Arbeitsbereich Ökonomie

IOS Working Papers

No. 360 October 2016

Determinants of export sophistication: Evidence from Monte Carlo simulations

Karen Poghosyan * and Evžen Kočenda**

* Central Bank of Armenia, Economic Research Department, Yerevan, Armenia. Email: karen.poghosyan@cba.am. ** Institute of Economic Studies, Charles University, Opletalova 26, 110 00, Prague, Czech Republic; Institute of Theory of Information and Automation, Czech Academy of Sciences, Prague; CES, Munich; IOS, Regensburg. E-mail: evzen.kocenda@fsv.cuni.cz.

Institut für Ost- und Südosteuropaforschung
Institute for East and Southeast European Studies
Landshuter Straße 4
D-93047 Regensburg

Telefon: (09 41) 943 54-10
Telefax: (09 41) 943 54-27
E-Mail: info@ios-regensburg.de
Internet: www.ios-regensburg.de
ISSN: 2199-9465

Contents

List of Tables

Abstract

In this paper we analyze the determinants of export sophistication based on a large panel dataset (2001–2014; 101 countries) and using different estimation algorithms. Using Monte Carlo simulations we evaluate the bias properties of estimators and show that GMM-type estimators outperform instrumental-variable and fixed-effects estimators. We show that when we apply the panel data over a different period and different set of countries, the findings of Hausmann et al. (2007) remain robust. We provide new evidence of export sophistication path-dependency and confirm that GDP per capita and the size of the economy exert significant and positive effects on export sophistication. Institutional quality positively affects only countries with low institutional quality. The high persistence of export sophistication is also a sign that export diversification promotes not only productivity and sustainable economic growth but also resistance during economic downturns.

JEL-Classification: C52; C53; F14; F47; O19

Keywords: international trade; export sophistication; specialization; dynamic panel data; Monte Carlo simulation; panel data estimators

Research support from GAČR grant No. 16-09190 is gratefully acknowledged.

1 Introduction

There is strong evidence that diversified exports promote productivity and sustainable economic growth in the long run (Hausmann et al., 2011; Hidalgo et al., 2007). In terms of export composition, Hausmann et al. (2007) persuasively show that what you export matters: a country with a higher human-capital level can produce goods at a higher level of productivity ("sophistication"). They also theoretically identify several key determinants to explain the variation of such export sophistication and empirically find that GDP per capita and human capital have a significant effect on the variation of export sophistication. However, when other theoretical factors are added, the results are mixed. Further, we believe that the estimation procedure matters as well. In this paper we extend the above seminal work in three ways. First, since the structure of exports does not change rapidly, we modify the econometric specification of export sophistication to account for path dependency in the structure of exports. Second, we employ several estimators and perform Monte Carlo simulations to obtain the most accurate and effective estimates on determinants of export sophistication. Third, we cover most of the economies in the world over a period that allows assessment before as well as after the global financial crisis.

What are the motivations and motivating details behind our extensions? First, theoretically motivated determinants of export sophistication do not exactly match the empirical estimation of Hausman et al. (2007) in terms of statistical significance. Two key ones are GDP per capita and human capital. However, when other explanatory variables are included in the estimated specification, the results become mixed. When the law of rule index (alone or in combination with population and land area) is added, human capital loses its significance but GDP per capita still impacts export sophistication significantly.

Second, exporting is usually not a one-shot activity. Exports need the development of trade links among partners. Hence, exports and related activities are likely to exhibit path-dependency (Hidalgo et al., 2007; Mehta and Felipe, 2014). Further, Córcoles et al. (2014) show that the risk of destabilization of trade links decreases with the complexity of the involved products. On the other hand, exports do not automatically generate further exports because importers play an important and often dominant role in initiating and affecting international trade transactions and thus choosing their exporting partners (Liang and Parkhe, 1997).[1] To account for such path-dependency we modify the econometric specification of the export sophistication model by including the lagged value of the dependent variable. Beside the path-dependency issue, there

[1] Liang and Parkhe (1997; p. 520) argue that "often it is the importers who drive exports, by choosing exporters and export countries, rigidly specifying the product to be exported, handling all export marketing functions in the import country, and even entering into joint ventures with exporters".

is an additional reason to include lagged values of the dependent variables in the model; it is then possible to solve some econometric problems. In particular, we are able to better deal with the autocorrelation of the disturbances in the panel estimation, with time-invariant country characteristics correlated with explanatory variables, and with some regressors that may be predetermined variables and not strictly exogenous.

