Online Appendix to "Trade Openness and Growth: A Network-Based Approach"*

Georg Duernecker[1], Moritz Meyer[2], and Fernando Vega-Redondo[3]

[1]Goethe-University Frankfurt, CEPR
[2]World Bank
[3]Bocconi University

July 26, 2022

A Computation of the Globalization Index

Our starting point is an $(n \times n)$-matrix A, which is row-stochastic, as is the one constructed in Section 2.2. We think of it as the adjacency matrix of a weighted directed network over n nodes. Thus, each entry a_{ij} is the relative weight with which node i connects to node j. Viewing such normalized weights as probabilities, the directed distance φ_{ij} from i to j is then identified as the expected number of steps required to reach j from i when, at every node $k = 1, 2, ..., n$, each possible link kl is chosen with probability a_{kl}. In our model, those paths reflect the transfers of information (or know-how) from one country to another, which occur with intensities that are proportional to the trades in the goods and services that embody that information.

To compute such expected magnitude, it is useful to consider the $(n-1) \times (n-1)$ matrix A_{-j} obtained from A by deleting the jth row and the jth column. (This matrix, of course, is no longer a stochastic matrix.) Then, it can be easily seen that the probability that a path that started at i is at $k \neq j$ after r steps is simply $[(A_{-j})^r]_{ik}$, where $(A_{-j})^r$ is the rth-fold composition of A_j with itself, and $[\cdot]_{ik}$ stands for the ik-entry of the matrix $[\cdot]$. Thus, the probability that it visits node j for the first time in step $r+1$ is simply

$$\gamma_{ij}(r+1) = \sum_{k \neq j} [(A_{-j})^r]_{ik} \ a_{kj}.$$

*Corresponding author: Georg Duernecker (duernecker@econ.uni-frankfurt.de)

Therefore, the expected number of steps φ_{ij} can be obtained as follows:

$$\begin{aligned}\varphi_{ij} &= \sum_{r=1}^{\infty} r\ \gamma_{ij}(r) = \sum_{r=0}^{\infty}(r+1)\ \sum_{k\neq j}\left[(A_{-j})^r\right]_{ik}\ a_{kj} \\ &= \sum_{k\neq j}\sum_{r=1}^{\infty} r\left[(A_{-j})^{r-1}\right]_{ik}\ a_{kj} = \left[\left(\sum_{r=1}^{\infty} r\ (A_{-j})^{r-1}\right)_{ik}\right]_{\substack{k=1,2,\ldots,n\\ k\neq j}} \left(a_{kj}\right)_{\substack{k=1,2,\ldots,n\\ k\neq j}}\end{aligned} \tag{1}$$

Using now a standard formula from linear algebra, we have:

$$\sum_{r=1}^{\infty} r\ (A_{-j})^{r-1} = (I - A_{-j})^{-2}$$

so that, in an integrated matrix form, the (column) vector $\left(\varphi_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}}$ can be written as follows

$$\left(\varphi_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}} = (I - A_{-j})^{-2}\ \left(a_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}}.$$

Finally, note that, because A is a row-stochastic matrix, it follows that

$$a_{ij} = 1 - \sum_{k\neq j} a_{ik}$$

and therefore

$$\left(a_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}} = (I - A_{-j})\ e$$

where e is the column vector $(1, 1, ..., 1)^{\top}$. Hence the vector $\left(\varphi_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}}$ can be computed from the following simple expression:

$$\begin{aligned}\left(\varphi_{ij}\right)_{\substack{i=1,2,\ldots,n\\ i\neq j}} &= (I - A_{-j})^{-2}(I - A_{-j})\ e \\ &= (I - A_{-j})^{-1}\ e.\end{aligned}$$

B Empirical model

We follow the approach developed in Moral-Benito (2013, 2016) and augment the dynamic panel model of Section 5 by a feedback process that relates the predetermined variables to all lags of the explained variable, all lags of the predetermined variables, and the exogenous variables. Moreover, we transform the augmented model to obtain a simultaneous-equation representation. This representation has proven useful because it facilitates the estimation of the model by allowing a concentration of the parameters of the model's log-likelihood. Thus, for each country i, the model consists of a system of $T + (T-1)k$ equations, where T is the total number of time periods. Using matrix notation, we can write the model compactly as:

$$A\boldsymbol{R}_i = B\boldsymbol{Z}_i + \boldsymbol{U}_i \tag{2}$$

where the following definitions are used:

$$\boldsymbol{R}_i = (\boldsymbol{y}_i, \mathbf{x}_i)' \qquad \boldsymbol{y}_i = (y_{i1}, y_{i2}, ..., y_{iT})'$$
$$\mathbf{x}_i = (\mathbf{x}_{i2}, \mathbf{x}_{i3}, ..., \mathbf{x}_{iT})' \qquad \mathbf{x}_{it} = \left(x_{it}^1, x_{it}^2, ..., x_{it}^k\right)'$$

$$\boldsymbol{Z}_i = (y_{i0}, \mathbf{x}_{i1}, \mathbf{z}_i)' \qquad \mathbf{z}_i = \left(z_i^1, z_i^2, ..., z_i^m\right)'$$

$$\boldsymbol{U}_i = (\epsilon_i + \mathbf{v}_i, \boldsymbol{\xi}_i)' \qquad \mathbf{v}_i = (v_{i1}, v_{i2}, ..., v_{iT})'$$
$$\boldsymbol{\xi}_i = (\boldsymbol{\xi}_{i2}, \boldsymbol{\xi}_{i3}, ..., \boldsymbol{\xi}_{iT})' \qquad \boldsymbol{\xi}_{it} = \left(\xi_{it}^1, \xi_{it}^2, ..., \xi_{it}^k\right)'$$

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & \mathbf{I} \end{pmatrix} \qquad A_{11} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\alpha & 1 & 0 & \cdots & 0 \\ 0 & -\alpha & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -\alpha & 1 \end{pmatrix} \qquad A_{12} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ -\boldsymbol{\beta} & 0 & \cdots & 0 \\ 0 & -\boldsymbol{\beta} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\boldsymbol{\beta} \end{pmatrix}$$

where $\mathbf{I}$ is an identity matrix of dimension $(T-1)k \times (T-1)k$, ϵ_i can be interpreted as an individual-specific effect, and $\boldsymbol{\xi}_{it}$ is a $k \times 1$ vector of prediction errors. Furthermore, we have:

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \qquad B_1 = \begin{pmatrix} \alpha + \gamma_y & \boldsymbol{\beta} + \boldsymbol{\gamma} & \boldsymbol{\delta} \\ \gamma_y & \boldsymbol{\gamma} & \boldsymbol{\delta} \\ \vdots & \vdots & \vdots \\ \gamma_y & \boldsymbol{\gamma} & \boldsymbol{\delta} \end{pmatrix} \qquad B_2 = \begin{pmatrix} \boldsymbol{\pi}_{2y} & \boldsymbol{\pi}_{2x} & \boldsymbol{\pi}_{2z} \\ \boldsymbol{\pi}_{3y} & \boldsymbol{\pi}_{3x} & \boldsymbol{\pi}_{3z} \\ \vdots & \vdots & \vdots \\ \boldsymbol{\pi}_{Ty} & \boldsymbol{\pi}_{Tx} & \boldsymbol{\pi}_{Tz} \end{pmatrix}$$

$$\boldsymbol{\beta} = \left(\beta^1, \beta^2, ..., \beta^k\right) \qquad \boldsymbol{\gamma} = \left(\gamma^1, \gamma^2, ..., \gamma^k\right) \qquad \boldsymbol{\delta} = \left(\delta^1, \delta^2, ..., \delta^m\right)$$

$$\boldsymbol{\pi}_{ty} = \begin{pmatrix} \pi_{ty}^1 \\ \pi_{ty}^2 \\ \vdots \\ \pi_{ty}^k \end{pmatrix} \qquad \boldsymbol{\pi}_{tx} = \begin{pmatrix} \pi_{tx}^{11} & \pi_{tx}^{12} & \dots & \pi_{tx}^{1k} \\ \pi_{tx}^{21} & \pi_{tx}^{22} & \dots & \pi_{tx}^{2k} \\ \vdots & \vdots & & \vdots \\ \pi_{tx}^{k1} & \pi_{tx}^{k2} & \dots & \pi_{tx}^{kk} \end{pmatrix} \qquad \boldsymbol{\pi}_{tz} = \begin{pmatrix} \pi_{tz}^{11} & \pi_{tz}^{12} & \dots & \pi_{tz}^{1m} \\ \pi_{tz}^{21} & \pi_{tz}^{22} & \dots & \pi_{tz}^{2m} \\ \vdots & \vdots & & \vdots \\ \pi_{tz}^{k1} & \pi_{tz}^{k2} & \dots & \pi_{tz}^{km} \end{pmatrix}.$$

