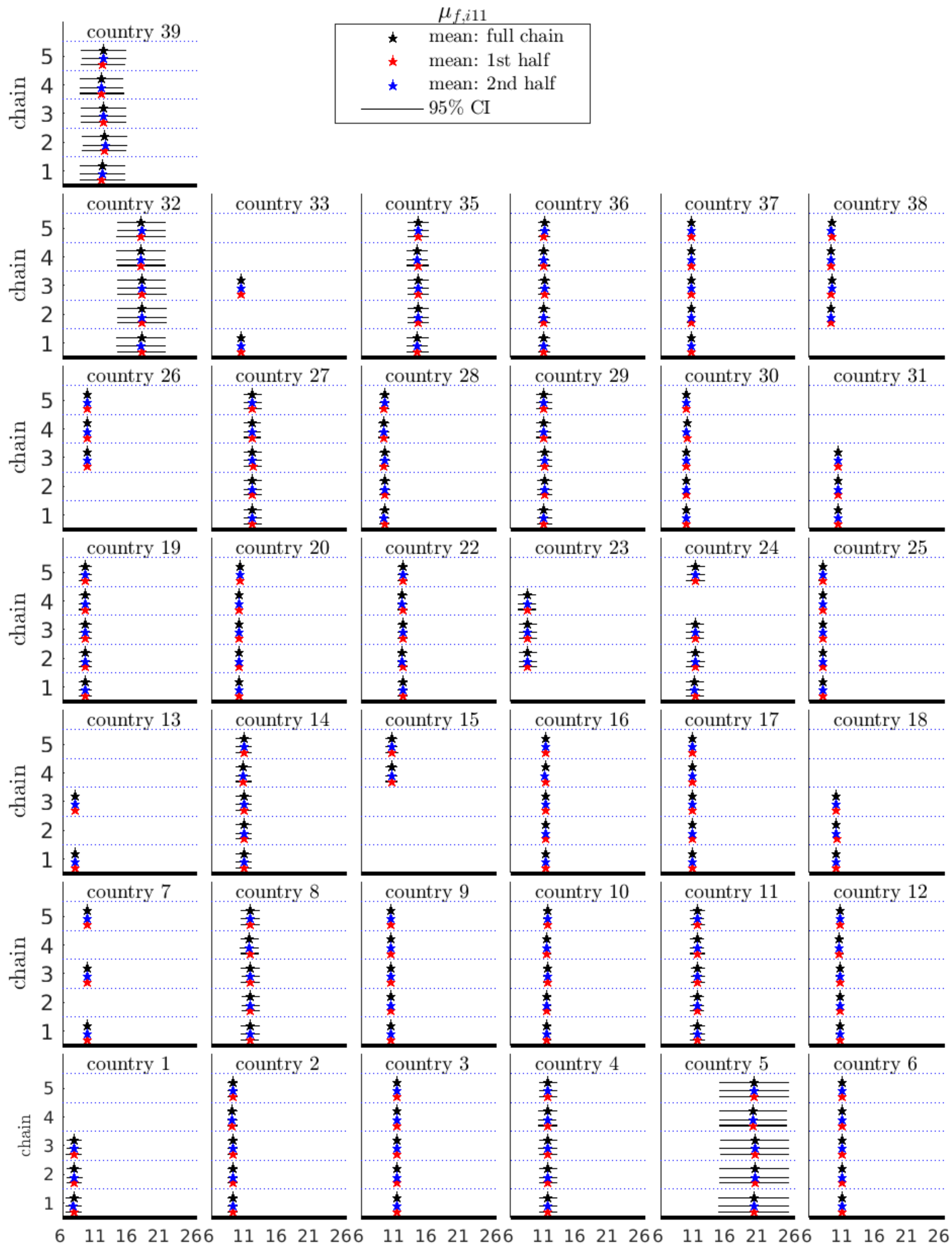


Figure L27: Summary plots for QBE chains: $\mu_{f,i12}$

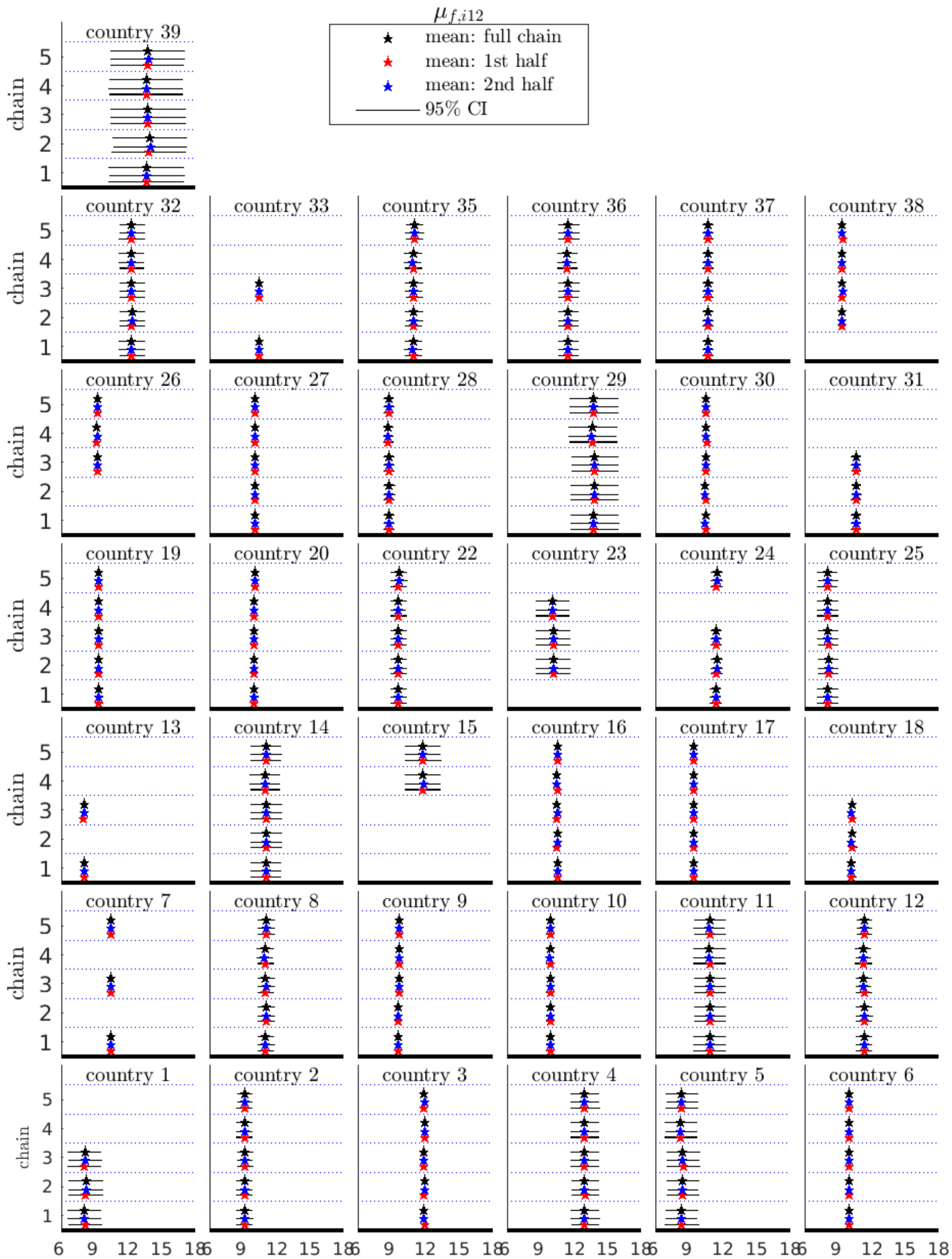