Third, from the econometric perspective, the inclusion of lagged values of export sophistication (dependent variable) changes the econometric specification from a static to a dynamic panel-data model. From the modern econometric literature it is known that when lagged values of dependent variables are included in a dynamic panel model, then fixed and random effect estimators become biased (Nickel, 1981). Thus, the inconsistency of the static panel data estimation algorithms leads to the implementation of consistent dynamic panel data estimation algorithms. A number of dynamic panel data estimators have been developed, such as the IV-type estimators in Anderson and Hsiao (1982) and the GMM-type estimators proposed by Arellano and Bond (1991) and Blundell and Bond (1998).[2] To evaluate the performance of the above-mentioned estimators with an application to the export sophistication model we conduct Monte Carlo simulations. In this paper, Monte Carlo experiments are based on the actual panel dataset, which from our point of view can increase the realism of the experiments. Based on the Monte Carlo simulation experiments and using Bias and RMSE criteria we conclude that GMM-type estimators (Arellano-Bond and Blundell-Bond), compared with fixed-effects and IV-type estimators, performs relatively well, even when we increase the coefficient on the lagged dependent variables and the standard deviation of the individual effects of the countries.

Fourth, a topical aspect of our analysis is data coverage: we employ data across 101 countries over the years 2001–2014. Adjustments in our econometric specifications allow isolating the effect of the global financial crisis. A number of papers show that during the global financial crisis the qualitative structure of exports does not change and remains relatively stable (Shelburne, 2010; Da Costa Neto and Romeu, 2011). One can therefore assume that the main determinants continued to have the same effect on export sophistication even after the post-crisis period. Hence we can expect that the global financial crisis did not change the composition of the main determinants that play an important role in the export sophistication level of the countries.

Our contribution points at the use of the most accurate estimators and thereby better evaluation of the key determinants of export sophistication and their parameters; subsequently we compare our estimates with results from Hausmann et al. (2007). Therefore, this paper can be

[2] There is also another class of estimators, particularly the class of direct bias correcting estimators suggested by Kiviet (1995), Hansen (2001), and Bun and Carree (2005). In this paper we do not consider the class of direct bias correcting estimators, which can be an area for future research. Here we concentrate our attention mainly on the evaluation of the bias properties of IV- and GMM-type estimators.

considered an extension based on the new panel dataset and better results from improved estimation algorithms. The paper also represents a methodological contribution to the problem of choosing the right algorithm to estimate an export sophistication dynamic panel data model. In this respect, our results should be of interest to practical macroeconomic policymakers, especially from developing countries. This is because for developing countries there is a large scope for structural transformation of the economy and improvement of export sophistication. The estimated parameters of the export sophistication regression can be used for calibration purposes in the process of developing other structural models. Also, our results should be of interest to practical policy econometricians who engage in model estimation and evaluation. The Monte Carlo simulations that we use in this paper can be implemented in the process of the evaluation of other types of dynamic panel models.

The paper is organized as follows. In section 2 we briefly review the literature related to the researched topic. In section 3, we introduce a modification of the existing specification for the export sophistication regression and provide an overview of the IV- and GMM-type estimators employed. In section 4 we present the data, descriptive statistics, and comparative tables. We present our estimation results in section 5. These are followed by the Monte Carlo simulation results and discussion on the optimal estimator in section 6. Brief conclusions are summarized in section 7.

2 Literature review

The link between the nature of exports and the performance of open economies is an important empirical question. In their seminal paper, Hausmann et al. (2007) show that the composition of exports determines the level of export sophistication, which indicates the similarity of export bundles of a country with exports of high income countries. They also argue that the mix of goods that a country produces may have important implications for economic growth. Based on a theoretical model, they demonstrate this proposition formally and to some extent support it empirically; their index of the "income level of a country's exports" is shown to predict subsequent economic growth. Their index is used to measure how a country with a higher human-capital level can produce goods of higher productivity ("sophistication"). The measurement of export sophistication has attracted the attention of other researchers as well. Much earlier, Michaely (1984) constructed a similar index ("income level of exports") in which a different weighting scheme disproportionally favored large countries. More recently, Lall et al. (2005) developed another similar measure to assess the "sophistication level of exports."