Under normality of the random disturbances, the model in (2) gives rise to the following log-likelihood function:

$$\mathcal{L}(\mathbf{y},\mathbf{X}|\mathbf{Z},\boldsymbol{\theta}) \propto -\frac{N}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}tr\left(\boldsymbol{\Omega}^{-1}\mathbf{U}\mathbf{U}'\right) \tag{3}$$

where $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{Z}$ are the observations on $\mathbf{y}_i$, $\mathbf{x}_i$ and $\mathbf{z}_i$ for all N countries in the sample, $\boldsymbol{\theta}$ is the vector of model parameters, and $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_N]$. Moreover, $\boldsymbol{\Omega}$ is the variance-covariance matrix of $\mathbf{U}$, and $tr(\cdot)$ denotes the trace of the corresponding matrix. Notice that the following simplification was made: $\sum_{n=1}^{N}\mathbf{U}_n'\boldsymbol{\Omega}^{-1}\mathbf{U}_n = tr(\boldsymbol{\Omega}^{-1}\mathbf{U}\mathbf{U}')$. Also, notice that the determinant of A is equal to unity.

C Integrated likelihood

The integrated likelihood used in Equation (13) is defined as follows:

$$p(\mathbf{y}|M_j) = \int p(\mathbf{y}|M_j,\boldsymbol{\theta})f(\boldsymbol{\theta}|M_j)d\boldsymbol{\theta} \tag{4}$$

where $p(\mathbf{y}|M_j,\boldsymbol{\theta})$ is the conditional likelihood of the data. The expression in (4) is typically hard to evaluate, but a simple and accurate approximation, the BIC approximation, makes use of Laplace's method. Let $m(\boldsymbol{\theta}) = \log(p(\mathbf{y}|M_j,\boldsymbol{\theta})f(\boldsymbol{\theta}|M_j))$ denote the posterior mode and construct a Taylor-series expansion of $m(\cdot)$ around $\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} m(\boldsymbol{\theta})$:

$$m(\boldsymbol{\theta}) = m(\tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'m'(\tilde{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'m''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}) \tag{5}$$

where m' and m'' are the first and second derivatives of m, respectively. $m(\boldsymbol{\theta})$ reaches its maximum at $\tilde{\boldsymbol{\theta}}$; therefore $m'(\tilde{\boldsymbol{\theta}}) = 0$, and Equation (5) becomes

$$m(\boldsymbol{\theta}) = m(\tilde{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'m''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}) \tag{6}$$

Inserting (6) into the integral gives:

$$p(\mathbf{y}|M_j) = \int e^{m(\tilde{\boldsymbol{\theta}})+\frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'m''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})}d\boldsymbol{\theta} = e^{m(\tilde{\boldsymbol{\theta}})}\int e^{\frac{1}{2}(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})'m''(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}})}d\boldsymbol{\theta} \tag{7}$$

The integral is a Gaussian integral and, therefore, we obtain the following expression:

$$p(\mathbf{y}|M_j) = e^{m(\tilde{\boldsymbol{\theta}})}(2\pi)^{\frac{k}{2}}|-m''(\tilde{\boldsymbol{\theta}})|^{-\frac{1}{2}} \tag{8}$$

where k and $|-m''(\tilde{\boldsymbol{\theta}})|$ are, respectively, the rank and the determinant of $-m''(\tilde{\boldsymbol{\theta}})$. In a large sample, $\tilde{\boldsymbol{\theta}} \approx \hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$. By taking logs, we obtain:

$$\log p(\mathbf{y}|M_j) = \log p(\mathbf{y}|M_j,\hat{\boldsymbol{\theta}}) + \log f(\hat{\boldsymbol{\theta}}|M_j)) + \frac{k}{2}\log(2\pi) - \frac{1}{2}\log|-m''(\tilde{\boldsymbol{\theta}})| \tag{9}$$

Following Raftery (1995), in large samples, $-m''(\hat{\boldsymbol{\theta}}) \approx N\mathbf{I}$, where N is the number of observations and $\mathbf{I}$ is the expected Fisher information matrix. Using that, we obtain $|-m''(\hat{\boldsymbol{\theta}})| \approx N^k|\mathbf{I}|$ and:

$$\log p(\mathbf{y}|M_j) = \log p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}}) + \log f(\hat{\boldsymbol{\theta}}|M_j)) + \frac{k}{2}\log(2\pi) - \frac{k}{2}\log N - \frac{1}{d}\log|\mathbf{I}| \tag{10}$$

The first and the fourth term on the right-hand side of this expression are of order N and $\log N$, respectively, whereas all other terms are of order 1 or less. When we remove these terms, we arrive at the following expression for the (approximated) integrated likelihood:

$$\log p(\mathbf{y}|M_j) = \log p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}}) - \frac{k}{2}\log N \tag{11}$$

This expression is well known, and it is very similar to the Akaike information criterion. With this expression at hand, we are almost ready to compute the posterior model probability given in (13). One more step is required because the model in (2) does not give us $p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}})$ but rather $p(\mathbf{y}, \mathbf{X}_j|M_j, \hat{\boldsymbol{\theta}})$, which is the joint conditional likelihood of $(y, \mathbf{X}_j)$, with M_j containing the relevant $\mathbf{Z}$-regressor variables.

In the BMA, we consider different models each consisting of a particular combination of regressor variables. If we were to use the joint likelihood $p(y, \mathbf{X}_j|\cdot)$, we would compare different likelihoods, for instance, $p(\mathbf{y}, \mathbf{X}^1, \mathbf{X}^2, ...\mathbf{X}^k|\cdot)$ and $p(\mathbf{y}, \mathbf{X}^4, \mathbf{X}^5, ...\mathbf{X}^k|\cdot)$, which are, in fact, not comparable. Thus, instead, we proceed as follows. For a given model M_j, we first maximize (3) to obtain the maximum likelihood estimate of θ_j. Then, we compute the likelihood of the outcome variable y conditional on the estimated model, that is $p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}}_j)$. Most importantly, this statistic is comparable across the different models, and hence we can use this expression to compute the posterior probability of the underlying model. The conditional likelihood $p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}}_j)$ can be obtained in a relatively straightforward manner by transforming the model given in (2) as follows:

Given $\hat{\boldsymbol{\theta}}$, we first substitute the feedback process into the outcome-equation that yields:

$$y_{n,1} = (\hat{\alpha} + \hat{\gamma}_0)y_{n,0} + \left(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\beta}}\right)\mathbf{x}_{n,1} + \hat{\boldsymbol{\delta}}\mathbf{z}_n + \epsilon_n + v_{n,1} \tag{12}$$

and for $t = 2, ..., T$, we get:

$$y_{n,t} = \hat{\alpha}y_{n,t-1} + \left[\hat{\gamma}_0 + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\pi}}_{t0}\right]y_{n,0} + \left[\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\pi}}_{t1}\right]\mathbf{x}_{n,1} + \left[\hat{\boldsymbol{\delta}} + \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\pi}}_{t2}\right]\mathbf{z}_n + \hat{\boldsymbol{\beta}}\boldsymbol{\xi}_{n,t} + \epsilon_n + v_{n,t} \tag{13}$$