Figure L28: Summary plots for QBE chains: $\mu_{f,i13}$

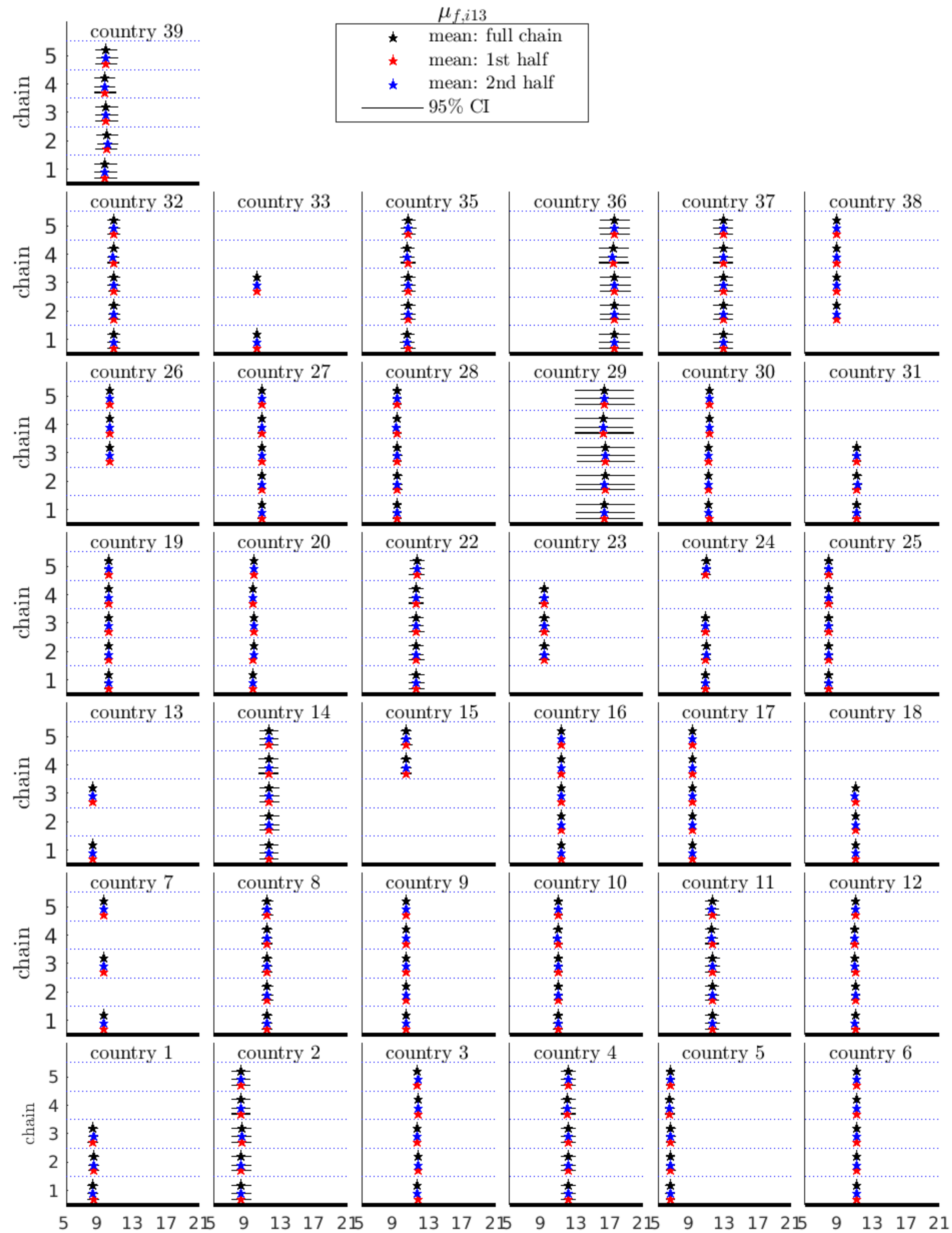


Figure L29: Summary plots for QBE chains: $\mu_{f,i14}$

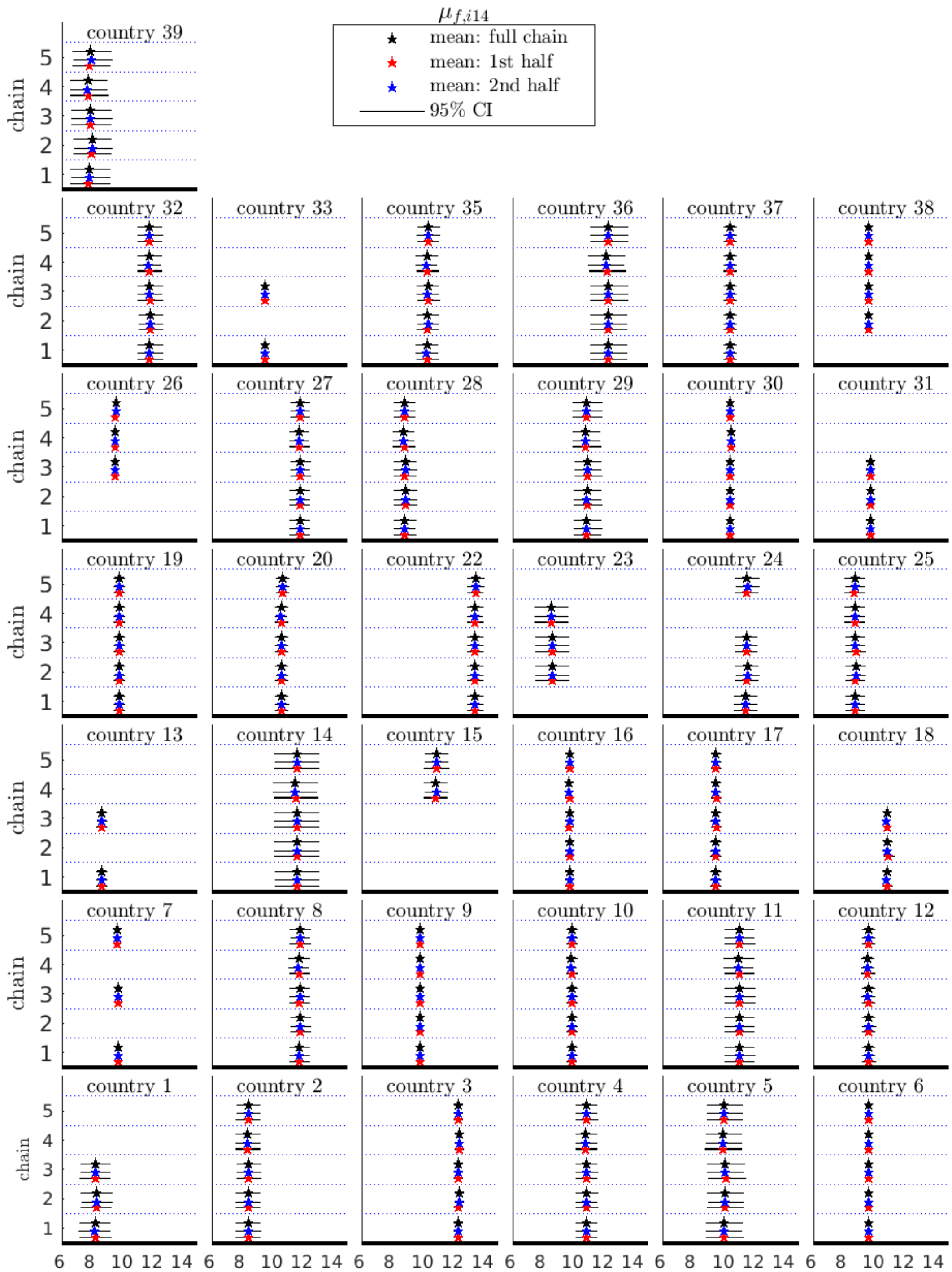