The approach of Hausmann et al. (2007) received justified attention because it offered a theoretical structure to explain export sophistication along with an adequate empirical treatment. GDP per capita, human capital, the rule of law index, population, and land area were identified as potential determinants for the explanation of export sophistication variation across countries. Formally, the productivity level associated with the export basket of a specific country (EXPY; export sophistication) was regressed on one or more of the above determinants. Subsequently, four different models for export sophistication were estimated with different sets of theoretically motivated explanatory variables. In the first model the log of GDP per capita was used as an explanatory variable showing that a 1% change in GDP per capita can cause a 0.354% change in the export sophistication index. In the second model, with the log of GDP per capita and the log of human capital, the estimation shows that both variables have a positive and significant impact on export sophistication: 0.298 and 0.281, respectively. In the third model with three explanatory variables (log of GDP per capita, log of human capital, and the rule of law index), all three variables are positively correlated with the export sophistication index and only the log of GDP per capita has a significant effect on the dependent variable, while the other two explanatory variables are not significant. In the fourth model, five explanatory variables were included. According to the estimation results the log of GDP per capita (0.282) and the log of population (0.089) have positive and statistically significant effects on the export sophistication index while the log of land area (–0.032) exhibits a negative effect; the coefficients of the other two variables (log of human capital and the rule of law index) are positive but statistically insignificant.

Zhu et al. (2010) further explore the idea of export sophistication and regress the EXPY variable on an extended set of explanatory variables. The first group includes variables that are related to a country's natural resources (capital-labor ratio and land area per capita). The second group includes variables that are related to human capital (gross tertiary enrollment and proportion of R&D expenditure in GDP). The third group of variables is related to foreign direct investment (FDI), economy size (population), and country institutional quality (rule of law index). Like Hausmann et al. (2007), in this paper the log of land area has a negative impact on EXPY and the capital-labor ratio has a significant and positive impact. The relationship between EXPY and human capital is significant and positive. For example the effect of education is significant in the low-income country group, while the effect of R&D is significant in high income countries. Population size has a significant and positive impact on EXPY both in high- and low-income countries. The institutional quality has a negative effect on EXPY both in high- and low-income countries. In low-income countries it has a significant effect. According to this paper, the export sophistication of countries is enhanced by capital intensity and an engagement in knowledge creation and transfer via investment in education, R&D, foreign direct investment, and imports. On the other hand, the effect of natural resources on the export sophistication level depends on the quality of the institutions in a particular country. That is, if the particular country has effective institutions, then there could be a positive effect of natural resources on export sophistication and vice versa.

Other researchers amended the issue of export sophistication research with additional contributions. Cabral and Veiga (2010) find that GDP per capita and the size of the economy are positively correlated with EXPY. Also, they found that improvements in institutional, political, and educational factors may play an important role in enhancing better export sophistication in Sub-Saharan Africa. Further, they show that a high level of corruption is an important factor in limiting the level of export sophistication. Finally, increases in human capital are found to be positively correlated with export sophistication. Anand et al. (2012) indicate that the relationship between GDP growth and export sophistication is significant and positive. Overall, their results indicate that an educated workforce, external liberalization, and good information flows are all significantly associated with a high level of export sophistication across a broad range of different specifications.

The most recent contribution is Weldemicael (2014), which explores the relative importance of technology and trade costs on export sophistication and welfare in a general equilibrium framework. The results show that GDP per capita, human capital, and country size maintain their significant and positive impact on EXPY. In addition to this, lagged EXPY has a significant positive effect on the current value, and according to this paper it has a dominant effect in comparison with other explanatory and control variables. Using cross-country panel data in the paper, it was shown that foreign direct investment has a positive effect and the effect is greater for countries with low institutional quality. From the other side, the remoteness (distance) from main markets has a strong negative effect on export sophistication. Regarding institutional quality, its effect on export sophistication is low and insignificant.

Thus in the current paper we extend Hausmann et al. (2007) and differentiate our contribution from the existing research in three ways. First, since the structure of exports does not change rapidly, we modify the econometric specification of export sophistication to account for path dependency in the structure of exports. Second, we employ several estimators and perform Monte Carlo simulations to obtain the most accurate estimates of the determinants of export sophistication. Third, we cover most of the economies in the world over a period that allows assessment before as well as after the global financial crisis.

3 Model modification and estimation algorithms

In this paper we aim to assess the robustness of the determinants introduced by Hausman et al. (2007): whether the determinants have the same effect (in terms of sign and significance) on export sophistication after modifying the model specification. To do this, we aim to obtain the most accurate estimates by performing estimations via suitable estimation procedures. The procedures are briefly introduced below and more formally in the Technical Appendix. As a second step we conduct Monte Carlo simulations to investigate the performance of the different estimation algorithms in terms of minimizing bias and root mean square error (RMSE); see section 6 for details.