For each country observation i, the model in (12)-(13) is a system of T equations that can be compactly written as:

$$\mathcal{A}\mathbf{y}_i = \mathcal{B}\mathbf{Z}_i + \mathcal{C}\mathbf{U}_i \tag{14}$$

where the following definitions are applied:

$$\mathcal{A} = \hat{A}_{11} \qquad \mathcal{B} = \begin{bmatrix} 0 \\ \left(\mathbf{I}_{T-1} \otimes \hat{\boldsymbol{\beta}}\right) \hat{B}_2 \end{bmatrix} + \hat{B}_1 \qquad \mathcal{C} = \left[\mathbf{I}, -\hat{A}_{12}\right].$$

$\mathbf{I}_{T-1}$ is an identity matrix of order $T-1$. The variables $\mathbf{y}_i$, $\mathbf{Z}_i$ and $\mathbf{U}_i$ are defined as above in (2) together with the matrices $\hat{A}_{11}$, $\hat{\boldsymbol{\beta}}$, $\hat{B}_2$, $\hat{B}_1$, $\hat{A}_{12}$ that are evaluated at the ML-estimate $\hat{\boldsymbol{\theta}}$. Finally, we write the log-likelihood of observation $\mathbf{y}$, conditional on $\mathbf{Z}$ and $\hat{\boldsymbol{\theta}}$ as follows:

$$\log p(\mathbf{y}|M_j, \hat{\boldsymbol{\theta}}) \propto -\frac{N}{2} \log |\mathcal{C}\hat{\boldsymbol{\Omega}}\mathcal{C}'| - \frac{1}{2} tr(\hat{\boldsymbol{\Omega}}^{-1}\mathbf{U}\mathbf{U}'). \tag{15}$$

The expression in (15) is substituted into (11) to obtain the approximated integrated likelihood.

D Data

		Mean	Median	Std	Min	Max
1.	Logarithm of real GDP per capita	8.35	8.39	1.30	5.19	10.72
	Total population (mill.)	38.9	7.98	124	0.35	1148
	Annual growth rate of population	0.02	0.02	0.01	−0.01	0.06
	Price level of investment	0.93	0.64	2.06	0.11	21.5
	Exports plus imports as a share of GDP	0.53	0.46	0.36	0.04	2.90
	Consumption share of GDP	0.72	0.72	0.15	0.23	1.32
	Investment share of GDP	0.22	0.21	0.09	0.03	0.57
	Government share of GDP	0.10	0.08	0.06	0.02	0.39
	Labor force participation rate	0.39	0.39	0.08	0.19	0.57
2.	Life expectancy at birth, in years	59.9	61.8	12.0	30.3	78.8
	Population density, people per km^2	120	37	391	1.4	4547
	Urban population, % of total	0.45	0.43	0.24	0.02	1.00
	Population aged 0–14, % of total	0.38	0.41	0.09	0.16	0.50
	Population share aged 65+, % of total	0.06	0.04	0.04	0.01	0.18
3.	Sachs–Warner openness measure	0.50	0.50	0.50	0.00	1.00
4.	Democracy index	0.58	0.70	0.38	0.00	1.00
5.	1/0 dummy for former Spanish colony	0.21	0.00	0.41	0.00	1.00
6.	1/0 dummy for armed conflict	0.16	0.00	0.37	0.00	1.00
7.	Land area in km^2 (thousand)	1026	272	2115	0.61	9590
	Land share in the geographic tropics	0.55	0.95	0.48	0.00	1.00
	Population share in the geographic tropics	0.51	0.78	0.49	0.00	1.00
	Land area within 100km of navigable water	0.48	0.38	0.37	0.00	1.00
	Land share in Koeppen–Geiger tropics	0.38	0.06	0.42	0.00	1.00
	1/0 dummy for landlocked country	0.16	0.00	0.37	0.00	1.00
	Air distance to NYC, Rotterdam, Tokyo	4205	4065	2594	140	9590
	Timing of national independence	0.96	1.00	0.97	0.00	2.00
	1/0 dummy for socialist rule in 1950–1995	0.10	0.00	0.30	0.00	1.00
8.	Average years of primary schooling	2.87	2.63	1.79	0.02	7.51
	Average years of secondary schooling	1.06	0.72	1.05	0.01	5.09
9.	Globalization Index	0.56	0.55	0.08	0.39	0.76
10.	1/0 dummy for Western European country	0.18	0.00	0.39	0.00	1.00
	1/0 dummy for Latin-American country	0.26	0.00	0.44	0.00	1.00
	1/0 dummy for East Asian country	0.11	0.00	0.31	0.00	1.00
	1/0 dummy for Sub-Saharan country	0.26	0.00	0.44	0.00	1.00

Data sources: 1. Penn World Tables, 2. World Development Indicators, 3. Sachs and Warner: "Trade Openness Indicators," Data set: sachswarneropen.xls, 4. Polity IV Project: Regime Authority Characteristics and Transitions Datasets: p4v2010.xls, 5. Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) geo_cepii.xls, 6. Uppsala Conflict Data Program (UCDP), Data set: 64464_UCDP_PRIO_ArmedConflictDataset_v42011.xls, 7. Gallup, Mellinger, Sachs, Harvard University Center for International Development, Data sets: physfact_rev.csv (Physical geography and population), kgzones.csv (Köppen–Geiger Climate zones), geodata.csv (Geography and Economic Development), 8. Barro and Lee 2000, Data set: appendix_data_tables_in_panel_set_format.xls, 9. UN Comtrade

Table 1: Data: Sources and descriptive statistics.

Asia: Afghanistan, Armenia, Azerbaijan, Bahrain, Bangladesh, Bhutan, Brunei, Cambodia, *China*, Georgia, Hong Kong, *India*, *Indonesia*, *Iran*, Iraq, *Israel*, *Japan*, *Jordan*, Democratic Republic of Korea, *Republic of Korea*, Kuwait, Kyrgyzstan, Laos, Lebanon, Macao, *Malaysia*, Maldives, Mongolia, Myanmar, *Nepal*, Oman, *Pakistan*, *Philippines*, Qatar, Saudi Arabia, Singapore, *Sri Lanka*, *Syria*, Tajikistan, *Thailand*, *Turkey*, Turkmenistan, United Arab Emirates, Uzbekistan, Vietnam, Yemen, Former Yemen	
Europe: Albania, Andorra, *Austria*, Belarus, *Belgium*, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Former Czechoslovakia, Czech Republic, *Denmark*, Estonia, *Finland*, *France*, Germany, East Germany, Former USSR, Gibraltar, *Greece*, Hungary, Iceland, *Ireland*, *Italy*, Kazakhstan, Latvia, Lithuania, Luxembourg, Macedonia, Malta, Moldova, *Netherlands*, *Norway*, Poland, *Portugal*, Romania, Russia, San Marino, Serbia-Montenegro, Slovakia, Slovenia, *Spain*, *Sweden*, *Switzerland*, Ukraine, *United Kingdom*, Former Yugoslavia	
Africa: *Algeria*, Angola, Benin, Botswana, Burkina Faso, Burundi, *Cameroon*, Cape Verde, *Central African Republic*, Chad, Comoros, *Democratic Republic of Congo*, *Republic of Congo*, Cote d'Ivoire, Djibouti, *Egypt*, Equatorial Guinea, Eritrea, Ethiopia, Gabon, *Gambia*, *Ghana*, Guinea, Guinea-Bissau, *Kenya*, Kiribati, Lesotho, Liberia, Libya, Madagascar, *Malawi*, *Mali*, Mauritania, *Mauritius*, Morocco, *Mozambique*, Namibia, *Niger*, Nigeria, *Rwanda*, *Senegal*, Seychelles, *Sierra Leone*, Somalia, *South Africa*, Sudan, Swaziland, Tanzania, *Togo*, *Uganda*, *Tunisia*, *Zambia*, *Zimbabwe*	
North America: Antigua & Barbuda, Bahamas, Barbados, *Belize*, Bermuda, *Canada*, *Costa Rica*, Cuba, Dominia, *Dominican Republic*, *El Salvador*, Grenada, Greenland, *Guatemala*, *Haiti*, *Honduras*, *Jamaica*, *Mexico*, Netherlands Antilles, *Nicaragua*, Former Panama, *Panama*, Saint Kitts-Nevis, Saint Lucia, Saint Vincent, *Trinidad-Tobago*, *United States*	
South America: *Argentina*, Aruba, *Bolivia*, *Brazil*, *Chile*, *Colombia*, *Ecuador*, *El Salvador*, Guyana, *Paraguay*, *Peru*, Suriname, *Uruguay*, *Venezuela*	
Australia: *Australia*, Fiji, French Polynesia, Marshall Islands, Micronesia, New Caledonia, *New Zealand*, Palau, *Papua New Guinea*, Solomon Islands, Samoa, Tonga, Tuvalu	
Countries in *italics* are included in the Bayesian model averaging analysis.	