Figure L30: Summary plots for QBE chains: $\mu_{f,i15}$

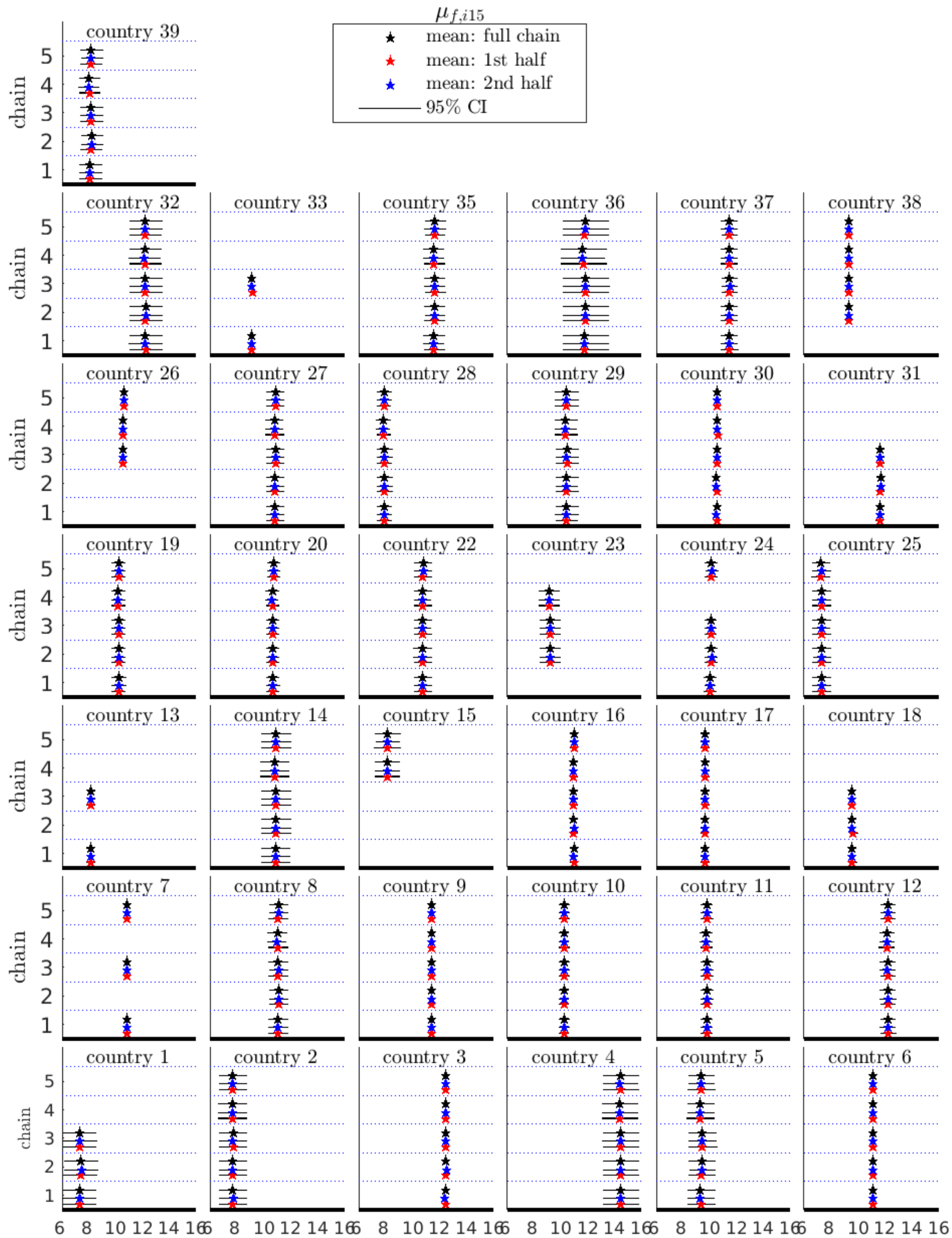


Figure L31: Summary plots for QBE chains: $\bar{\sigma}_{\varphi,i}$

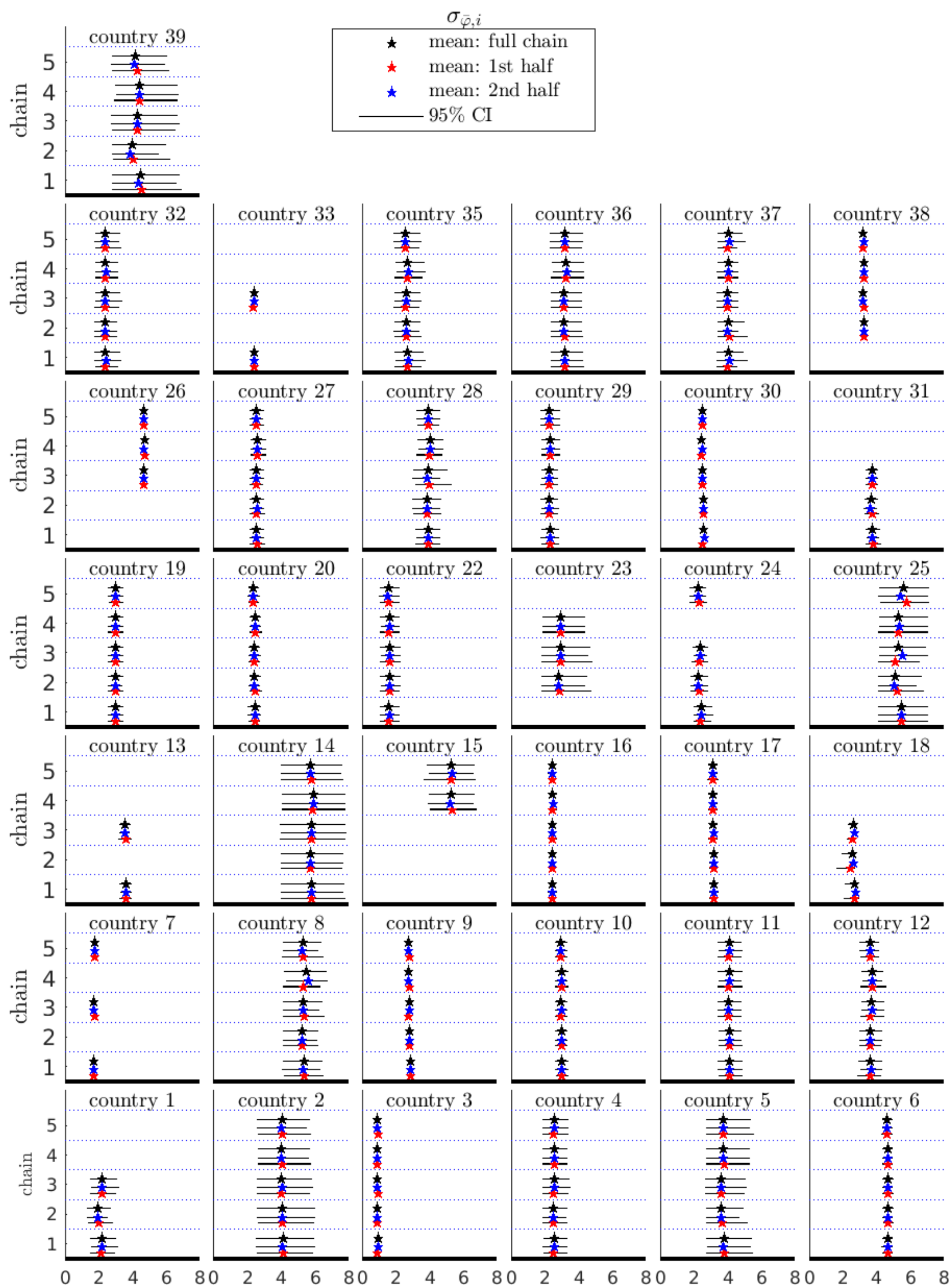


Figure L32: Summary plots for QBE chains: $\sigma_{\alpha,i}$

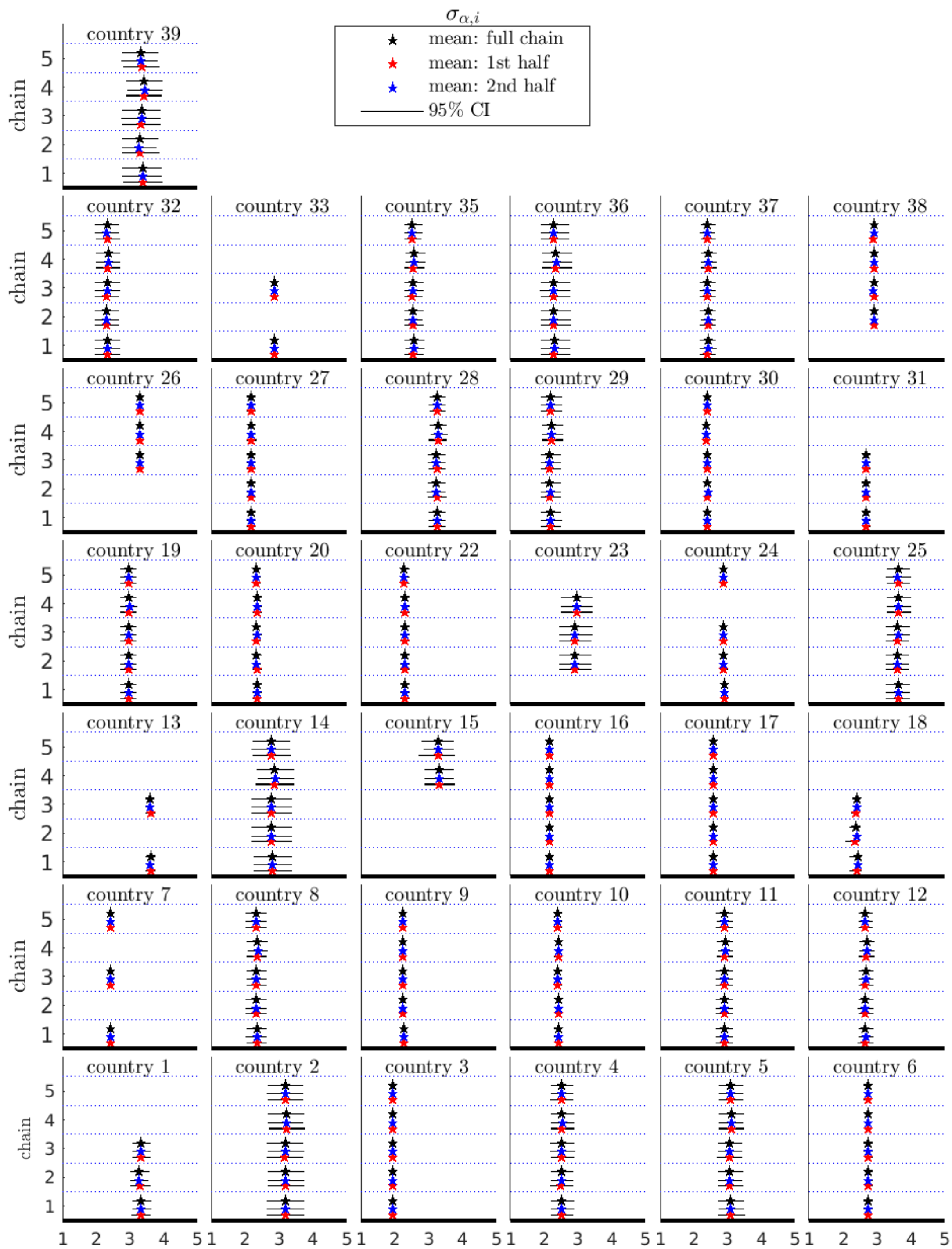