3.1 Econometric model

A preliminary conclusion based on Hausman et al. (2007) is that the log of GDP per capita is a robust determinant of export sophistication and its values remain stable through different model specifications. The coefficients of other determinants change their value and statistical significance depending on what variables are included in the econometric specification. An important observation is that the constant is large and statistically significant. Further, when the initial level of export sophistication is added to the regression, the coefficient of this variable is also relatively large and statistically significant. Both cases indicate that export sophistication might exhibit an important degree of path dependency. This is a reasonable assumption because exporting activities and exports themselves are usually longer-term projects. Exports are to some extent costly because they require market exploration and the development of trade links among partners. However, exports do not automatically generate further exports because exporters are not necessarily always the driving force behind international trade transactions. Evidence shows that much international exchange is better conceptualized as buyer-coordinated importing rather than producer-initiated exporting (Liang and Parkhe, 1997). Still, in both cases stable international trade activities mitigate the costs (Hidalgo et al., 2007; Mehta and Felipe, 2014) and such stability results in a certain degree of path dependency.

We account for theoretically as well as empirically motivated path dependency by including the lagged value of the dependent variable (export sophistication). Hence, the original specification of Hausman et al. (2007) is modified in the following way:

$$\begin{aligned}\ln(EXPY_{it}) &= \beta_0 + \beta_1 \ln(EXPY_{it-1}) + \beta_2 \ln(GDPpc_{it}) + \beta_3 (HC_{it}) + \beta_4 (RofL_{it}) \\ &+ \beta_5 \ln(POP_{it}) + \beta_6 \ln(Area_{it}) + \eta_i + \nu_{it}\end{aligned}, \qquad (1)$$

where $\ln(EXPY_{it})$ is the logarithm of the export sophistication index that is formally defined in full detail in section 4, $\ln(GDPpc_{it})$ is the logarithm of GDP per capita, HC_{it} is a measure of human capital, $RofL_{it}$ is the rule of law index, $\ln(POP_{it})$ is the logarithm of population, and $\ln(Area_{it})$ is the logarithm of land area. Subscript *i* denotes countries and subscript *t* denotes time periods (years).

We hypothesize that the lagged value of export sophistication will have a positive and significant effect on the current value due to the path-dependency of exporting activities. Further, the inclusion of the lagged value of the dependent variable enables us to solve some econometric problems. Specifically, we are able to better deal with: (i) the autocorrelation of disturbances in the panel estimation, (ii) time-invariant country characteristics correlated with explanatory variables, and (iii) some regressors that may be predetermined variables and not strictly exogenous.

3.2 Estimation algorithms and estimators

The inclusion of the lagged values of export sophistication (dependent variable) changes the econometric specification from a static to a dynamic panel data model and fixed- and random-effect estimators become biased (Nickel, 1981). The inconsistency of the within-group estimator has led to the development of consistent dynamic panel data estimation algorithms. In modern panel data econometrics there are a number of algorithms that allow us to consistently estimate this parameter.

As alternative estimation algorithms we use a static panel data estimation algorithm (fixed effect; FE) and different dynamic panel data estimation algorithms, particularly IV-type estimator (Anderson-Hsiao level and difference), GMM (Arellano-Bond), and system GMM (Blundell-Bond) algorithms. The algorithms are formally described in the Technical Appendix.

All the above-mentioned algorithms are used to estimate the export sophistication panel regression. But as mentioned above, we are interested not only in estimation but also in comparison of FE-, IV-, GMM-, and System GMM-type estimators. A proper comparison can be done through a simulation experiment, where we know the data-generating process (DGP) and can therefore relate the estimates to the true values. The simulations and their outcomes are presented in section 6.

4 Data and descriptive statistics

Our working balanced panel dataset consists of six macroeconomic variables on a yearly frequency from 2001 to 2014; in total we have 1414 observations per each macroeconomic variable (101 countries times 14 years). Our panel includes both the pre- and post-crisis periods. This should be important when we compare our estimates with those of Hausmann et al. (2007).[3] The data include a measure of export sophistication (the dependent variable that is defined presently) and five macroeconomic determinants: GDP per capita, human capital, a rule of law index, population, and land area. All of the variables, with the exception of human capital and the rule of law index, are in logarithms. All variables correspond to those used in Hausman et al. (2007).