Table 2: Sample of countries

E Markov chain - Monte Carlo - Model Composition

Here we describe the first-order Markov chain that, as explained in Section 5, approximates the posterior probability distribution induced by our BMA analysis. This Markov chain evolves according to the following transition kernel. Suppose the current state of the chain is M_j. Then, a candidate model is sampled from the neighborhood of M_j, where the neighborhood consists of the set of models with either one variable more or one variable less than in M_j. The candidate model, denoted by $M_{j'}$, is then "compared" to M_j, and it is accepted with probability $\min\{1, \frac{P(M_{j'}|\mathbf{y})}{P(M_j|\mathbf{y})}\}$. If the candidate model is accepted, then the Markov chain moves to $M_{j'}$; otherwise, it stays at M_j. The ratio $\frac{P(M_{j'}|\mathbf{y})}{P(M_j|\mathbf{y})}$ is the posterior odds (= prior odds $\times$ Bayes factor), and it measures how much the data support one model over the other. The posterior odds for M_j and $M_{j'}$ are given by:

$$\frac{p(M_{j'}|\mathbf{y})}{p(M_j|\mathbf{y})} = \frac{p(\mathbf{y}|M_{j'})}{p(\mathbf{y}|M_j)} \times \frac{p(M_{j'})}{p(M_j)}$$

Here, $p(\mathbf{y}|M.)$ and $p(M.)$ are the integrated likelihood and the prior probability of a given model, respectively.

We compute the following diagnostic statistics to check the mixing and convergence properties of the simulated chain. First, we compute the statistic $Corr(\Pi, Freq)$ tests for convergence of the Markov chain, which consists of the following steps: (1) discard the first S_0 steps of the simulated Markov chain to eliminate possible effects from influential starting values; (2) split the remaining chain into two parts: the first S_1 steps and the subsequent S_2 steps; (3) compute the transition matrix T_1, where an element of T_1, say t_{ij}, records how many times the chain has moved from model m_i to model m_j. The dimension of T_1 is equal to the number of different models in S_1; (4) convert T_1 into the transition probability matrix P_1. An element of P_1, say p_{ij}, is determined as $t_{ij}/\sum_{k=1}^{\dim(T)} t_{ik}$, and it measures the probability of the chain moving from m_i to m_j, conditional on being in m_i; (5) calculate the ergodic probability of being in m_i (from P_1^{∞}), which gives the unconditional probability of observing model m_i; (6) derive, for every $m_i \in S_1$, the empirical frequency in S_2 as $c_i/\dim(S_2)$, where c_i counts how often model m_i is visited in S_2; (7) denote by $Corr(\Pi, Freq)$ the correlation coefficient between the ergodic probabilities of all models in S_1 and their empirical frequencies in S_2. $Corr(\Pi, Freq)$ approaches one when the Markov chain reaches stationarity. This is because any two subsets of a stationary chain give rise to the same stationary distribution, and the stationary distribution is (in a large sample) identical to the empirical frequency of each state.

Second, we also compute the statistic $Corr(Bayes, Freq)$, which is another stationarity test that involves the following steps: (1) eliminate a burn-in period from the simulated Markov chain and identify the model with the highest posterior probability, denoting it by $\bar{m}$; (2) compute the empirical frequency for each model in the chain and denote it by f_i; (3) calculate the relative frequency for each model with respect to the best model: $f_i/f_{\bar{m}}$; (4) determine the Bayes factor for each model with respect to the best model: $b_i/b_{\bar{m}}$ (the Bayes factor is the ratio of the posterior probabilities of two models); (5) compute the correlation coefficient $Corr(Bayes, Freq)$ between $f_i/f_{\bar{m}}$ and $b_i/b_{\bar{m}}$. $Corr(Bayes, Freq)$ approaches 1 as the chain reaches stationarity. This is because the model selection along the chain is based upon the Bayes factor (the probability that the chain accepts moving to a candidate model is equal to the Bayes factor between the current model and the candidate model), and as a result, the chain visits those models more often that have a high posterior probability.

Third, we derive the **Raftery–Lewis** dependence factor that is a measure for the mixing behavior of the Markov chain. Dependence factors above 5 are critical and indicate bad mixing of the chain or influential starting values—see Raftery and Lewis (1992) for details (the parameter values required in the test are as in Raftery and Lewis (1992) and given by $q = 0.025$, $r = 0.005$, $s = 0.95$, $\epsilon = 0.001$). To obtain an accurate representation of the posterior distribution, it is important that the chain explores those areas in the model space that have a high probability mass. We follow George and McCulloch (1997) and use a capture-recapture algorithm to estimate what fraction of the total posterior probability mass the Markov chain has visited.

In Table 3, we report a number of statistics describing the properties of the simulated Markov chain. *Markov steps* refers to the total number of steps (in 1000) of the simulated chain. *Posterior model size* refers to the posterior model size. *Models covering* 50% is the number of models with the highest posterior model probability that, in sum, account for 50% of the posterior model probability. *P(max)* is the maximum posterior model probability achieved by a single model. *Visited probability* refers to the estimated fraction of the total posterior probability mass that the Markov chain has visited. This number is computed using the capture-recapture algorithm described in George and McCulloch (1997). The remaining statistics describe the convergence and mixing properties of the simulated chain. Generally, the values of these indicators indicate very good mixing and convergence properties of the simulated Markov chain. For example, the values of Corr(Π,Freq) and Corr(Bayes,Freq) are very close to unity, suggesting that the simulated Markov chain has reached stationarity. Furthermore, we obtain a Raftery–Lewis factor equal to 3.38, which indicates fast mixing of the process. Factors above 5 are critical and indicate bad mixing of the chain or influential starting values. Lastly, the estimate for the total posterior probability mass that the Markov chain has visited is very high and equal to 98%. The high value is reassuring because an accurate representation of the posterior distribution requires that the Markov chain reaches the areas in the model space with high probability mass.

	Benchmark	**Higher-order trade**
Markov steps (× 1000)	836	996
Posterior model size	8.7	8.3
Models covering 50%	58	97
Pr(best model)	7.20	4.78
Visited probability	98.0	95.7
Corr(Π,Freq)	0.997	0.909
Corr(Bayes,Freq)	0.998	0.966
Raftery–Lewis factor	3.38	3.42

Table 3: MC^3 statistics.