Figure L33: Summary plots for QBE chains: $\sigma_{f,i}$

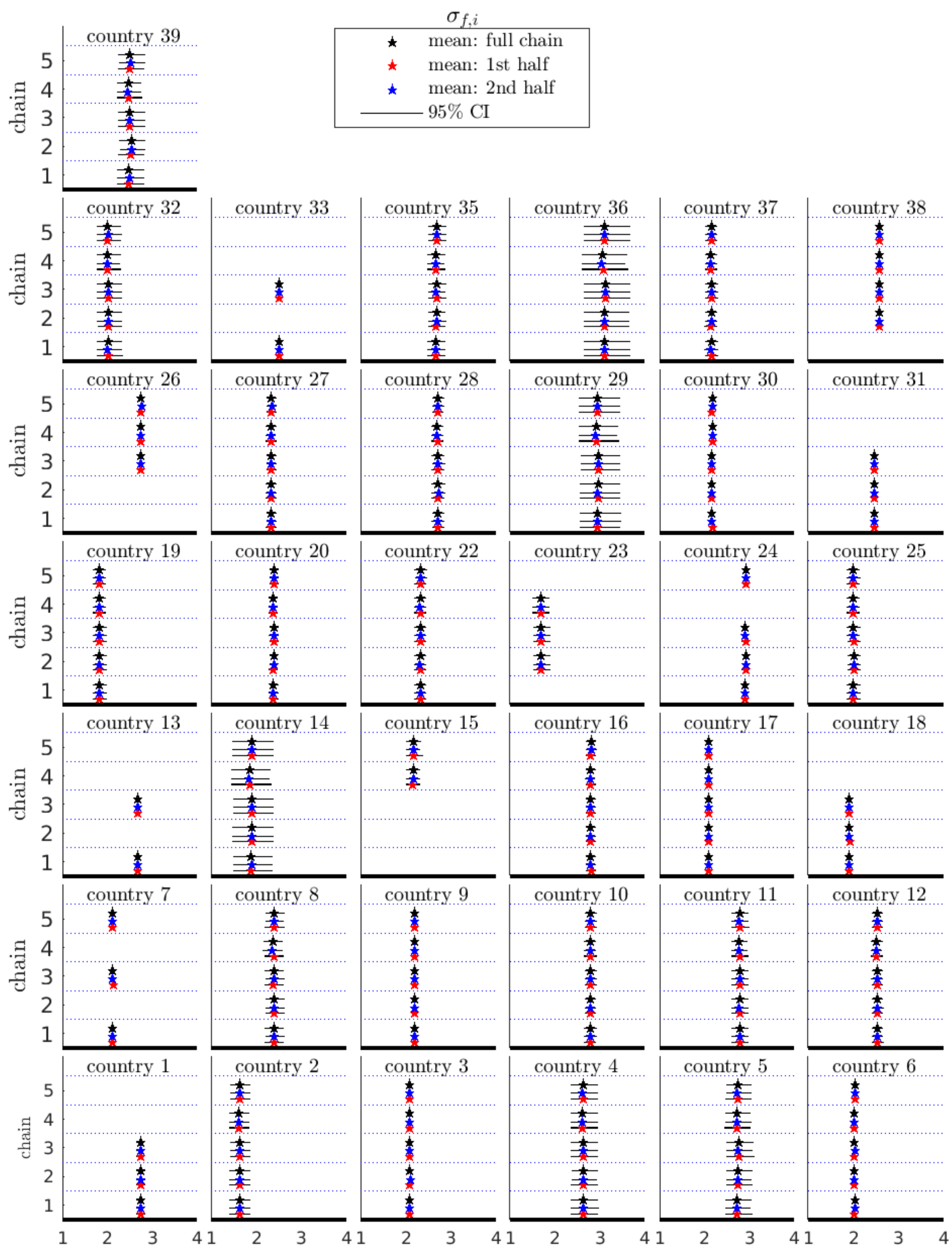
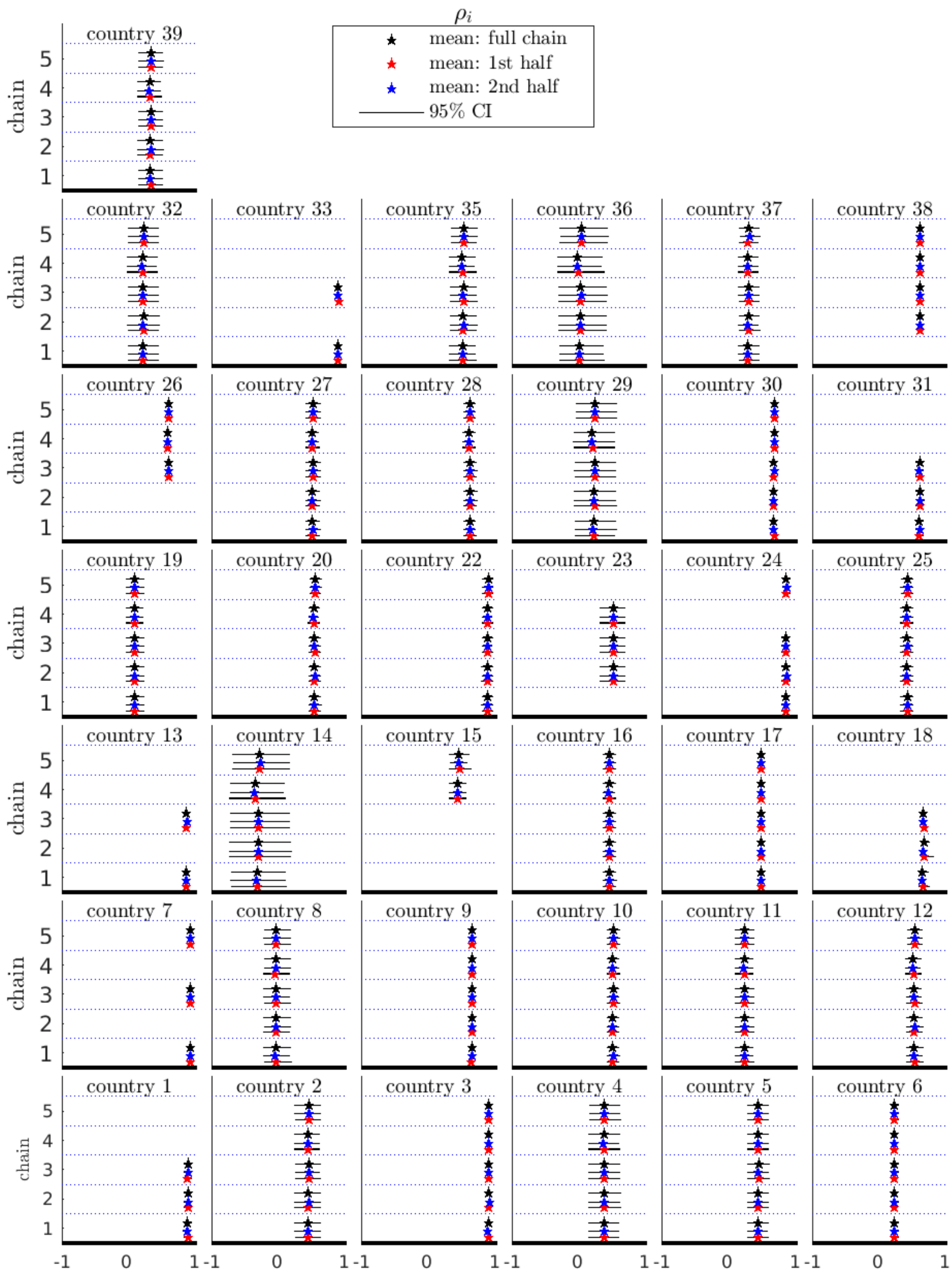