The GDP per capita (GDPpc) is the gross domestic product converted to international dollars using purchasing power parity rates in order to provide a comparable perspective. Human capital is proxied by the gross tertiary enrolment ratio (both sexes) in percent. A high ratio indicates a high degree of current tertiary education and a higher level of human capital in the economy.[4] Data on population (POP), as a proxy for the size of the economy/market, are reported in millions of inhabitants. Data on the area (AREA) are reported in square kilometers.

All of the variables above were obtained from the World Development Indicators (WDI) database of the World Bank. The Rule of Law index data (RofL) are collected from the World Bank's Worldwide Governance Indicators database (WGI). This database reports the rule of law, government efficiency, and other indices of institutional quality in the years of 1996, 1998 and 2000–2014. The rule of law index is commonly used to characterize institutional quality. The index ranges from –2.5 to 2.5 and a higher value represents better governance.

The measure of export sophistication (denoted as EXPY) is defined as the average income associated with a country's export bundle. We follow Hausman et al. (2007) and construct the export sophistication index in two steps. First, we compute the productivity level associated with each product separately. Second, we compute the average productivity level that corresponds to a country's total export basket.

Formally, let's assume that X_i^k represents the exports of product *k* from country *i*. Then, the total export of country i is $X_i = \sum_k X_i^k$. The income (productivity) level (PRODY) associated with each product *k* in the export basket is then calculated as:

[3] We formally check for the presence of a structural break and provide more details in section 4.

[4] Our data set originally consisted of 112 countries. However, the data for the gross tertiary enrollment ratio are not available for 11 countries out of the 112 for which export sophistication is possible to construct. Therefore, and in order to work with a balance panel, we reduced the size of the data set slightly (112 – 11 = 101) and perform the estimation and simulation on a balanced panel of 101 countries.

$$PRODY^k = \sum_i \left\{ \frac{\left(X_i^k / X_i\right)}{\sum_i \left(X_i^k / X_i\right)} Y_i \right\},$$

where Y_i denotes the GDP per capita of country i.

In the second step, we calculate the average productivity level that corresponds to a country's total export basket (EXPY) as

$$EXPY_i = \sum_k \left\{ \frac{X_i^k}{X_i} PRODY^k \right\}.$$

In order to calculate the export sophistication index, the export data from the International Trade Center (ITC, http://www.intracen.org/) database have been used. The HS02 4-digit-level classification incorporating 1258 products were used as the basis. The value of exports is measured in thousands of current U.S. dollars. The number of countries that report trade data vary considerably from year to year. However, we construct the PRODY and EXPY measures for a balanced sample of 101 countries that report trade data in each year during 2001–2014.

In Table 1 we report summary statistics for export sophistication (EXPY) dynamics. As we can see there is a high variation in EXPY among countries as evidenced by the high standard deviation (Table 1, column 4). During the period under research EXPY increased significantly. In column 3 we report the average export sophistication values for all countries included in the sample (101 countries). As we can see, the EXPY value increased steadily from 13831.4 USD in 2001 to 23168.0 USD in 2014, with an average 4.0% growth per year calculated by geometric mean. On the other hand, the difference between the minimum and maximum values of EXPY also increased, particularly from 22128.9 USD in 2001 to 37412.4 USD in 2013 (Table 1, column 7).

Table 1: Descriptive statistics for EXPY (USD)

Year	Obs.	Mean	SD	Minimum	Maximum	Range
1	2	3	4	5	6	7
2001	101	13831.4	4936.0	3559.3	25688.2	22128.9
2002	101	14361.8	5252.8	3436.3	28008.9	24572.7
2003	101	14847.3	5269.5	3210.1	25723.0	22512.8
2004	101	15843.8	5678.5	4505.4	28138.1	23632.7
2005	101	16616.4	5626.6	4440.0	27684.3	23244.3
2006	101	18151.4	6063.7	4661.9	29678.5	25016.6
2007	101	19331.5	6385.9	5407.2	31686.1	26278.8
2008	101	20185.5	6906.5	4674.2	34976.1	30301.9
2009	101	19776.8	6504.3	4617.4	35576.1	30958.7
2010	101	20593.1	6598.5	5391.9	37882.5	32490.6
2011	101	21580.1	7107.7	6472.6	41909.5	35436.9
2012	101	22055.6	7164.2	5234.1	42210.1	36976.0
2013	101	22637.6	7174.9	5974.2	43386.6	37412.4
2014	101	23168.0	6774.1	7281.9	44594.9	37313.0

Using data from columns 5 and 6 of Table 1 we calculate that the minimum value of EXPY increased on average by 5.7%, while the maximum value of EXPY increased on average by 4.3% (both calculated by geometric mean). This is an indication that the EXPY value for some low-income countries grows during 2001–2014 more rapidly than in high- and middle-income countries. This is quite an optimistic observation as it suggests that some countries with a low level of export sophistication continually increased their production and export diversification and as a result their export sophistication steadily increased as well.