F Robustness

As advanced in Section 8, here we explore the sensitivity of the findings in Section 6 to variations in the data input, to modifications of the underlying model assumptions, and to alternative measures of network centrality.

F.1 Data

Raw data vs. cleaned data: In our baseline approach, we use the raw trade data from the UN Comtrade to compute the GI. However, a National Bureau of Economic Research project led by Robert Feenstra has systematically cleaned a number of inconsistencies from the UN Comtrade data. The resulting data set is available from the Center for International Data, and a detailed description is provided in Feenstra et al. (2005). As a robustness check, we use these data instead of the raw trade data to compute our GI. Then, we perform a BMA analysis where we include this new measure. The row labeled *Feenstra* in Table 4 shows the resulting findings are very similar to the baseline results.

	$E(\theta_k\|\mathbf{y})$	PIP	$\%_{sig}$
Benchmark	**6.289*****	**85**	**99**
Feenstra	5.497***	77	97
IMF DOTS	6.084***	88	96
PWT 6.2	5.589***	83	99
PWT 6.3	4.387**	95	88
PWT 7.0	7.318***	80	97
PWT 7.1, 1960–2009	5.823**	71	96
PWT 7.1, 5 yrs	2.169***	79	97

Table 4: Robustness: Data.

IMF DOTS: The International Monetary Fund (IMF) publishes the Direction of Trade Statistics (DOTS), which provides detailed data on bilateral trade flows. We use these data instead of the UN Comtrade data to compute the GI. Again, the results are largely unchanged—see row "IMF DOTS" in Table 4.

Penn World Tables: Several variables included in the empirical analysis are constructed from data taken from the Penn World Tables (PWT). Ciccone and Jarocinski (2010) raise the important concern that the results of growth empirics are often sensitive to revisions in the PWT data. We address this concern by using different releases of the PWT to compute the relevant variables. Table 4 compares the results. By and large, our baseline findings are robust to revisions of the PWT. An advantage of the recent releases of the PWT is that they extend the time period covered by the data, which allows us to consider a longer period in the BMA. Specifically, we can use the period from 1960–2010, which gives us a total of five observations for each country. Again, the results are very similar to the baseline findings. As yet another check, we also organize the data into five-year time intervals (instead of using 10-year intervals), giving us a total of 10 country observations. As can be seen in Table 4, the higher-frequency data do not lead to noteworthy changes in the sign and significance levels of the results.

F.2 Model specification

In the baseline approach, we use a binomial-beta structure as the model prior distribution. We test the robustness of this choice by using a uniform prior as in Moral-Benito (2016). Accordingly, all models are equally likely a priori, and $P(M_j) = 2^{-K}$, where K is the number of potential regressor variables. The results of the BMA analysis with the uniform prior are very close to the results of the baseline case, as can be seen from Table 5. Hence, we conclude that the assumption on the model prior distribution does not have a significant effect on the results.

	$E(\theta_k\|\mathbf{y})$	PIP	$\%_{sig}$
Benchmark	**6.29*****	**0.85**	**0.99**
Uniform prior	6.46**	0.95	0.97
Rich	9.52***	0.98	0.99
Poor	4.18**	0.79	0.84
15 best covariates	6.41***	0.99	0.99
10 best covariates	6.37***	0.99	1.00
15 worst covariates	4.93***	0.99	0.98
10 worst covariates	4.61***	0.99	1.00

Table 5: Robustness: Model.

Next, we note a large degree of heterogeneity among the countries in our sample. To account for this heterogeneity, we include into the empirical analysis a large set of covariates and a full set of dummy variables to control for country, region, and time fixed effects.

Despite these efforts, we cannot exclude the possibility of additional dependencies that we have not properly controlled for. Two particularly relevant concerns are the existence of spatially correlated shocks that affect geographically-proximate countries and the potentially differential effect of the covariate for developed and undeveloped countries. We address these concerns in various ways. First, we cluster the data by splitting the sample into rich and poor countries. More concretely, we consider countries as poor (rich) if their GDP per capita in 1960 was less (more) than 1/5 of the level of the United States. The resulting samples consist of 48 poor countries and 34 rich countries. Then, we perform the BMA analysis on both samples separately and report the results in Table 5. Importantly, the positive relationship between openness and growth is found to be very robust for both high-income and poor countries, but it seems somewhat stronger for the rich.

In a similar vein, we have also interacted in a separate experiment the trade share with the region fixed effect to assess whether the relationship between the traditional openness measure and growth varies across regions. Interestingly, we find that the insignificant relationship between the trade share and growth that arises in the baseline case also holds across all five regions that we consider. Moreover, we interacted the region-fixed effect with time-fixed effects to account for spatially correlated shocks. Again, we conclude that the results are very similar to the baseline results, especially for the estimate of the globalization measure. For conciseness, we do not report the results of the last two experiments here, but they are available upon request.

As an additional robustness test, we check whether our main findings are sensitive to the number of regressor variables included in the empirical model. In the baseline case, we consider 34 candidate regressors. Now, we include only a subset of these variables into the model. In particular, we pick those 10 (15) variables that had the highest posterior inclusion probability in the baseline case. As an additional experiment, we select—together with the GI and initial GDP per capita—those 10 (15) variables that had the lowest posterior inclusion probability.

The results of the BMA analysis are in Table 5, and again, we observe no significant change with respect to the baseline findings.

Finally, we recall that the analysis in Section 6 reveals a weak relationship between the traditional measures of openness—such as the trade share and the Sachs–Warner index—and economic growth, as indicated by low values of their posterior inclusion probability. Now we want to address the concern that this result may be driven by a potential dependence between our GI and the traditional measures. Table 6 shows the posterior mean and the inclusion probability (rows) of the openness measures for the baseline case and for different combinations of included variables (column). The results in the table do not reveal any notable dependencies between the different measures.

		Baseline	Include			
			GI S&W	GI TS	GI	TS S&W
GI	$E(\theta_k\|\mathbf{y})$	6.29***	6.21***	6.15***	6.11***	
	PIP	0.85	0.83	0.81	0.80	
TS	$E(\theta_k\|\mathbf{y})$	−0.07		−0.04		0.10
	PIP	0.04		0.04		0.05
S&W	$E(\theta_k\|\mathbf{y})$	0.18***	0.18***			0.13**
	PIP	0.16	0.36			0.07

Table 6: Robustness: Openness measures.

F.3 Alternative measures of network centrality

Our GI reflects a notion of network centrality that is known as closeness centrality. Other prominent notions of centrality considered in the literature are PageRank, Bonacich, eigenvalue, or betweenness centralities—see, for example, Bloch, Jackson, and Tebaldi (2017). we focus on PageRank centrality to avoid redundancy because they all behave quite similarly for the relevant parameter ranges. According to PageRank centrality, a central/influential node is identified as one that is largely connected to central/influential nodes. If we denote by $\nu = (\nu_i)$ the vector specifying such an "impact" for every node i, the centrality condition can then be written as

$$\nu = (\tilde{A})^T \nu,$$

where $\tilde{A}$ is a perturbation of the adjacency matrix A defined by $\tilde{A} = \alpha A + (1-\alpha)U$, where $0 < \alpha < 1$, and U is a (stochastic) matrix with entries all equal $1/n$. The matrix $\tilde{A}$ can still be formally interpreted as the transition probability matrix of a Markov process. Such a Markov process is clearly ergodic and thus has a unique invariant distribution. This allows PageRank to identify the centrality of any given node i as its weight in that invariant distribution so that we may write