Figure L34: Summary plots for QBE chains: ρ_i



M Unconstrained Melitz-Pareto model

In the main draft we present results of the estimated model with Pareto-distributed productivity shocks with the CDF given by $\Pr\left[\varphi_i \leq \varphi\right] = 1 - \left(\frac{\varphi}{b_i}\right)^{-\zeta_i}$, $\forall\ \varphi_i \geq b_i$, but with restriction $\frac{\zeta_i}{\sigma-1} \in [1.05, \infty]$. Below we present the results for the unconstrained version, in which we only required that $\frac{\zeta_i}{\sigma-1} \in [0, \infty]$. The IME is reported in Table M1 and the elasticity of estimated varaible and fixed trade costs with respect to distance are shown in Table M2. Figure M1 shows results that are analogous to those for the constrained model presented in Figure 5.

Table M1: Implied IME in full Melitz-Pareto models

	IME	95% CI
Data	0.67	[0.61, 0.73]
Full Melitz-lognormal model	0.63	[0.59, 0.67]
Full Melitz-Pareto model, constrained	0.63	[0.57, 0.70]
Full Melitz-Pareto model, unconstrained	0.71	[0.68, 0.74]

Note: The table presents the coefficient from the regression of log average exports on log total exports with origin and destination fixed effects implied by the simulated full Melitz-lognormal model, Melitz-Pareto constrained and unconstrained models. The sample includes 37 origins and 4 main destinations (USA, Germany, France, and Japan) in 2007. The IME in the data is estimated for the same sample. The point estimates and 95% confidence intervals are calculated based on 1,000 simulations based on random parameter draws from the generated Monte-Carlo Markov chain.

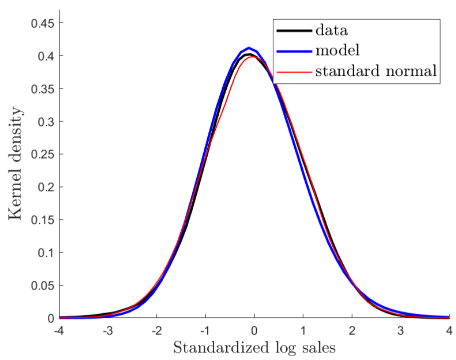
Table M2: Implied trade costs in simulated models

	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
	Melitz lognormal		Melitz-Pareto constrained		Melitz-Pareto unconstrained	
$corr\left(\widetilde{F}_{ij}, \widetilde{\tau}_{ij}\right)$	-0.31	[-0.45, -0.1]	-0.30	[-0.42, -0.19]	-0.52	[-0.72, 0.02]
Distance elasticity:						
Fixed costs	0.31	[0.18, 0.41]	0.50	[0.37, 0.65]	0.49	[0.29, 0.63]
Variable costs	0.34	[0.30, 0.37]	0.29	[0.25, 0.34]	0.16	[0.13, 0.20]

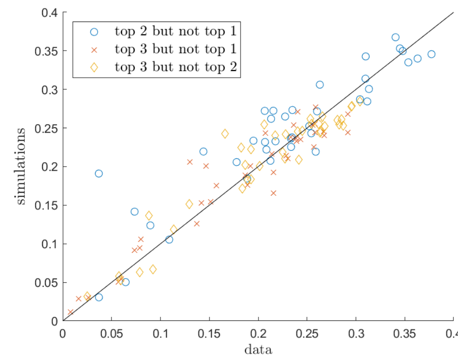
Note: The table presents the coefficient from the regression of log fixed and variable trade costs on distance, origin, and destination fixed effects implied by the simulated lognormal models. The sample includes 37 origins and 4 main destinations (USA, Germany, France, and Japan) in 2007. The point estimates and 95% confidence intervals are calculated based on 1,000 simulations based on random parameter draws from the generated Monte-Carlo Markov chain.