$$\nu = \frac{1-\alpha}{n}(I - \alpha A^\top)^{-1}\mathbf{e} \tag{16}$$

where n is the dimension of A and $\mathbf{e}$ is a column vector of all 1s. The notion of centrality given by (16) implicitly presumes that all nodes in the network are symmetric and command the same value. However, just as we did for our baseline measure introduced in Subsection 2.2, we want to account for the fact that countries are very different in relative size within the world economy. Again, this can be captured by replacing the uniform weighting embodied by the vector $\mathbf{e}$ by the alternative vector $\boldsymbol{\beta}$ (also used by our baseline measure) where each β_i captures the fraction of country i's GDP in the world economy. This leads to the following modified notion of PageRank centrality:

$$\nu = \frac{1-\alpha}{n}(I - \alpha A^\top)^{-1}\boldsymbol{\beta}, \tag{17}$$

which is the measure of integration we apply to our full sample of 200 countries and all years from 1962 to 2012. Table 7 below reports the outcome of the BMA exercise for different values of α. There we observe that the magnitude of the posterior mean estimate of the PageRank coefficient, the corresponding inclusion probability, and the $\%_{sig}$ statistic all grow monotonically with α, only achieving truly high values when this parameter is also high. These results are very much in line with those obtained for our benchmark measure of country integration because the parameter α plays in the present case a role analogous to δ for our benchmark integration measure. Here, α determines how much PageRank is dependent on the network architecture, hence depending on the full set of paths that directly and indirectly join each pair of nodes. The results of Table 7, therefore, are again a manifestation of the importance that long-range indirect connections have on growth even if integration were measured by the notion of PageRank centrality.

	$E(\theta_k \vert \mathbf{y})$	PIP	$\%_{sig}$
PageRank centrality			
$\alpha = 0.95$	2.6332**	0.63	76
$\alpha = 0.75$	2.0135*	0.34	71
$\alpha = 0.50$	0.7264	0.11	52
$\alpha = 0.25$	0.0506	0.08	44

Table 7: Global vs. local connections: PageRank centrality.

In addition to PageRank centrality, we have experimented with several other integration measures that belong to none of the aforementioned centrality concepts. Most noteworthy among those is the approach suggested by Arribas et al. (2009). One of the indicators they use to assess a country's integration is what they call *degree of connection* (DTC), which compares the trade of a given country in the actual world with what would prevail in an ideal and perfectly integrated one. More specifically, DTC measures whether a country's international flows match the weight of the other countries, being equal to 1 in case of a perfect match. This approach is conceptually very different from ours. Arribas et al. (2009) also consider the *degree of openness* (DO), which, for each country, is equivalent to 1 minus its corresponding diagonal element in our adjacency matrix A. These two different indicators capture a country's aggregate trade flows but not its architecture of first- and higher-order trade connections. Consequently, it is not surprising that the correlation between our integration indicator and DO and DTC is generally very low (as we showed to be the case with the traditional measures of openness in Section 4). For example, in 2004, it is equal to -0.03 and -0.05, respectively. We also find an insignificant role for these indicators when included in the BMA. For instance, the posterior mean associated with the indicator DO is not significant (even at the 10% level), and the posterior inclusion probability is only 6%.[1]

Lastly, we consider three different versions of random perturbations to the diffusion matrix A in order to address the criticism expressed in Keller (1998) that a spurious version of the trade network is likely to have the same implications for the global transmission of information as the actual trade network. At the same time, the analysis below allows assessing the importance of different dimensions of the network structure for the relationship between the GI and growth. In the first case, we keep the structure of the original matrix A as in the baseline case—in terms of the number of each country's links and the set of its partners—and we simply perturb the weight of existing links. In particular, we randomly assign a weight between 0 and 1 to each existing link and re-normalize the resulting matrix so that it is row-stochastic. This approach implies only a small modification to the original transition matrix A because the structure of the matrix is preserved. Using this modified version of the transition matrix, we compute the GI according to the approach described in Section 2.2. Clearly, the values of the GI depend on the realization of the random draws of the link weights. To eliminate the variation in the GI that is due to this randomness, we compute the GI for 100 different sets of realizations and average over the outcomes. Lastly, we include the resulting GI into the BMA analysis. The estimated coefficient of the GI is significant only at the 10% level, and the posterior inclusion probability drops from 85% in the baseline case to 42%.

The second case that we consider involves a more substantial modification of the matrix A. We keep the number and the weight of existing links for each country but assign the links to a randomly selected set of trading partners. That is, we reshuffle the existing links of a given country. As before, we use the perturbed transition matrix to compute the GI, then we average over 100 different realizations and include the resulting GI in the BMA. The estimated coefficient of the GI becomes insignificant, and the posterior inclusion probability of only 2%t is significantly below the baseline value. In the last case, we allocate the total weight of each country's links to a randomly selected set of trading partners. That is, we keep the outward orientation of countries

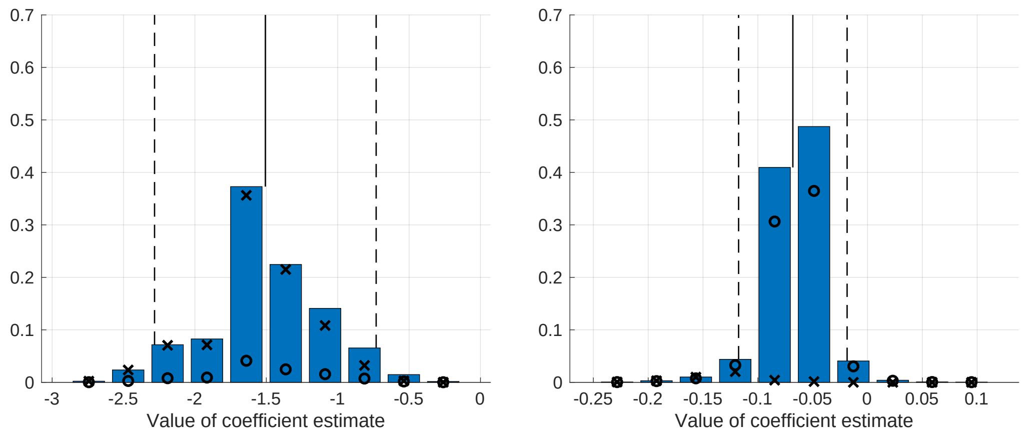
[1] In another experiment, we identify the first principle component (FPC) of trade openness and compare it to the GI. To conduct this comparison, we compute the correlation between the two variables and, in addition, include the FPC instead of the GI into the BMA. We find a correlation coefficient of -0.36, which is slightly higher (in absolute terms) than that for trade openness and the GI of -0.10. Still, the value is rather low, indicating a relatively weak relationship between the two variables. When including the FPC into the BMA, we find a posterior inclusion probability for this variable of less than 1%.

as in the baseline case but perturb the number of links. In this case, we also obtain in the BMA an insignificant coefficient estimate and very low posterior inclusion probability for the GI of 2%.

We interpret these findings as reflecting the importance of both the structure of the trade network—in terms of the number of links of a country and the set of its trading partners—and the intensity of trade connections between countries for the explanatory power of the GI. If we keep the structure but modify the intensity of trade connections (as in the first case), then the posterior inclusion probability of the GI declines substantially, but it is still higher than for 26 of 34 of the included covariates. Instead, if we perturb the set of trading partners (second case) and, in addition, also the number of links (third case), then the relationship between the modified GI and growth becomes very weak.

G Explaining discordance within the BMA analysis

As can be seen from Table 6 in the main text, there is a marked misalignment between the posterior inclusion probability and the $\%_{sig}$-statistic for several of the variables included in the BMA analysis. For example, the *Government share* has a PIP of only 11%, but the estimated coefficient is significant in 91% of the models. To understand this pattern it is useful to consider Figure 1, which focuses on the variables *Government share* and *Armed conflict*. It shows the posterior probability mass over the whole range of coefficient estimates (bars) and, for each value of the estimated coefficient, the share of models where the estimation is significant at the 5% level (crosses) and the posterior inclusion probabilities of the respective models (circles). The solid line and the broken lines represent the posterior mean and the 95% confidence bounds, respectively.