Figure M1: Melitz-Pareto model (unconstrained), goodness of fit

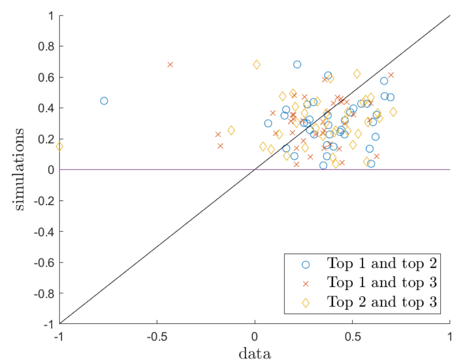
Panel A: PDF of log sales



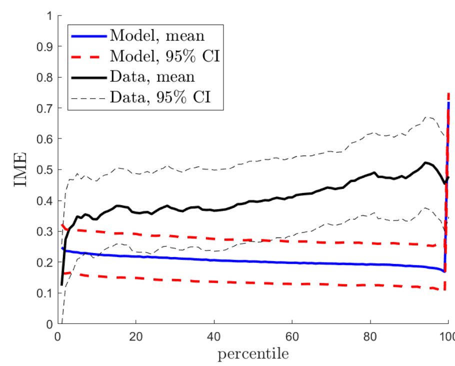
Panel B: Share of firms selling to destination X but not Y



Panel C: Correlation between log exports to top destinations



Panel D: IME for each percentile



Source: Exporter Dynamics Database and authors' calculations. Panel A: the black line corresponds to standardized log sales (demeaned, divided by standard deviation) pooled across different origin-destination cells. The blue line corresponds to standardized log sales pooled across different cells in the model. Panel B: each point corresponds to the share of firms exporting only to less popular markets in the data (horizontal axis) and according to the estimated model (vertical axis) for each origin. Panel C: each point corresponds for each origin and any two destinations among the three most popular ones, to the correlation in export value across all firms that sell in those two destinations in the data (horizontal axis) and according to the estimated model (vertical axis). Panel D: the x-axis represents percentiles; the blue solid line represents the coefficient from the regression of log average exports in each percentile on log total exports in the model; the dashed red lines represent the 95% confidence interval; the black solid line represents the coefficient from the regression of log average exports in each percentile on log total exports in the data; the dashed black lines represent the 95% confidence interval.

N Local trade elasticity in the full Melitz-lognormal model

Let $\bar{\varphi}$ be a combination of demand and productivity shocks (not demeaned and not in logs). Total exports from i to j is given by

$$X_{ij} = N_i \left(\bar{\sigma} w_i \tau_{ij}\right)^{1-\sigma} P_j^{\sigma-1} X_j \int_0^\infty \int_{\varphi_{0,ij}(\tau_{ij}|f)}^\infty \varphi^{\sigma-1} g(\varphi|f) d\varphi g(f) df, \tag{OA.44}$$

where the cutoff productivity for any f is given by

$$\varphi_{0,ij}(\tau_{ij}|f) = \bar{\sigma} w_i \tau_{ij} \left[\frac{\sigma w_i f}{P_j^{\sigma-1} X_j}\right]^{\frac{1}{\sigma-1}}. \tag{OA.45}$$

Assume that changes in τ_{ij} have no effect on P_j. Then trade elasticity is equal to

$$-\theta ij = (1-\sigma) + \frac{\partial \ln \int_0^\infty \int_{\varphi_{0,ij}(\tau_{ij}|f)}^\infty \varphi^{\sigma-1} g(\varphi|f) d\varphi g(f) df}{\partial \ln \tau_{ij}}. \tag{OA.46}$$

The first term is the intensive margin and the second term is the extensive margin (EM) of trade elasticity. Letting

$$\Gamma_{ij} = \int_0^\infty \int_{\varphi_{0,ij}(\tau_{ij}|f)}^\infty \varphi^{\sigma-1} g(\varphi|f) d\varphi g(f) df, \tag{OA.47}$$

then

$$EM = \int_0^\infty \frac{\partial \int_{\varphi_{0,ij}(\tau_{ij}|f)}^\infty \varphi^{\sigma-1} g(\varphi|f) d\varphi}{\partial \tau_{ij}} g(f) df \frac{\tau_{ij}}{\Gamma_{ij}} \tag{OA.48}$$

$$= -\int_0^\infty \varphi_{0,ij}(\tau_{ij}|f)^\sigma g(\varphi_{0,ij}(\tau_{ij}|f)|f) g(f) df \frac{1}{\Gamma_{ij}}, \tag{OA.49}$$

where we used

$$\frac{\partial \varphi_{0,ij}(\tau_{ij}|f)}{\partial \tau_{ij}} = \frac{\varphi_{0,ij}(\tau_{ij}|f)}{\tau_{ij}}. \tag{OA.50}$$

Now we can use the expression for the extensive margin of trade elasticity to calculate local trade elasticity implied by the full Melitz-lognormal model.

O Relation to Melitz and Redding (2015)

In this section we explore how the implications of our model relate to those in Melitz and Redding (2015). We start with a symmetric, two-country economy, with heterogeneous firms as in our Melitz-lognormal model. Following Melitz and Redding (2015), we set the domestic trade share to be 0.9, the share of exporters to be 0.18, and international trade costs to be $\tau = 1.83$. We set the parameters of the productivity and fixed cost shocks at their median values (across origin countries) from our estimation. We then conduct a trade liberalization experiment and compare welfare responses in the true model and the Melitz-Pareto models in the following way:

1. Increase trade costs to τ_{big} and calculate the local trade elasticity θ^s implied by the true lognormal model as outlined in Appendix Section N. This is similar to the "starting" trade elasticity in Melitz and Redding (2015).

2. Reduce trade costs from τ_{big} to 1.25 and calculate the following welfare responses:

   (a) Welfare gains in the true Melitz-lognormal model.

   (b) Welfare gains in the Melitz-Pareto model using the ACR formula with θ^s and changes in trade shares implied by the true model. Following ACR, we refer to this as an ex-post welfare analysis for the Melitz-Pareto model.

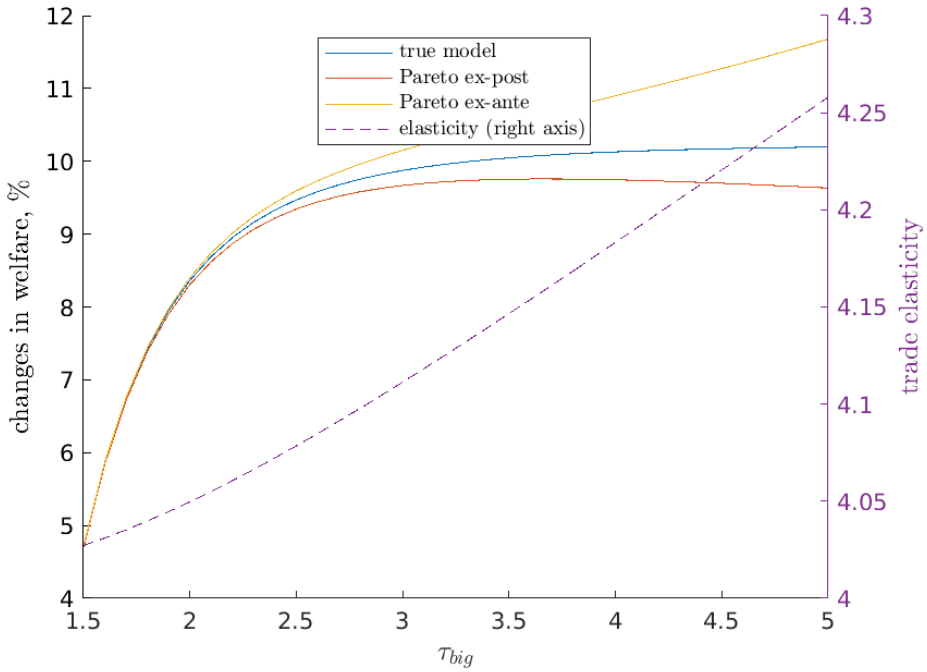
   (c) Welfare gains in the Melitz-Pareto model using the ACR formula with θ^s and changes in trade shares implied by the exact hat algebra in the Melitz-Pareto model. Following ACR, we refer to this as an ex-ante welfare analysis for the Melitz-Pareto model.

The results are reported in Figure O1. In contrast to Melitz and Redding (2015), we do not find large differences between the true model and the Pareto model. As the figure shows, the trade elasticity is below 4.3 even for high starting trade costs. Since the trade elasticity cannot fall below $\sigma - 1 = 4$, the starting trade elasticity θ^s is close to the average trade elasticity along the path of liberalization, and thus the differences between the welfare responses in the two models are not quantitatively large, both for the ex-post and the ex-ante analysis.

To highlight the importance of the variation in the trade elasticity for these results, we redo the exercise above but reducing the standard deviation of the sum of productivity and demand shocks from 4.15 used above to 1.4, which is far below the minimum standard deviation we find in out estimation. As can be seen in Figure O2, the trade elasticity becomes much more variable, and the differences in the welfare response become more pronounced. This establishes that our results are driven not by the nature of our exercise or the use of a lognormal distribution but by the discipline imposed by the estimation using the EDD data.

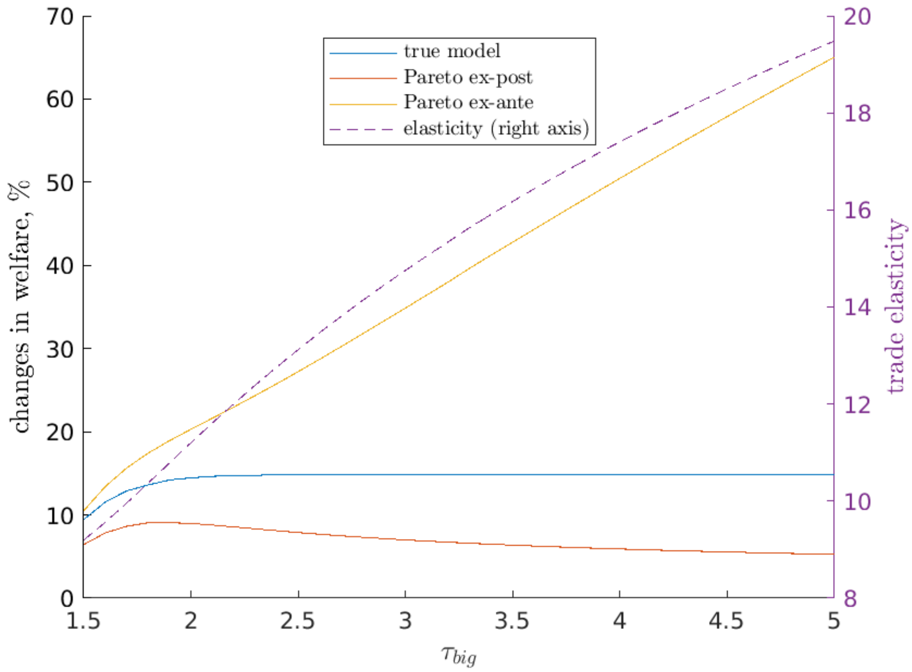
We thus conclude that while the the full Melitz-lognormal model in this paper can theoretically yield similar results to those in Table 1 in Melitz and Redding (2015), that does not happen when the model is estimated on the EDD data.

Figure O1: 2-country world: changes in welfare after trade liberalization I



Note: the figure presents the change in welfare in response to a reduction in trade costs in a symmetrical two-country world in spirit of Melitz and Redding (2015) (see Online Appendix section O for the definitions and calibration of parameters). The x-axis represents the level of trade costs at which trade liberalization starts. The y-axis (left) represents the changes in welfare in different models. The y-axis (right panel) represents the local trade elasticity in the true model at the corresponding level of trade costs.

Figure O2: 2-country world: changes in welfare after trade liberalization II



Note: the figure represents the change in welfare in response to a reduction in trade costs in a symmetrical two-country world in spirit of Melitz and Redding (2015) (see Online Appendix section O for the definitions and calibration of parameters). The x-axis represents the level of trade costs at which trade liberalization starts. The y-axis (left) represents the changes in welfare in different models. The y-axis (right panel) represents the local trade elasticity in the true model at the corresponding level of trade costs.

References

Bas, Maria, Thierry Mayer, and Mathias Thoenig, "From Micro to Macro: Demand, Supply, and Heterogeneity in the Trade Elasticity," *Journal of International Economics*, 2017, *108*, 1–19.

Bento, Pedro and Diego Restuccia, "Misallocation, Establishment Size, and Productivity," *American Economic Journal: Macroeconomics*, 2017, *9* (3), 267–303.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multiproduct Firms and Trade Liberalization*," *Quarterly Journal of Economics*, 2011, *126* (3), 1271–1318.

Chan, Jackie M.L., "Financial Frictions and Trade Intermediation: Theory and Evidence," *European Economic Review*, 2019, *119*, 567–593.

Chernozhukov, Victor and Han Hong, "An MCMC approach to Classical Estimation," *Journal of Econometrics*, 2003, *115* (2), 293–346.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 2011, *79* (5), 1453–1498.

__ , Samuel S. Kortum, and Sebastian Sotelo, "International Trade: Linking Micro and Macro," 2012.

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse, Toolkit, and Cookbook," *Handbook of International Economics*, 2014, *4*, 131–196.

__ , __ , and Mathias Thoenig, "Welfare and Trade without Pareto," *American Economic Review*, 2014, *104* (5), 310–316.

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures," 2011.

Melitz, Marc J. and Stephen J. Redding, "New Trade Models, New Welfare Implications," *American Economic Review*, 2015, *105* (3), 1105–46.

Silva, JMC Santos and Silvana Tenreyro, "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters*, 2011, *112* (2), 220–222.