(a) Government share ($PIP = 0.11$; $\%_{sig} = 91\%$) (b) Armed conflict ($PIP = 0.75$; $\%_{sig} = 6\%$)

Figure 1: Discordance between PIP and $\%_{sig}$.

A comparison of the two panels yields some useful insights. It illustrates, in particular, that the posterior inclusion probability of a variable and the share of significant coefficient estimates can be very different. In Panel (a), we observe that the coefficient associated with the *Government share* is typically estimated very precisely across models (crosses are close to the top of each bar), while the models that contained this variable are generally characterized by a low goodness-of-fit (circles close to the bottom). As a result, the posterior inclusion probability of these models is rather low across the entire range of the estimated coefficient. The opposite can be observed in Panel (b), which shows the same set of statistics for the variable *Armed conflict*. We find that the coefficient for *Armed conflict* is generally very imprecisely estimated, whereas the models that include it have a high goodness-of-fit and thus provide this variable with a high posterior inclusion probability.

In sum, the point here is that the identification of robust covariates according to the posterior inclusion probability (as done by the model averaging approach) can lead to conclusions that are very different from the traditional (single-equation) growth empirics that typically evaluates variables on the basis of the significance level of the estimated coefficient for a certain model specification. As a result of this practice, much of the empirical growth literature considers the variable *Government share* as robustly related to growth (see, for example, the work by Barro (1991, 1996) and Caselli et al. (1996)) whereas the results above lead to concluding the exact opposite. The same applies (but in reverse order) to the variable *Armed conflict*. With a posterior inclusion probability of 75%, this variable is found to be strongly related to growth. This result is in stark contrast to much of the existing empirical work that interprets the mostly insignificant coefficient estimates for this variable as evidence for a limited explanatory role. See, for example, Barro and Lee (1994) and Easterly and Levine (1997). Such contradictory assessment can also be established for several other candidate regressors, such as the *Investment price* (Easterly, 1993), the *Life expectancy* (Barro and Lee, 1994), *Democracy* (Barro, 1996; Dollar and Kraay, 2003), *Landlocked* (Easterly and Levine, 2001), or *Former Spanish colony* (Barro, 1996), all of which have been suggested to be important for economic growth. Instead, according to our results, these variables are characterized by low values of the posterior inclusion probability, hence indicating a weak relationship to growth. For yet other variables, our results are in line with the findings of the traditional empirical growth literature. This includes, for example, the *Investment share* and the dummy variable for *Sub-Saharan countries*.[2]

H Geography and the Globalization Index

H.1 Modified globalization index

The computation of the modified GI presented in Section 7.4 for a given country i involves the following steps. First, we denote by $\varphi_{m,j,-i}$ the expected number of steps required to reach j from any country $m \neq i$, conditional on **not** utilizing any of the links that involve country i.

[2] See Barro (1991, 1996), Barro and Lee (1994), Caselli et al. (1996), Easterly and Levine (1997), and Sala-i-Martin (1997a, 1997b).

$\varphi_{m,j,-i}$ can be derived as follows:

$$\varphi_{m,j,-i} = \sum_{k\neq i,j} \sum_{r=1}^{\infty} r \left[(A_{-i,-j})^{r-1} \right]_{m,k} a_{k,j} \tag{18}$$

Here $A_{-i,-j}$ is a $(n-2) \times (n-2)$ matrix obtained from the original adjacency matrix A by deleting the ith and the jth column and the ith and the jth row. $[\cdot]_{m,k}$ indicates the elements of the mth row and the kth column of the array $[\cdot]$. Rearranging Equation (18) yields the following expression:

$$\varphi_{m,j,-i} = \left[\left(\sum_{r=1}^{\infty} r \left(A_{-i,-j}\right)^{r-1} \right)_{m,k} \right]_{k=1,2,...,n;\, i\neq k\neq j} \left(a_{k,j}\right)_{k=1,2,...,n;\, i\neq k\neq j} \tag{19}$$

where $\left(a_{k,j}\right)_{k=1,2,...,n; k\neq i,j}$ is an $(n-2)\times 1$ vector that is obtained from the jth column of matrix A by deleting the ith and the jth element. We use $\sum_{r=1}^{\infty} r \left(A_{-i,-j}\right)^{r-1} = \left(I - A_{-i,-j}\right)^{-2}$ and substitute it into (19), to obtain

$$\varphi_{m,j,-i} = \left[\left(I - A_{-i,-j}\right)^{-2}_{m,k} \right]_{k=1,2,...,n;\, i\neq k\neq j} \left(a_{k,j}\right)_{k=1,2,...,n;\, i\neq k\neq j} \tag{20}$$

We compute $\varphi_{m,j,-i}$ for all combinations of (m, j), where $m = 1, 2, ..., n$ and $j = 1, 2, ..., n$, with $m \neq i$, $j \neq i$. This yields the $(n-1) \times (n-1)$ dimensional matrix $\left(\varphi_{m,j,-i}\right)^{n}_{m=1,j=1;\, m\neq i\neq j}$. An element of this matrix specifies the expected number of steps from any country j to each of country i's potential trading partners $m = 1, 2, ..., n, m \neq i$. The key difference to the related matrix in the benchmark case, that is, $\left(\varphi_{m,j}\right)^{n}_{m=1,j=1}$, is that here all connections from and to country i are disregarded. The remaining steps of the calculations involve the aggregation of $\varphi_{m,j,-i}$ using the distance-related weighting factors as described in the main text.

H.2 The Frankel–Romer approach

The volume of trade of a country is potentially affected by its rate of economic growth, which renders the matrix $A_t = \left(a_{ijt}\right)^{n}_{i,j=1}$ induced by the trade flows of year t and the resulting GI, Φ_{it}, possibly endogenous to growth. In this section, we take a step to alleviate this endogeneity issue. More concretely, in the spirit of the approach pursued by Frankel and Romer (1999), we construct a modified GI measure that is based on bilateral geographical distance alone and rely on it to instrument for Φ_{it}.

More concretely, the procedure implements the following steps. Let geo_{ij} denote the geographical distance (measured in kilometers) between countries i and j. In the first step, we replace the elements of the transition matrix, a_{ijt}, with the inverse of the geographical distance, $1/geo_{ij}$, between countries i and j. Naturally, after this step, the sum of each row is no longer equal to one. Thus, to make the matrix row-stochastic, we normalize the elements of each row by the sum of each row. Let $\tilde{A}_t$ denote this modified transition matrix. An element of this matrix, denoted by $\tilde{a}_{ijt}$, is given by $\frac{1/geo_{ij}}{\sum_k 1/geo_{ik}}$. Clearly, $\tilde{a}_{ijt}$ is exogenous to growth. Next, we use the modified transition matrix, $\tilde{A}_t$, to compute the GI as described in Equations (9) and (10). Let by $\tilde{\Phi}_{it}$ denote the value of the modified GI for country i in period t. A key step of our approach is to use $\tilde{\Phi}_{it}$ as an instrument for the potentially endogenous GI, Φ_{it}. Specifically, we estimate the following first-stage regression by OLS:

$$\Phi_{it} = \alpha + \gamma\tilde{\Phi}_{it} + \mu_i + \zeta_t + \epsilon_{it}$$

where μ_i and ζ_t represent country and time fixed effects. We also consider a version where we do not include fixed effects. We compute $\tilde{\Phi}_{it}$ for all countries in our sample and for all years, which gives us a total of 9553 observations. The estimated value of γ obtained from the regression is equal to 0.45 and is highly significant with a 95% confidence interval of $[0.42, 0.47]$. Moreover, the F-statistic of this regression is equal to 768.1 and far exceeds the value of 10 that is typically considered the critical value for indicating weak instruments. Let by $\widehat{\Phi}_{it}$ denote the predicted values of the regression. In the final step, we include $\widehat{\Phi}_{it}$ instead of the baseline GI measure, Φ_{it}, into the BMA. Importantly, the estimated coefficient of the modified GI is highly significant, and the posterior inclusion probability of 62% is only slightly below that of the baseline GI.

Two remarks are in order. First, even though the geographical distance between countries is time invariant, the values of the modified GI are not necessarily constant over time. This is because the number and the distribution of links in the trade network can change from year to year. Second, and relatedly, while the modified GI alleviates the endogeneity issue by using geographical distance as a measure of bilateral trade intensity, it does not completely remove it. Arguably, we cannot exclude the possibility that the number of a country's links is endogenous to its growth performance. That is, our approach does not tackle the endogeneity of whether two given countries engage in bilateral trade at all (extensive margin of trade) but only how much they trade (intensive margin). As a result of the latter observation, we do not interpret the results of the BMA with the modified GI as causal per se.

I Analysis of the patent data

In our analysis in Subsection 7.2, we focus on the patents originating in a sample of $n = 149$ countries that cite at least one other patent from a foreign country. That is, we disregard patents that (i) cite no other patent or (ii) cite only patents of the same country. The latter condition derives from the fact that we are interested in the flow of ideas between countries, and own-country citations do not contribute to that flow. The analysis has centered on two variables, Avg_{ij} and $Prob_{ij}^{inv}$, that measure, respectively, the average number of cited patents from j cited in every citing patent from i, and the fraction of cross-country patenting relationships that connect an inventor from i with another in j. Here we provide a precise description of how these variables are derived.

First, we explain the computation of each Avg_{ij}. It is based on two matrices, P and C, of the following form:

$$P = \begin{pmatrix} 0 & p_{12} & p_{13} & ... & p_{1n} \\ p_{21} & 0 & p_{23} & ... & p_{2n} \\ . & . & . & ... & . \\ p_{n1} & p_{n2} & p_{n3} & ... & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 0 & c_{12} & c_{13} & ... & c_{1n} \\ c_{21} & 0 & c_{23} & ... & c_{2n} \\ . & . & . & ... & . \\ c_{n1} & c_{n2} & c_{n3} & ... & 0 \end{pmatrix}.$$

The elements p_{ij} in matrix P represent the number of country-i patents that cite at least one country-j patent. Notice that, in general, $p_{ij} \neq p_{ji}$ and, of course, we may also find many elements in P for which $p_{ij} = 0$. That is, cross-country patenting need not be symmetric and the cross-citing patent network could be quite sparse. In our case the total number of elements for which $p_{ij} > 0$ is equal to 3376 (thus much lower than the maximum $n(n-1)$), whereas the total number of patents that cite a foreign patent is equal to $\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} = 2.98MM$.[3]

In contrast, the elements c_{ij} in matrix C count how many country-j patents are cited in total by country-i patents. Notice that this is a conditional statement as we include only those country-i patents in c_{ij} that cite at least one country-j patent. $\sum_{j=1}^{n} c_{ij}$ is the total number of foreign patents cited by country-i patents. For our sample, we obtain that the total number of citations to foreign patents is equal to $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} = 5.82mill$. The element-by-element division of both matrices C and P gives $Avg_{ij} = c_{ij}/p_{ij}$, which is the average number of country-j patents cited per country-i patents.

Next, we explain how the variables $Prob_{ij}^{inv}$ are obtained. Their computation relies on the following matrix:

$$T = \begin{pmatrix} 0 & t_{12} & t_{13} & ... & t_{1n} \\ t_{21} & 0 & t_{23} & ... & t_{2n} \\ . & . & . & ... & . \\ t_{n1} & t_{n2} & t_{n3} & ... & 0 \end{pmatrix}.$$

An element t_{ij} in matrix T specifies the total number of bilateral co-patenting relationships between inventors from countries i and j. To fix ideas, consider two patents: Patent 1 was created by a team of four U.S. inventors, two French inventors and two German inventors. Patent 2 was created by two U.S. inventors and three French inventors. Then, for this example, we would obtain $t_{US,FRA} = t_{FRA,US} = 8 + 6 = 14$, $t_{US,GER} = t_{GER,US} = 8$, $t_{FRA,GER} = t_{GER,FRA} = 4$. In our sample, the number of entries in the matrix T for which $t_{ij} > 0$ is equal to 1918 and the total number of collaborations between international inventors is $\sum_{i=1}^{n}\sum_{j=1}^{n} t_{ij} = 286,168$. Computing the fraction $t_{ij}/\sum_{j=1}^{n} t_{ij}$ for each $i, j = 1, 2, ..., n$, we arrive at the corresponding $Prob_{ij}^{inv}$.

[3]Note that if a country-i patent cites country-j and country-k patents, then this country-i patent will be counted in both p_{ij} and p_{ik}. Due to this multiple counting of patents, we find that the row-sum $\sum_{j=1}^{n} p_{ij}$ is higher than the total number of country-i patents.

References

[1] Arribas, I., Pérez, F., & Tortosa-Ausina, E. (2009). Measuring globalization of international trade: Theory and evidence. *World Development* **37**, 127–145.

[2] Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics* **106**, 407–433.

[3] Barro, R. (1996). Democracy and growth. *Journal of Economic Growth* **1 (1)**, 1–27.

[4] Barro, R., & Lee, J.-W. (1994). Sources of economic growth. *Carnegie-Rochester Conference Series on Public Policy*, **40**, 1–57.

[5] Bloch, F., Jackson, M., & Tebaldi, P. (2017). Centrality measures in networks, working paper.

[6] Caselli, F., Esquivel, G., & Lefort, F. (1996). Reopening the convergence debate: A new look at cross-country growth empirics. *Journal of Economic Growth* **1**, 363–389.

[7] Ciccone, A., & Jarocinski, M. (2010). Determinants of economic growth: Will data tell? *American Economic Journal: Macroeconomics* **2**, 222–246.

[8] Dollar, D., & Kraay, A. (2003). Institutions, trade, and growth. *Journal of Monetary Economics* **50**, 133–162.

[9] Easterly, W. (1993). How much do distortions affect growth? *Journal of Monetary Economics* **32(2)**, 187–212.

[10] Easterly, W., & Levine, R. (1997). Africa's growth tragedy: Policies and ethnic divisions. *Quarterly Journal of Economics* **112(4)**, 1203–1250.

[11] Easterly,W., & Levine, R. (2001). It's not factor accumulation: Stylized facts and growth models. *World Bank Economic Review* **15**, 177–219.

[12] Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., & Mo, H. (2005). World Trade Flows: 1962–2000. *NBER Working Papers* **11040**.

[13] Gallup, J., Mellinger, A., & Sachs, J. (2000). The economic burden of malaria. Harvard University, Center for International Development Working Paper No. 1.

[14] George, E.I., & McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339–373.

[15] Moral-Benito, E. (2013). Likelihood-based estimation of dynamic panels with predetermined regressors. *Journal of Business and Economic Statistics* **31**, 451–472.

[16] Moral-Benito, E. (2016) Growth empirics in panel data under model uncertainty and weak exogeneity. *Journal of Applied Econometrics*, **31**, 584–602.

[17] Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111–163.

[18] Raftery, A. E., & Lewis, S. M. (1992). How many iterations in the Gibbs sampler? *Bayesian Statistics* **4**, 763–773.

[19] Sala-i-Martin, X. (1997a). I just ran 4 million regressions. *NBER, Working Paper No. 6252.*

[20] Sala-i-Martin, X. (1997b). I just ran 2 million regressions. *American Economic Review* **87 (2)**, 178–183.