Further, in Table 2 we report countries with the smallest and largest EXPY values at the beginning, middle, and end of the sample period. The lowest values of EXPY are recorded by countries in Africa. On the other hand in 2001 only one European country (Luxembourg) and four oil exporting countries (Saudi Arabia, Oman, Algeria, and Qatar) were on the list with largest EXPY. But in 2014 already three European countries (Norway, Ireland, and Luxembourg) and only two oil exporting countries (Algeria and Qatar) were on the list with largest EXPY. Most oil-exporting countries are highly dependent on oil exports and therefore their EXPY is highly dependent on oil prices.

Table 2: List of countries with smallest and largest EXPY

Year	No	Country	EXPY	Country	EXPY
2001	1	Ethiopia	3559.3	Saudi Arabia	22018.3
	2	Burundi	3921.3	Luxembourg	22437.9
	3	Malawi	4067.4	Oman	22710.4
	4	Rwanda	4395.3	Algeria	25038.4
	5	Cambodia	4569.3	Qatar	25688.2
2007	1	Burundi	5407.2	Oman	28284.0
	2	Malawi	5626.3	Ireland	29062.1
	3	Cote d'Ivoire	5910.7	Luxembourg	29408.6
	4	Niger	6100.4	Algeria	30208.7
	5	Ethiopia	6330.8	Qatar	31686.1
2014	1	Burundi	7281.9	Ireland	32256.5
	2	Rwanda	7609.0	Norway	34570.0
	3	Malawi	9251.8	Luxembourg	36357.7
	4	Tanzania	10607.7	Algeria	36798.6
	5	Madagascar	10855.1	Qatar	44594.9

5 Estimation results

We estimate an export sophistication model with different estimation algorithms, particularly fixed-effects (FE), Anderson-Hsiao (with level and difference instruments AH-L and AH-D), Arellano-Bond GMM (GMM1), and Blundell-Bond system GMM (GMM1-SYS) estimators. Prior to estimation, we formally check for the existence of a structural break in the dynamics of the export sophistication data. We perform a series of Chow tests for each country in our dataset over the period 2001–2014. The results (not reported but available upon request) suggest that there is no structural break in the export sophistication dynamics. Hence, we estimate specification (1) without adjustments for a structural break.

Estimation results for the export sophistication dynamic panel model are presented in Table 3 for the whole sample of available data. Based on the results in Table 3, the lagged value of EXPY has a significant and positive effect on the current value of EXPY. The results based on specific procedures (i.e. FE, Anderson-Hsiao (Level), and system GMM) even have a dominant effect in comparison with other explanatory variables. The estimated coefficients of GDP per capita are positive and statistically significant. As we can see from Table 3 the results are robust across all estimation algorithms. Concerning the HC variable we see that the estimated coefficient is close to zero and its value is statistically insignificant. The coefficient of HC is small, which tells us that the causal effect may go from EXPY to human capital and not vice versa, which corresponds to the argument made by Hausmann et al. (2007). The index of institutional quality, which is proxied by the rule of law index, is also close to zero and its value is statistically insignificant (with the exception of the system GMM estimator). The estimated coefficients of country size are positive and statistically significant for almost all estimation algorithms (with the exception of Anderson-Hsiao (Level)). This is consistent with the findings of Hausmann, et al. (2007), who proxy country size using population. On the other hand, this result also corresponds to the argument that the number of horizontal varieties produced by a country is a function of its economic scale (Krugmann, 1980; Scott, 2008; Hummesl and Klenow, 2005). The last explanatory variable is the land area, which is negatively correlated with EXPY. The estimated parameter is statistically insignificant (with the exception of the system GMM). We see that this variable is not well associated with EXPY dynamics. This is because the land area of a majority of the countries does not change over the years and therefore has no variation, while EXPY dynamics can be characterized by a relatively high level of variation (see Table 1).

Further, we also estimate the relationship between EXPY and the explanatory variables after dividing the whole sample into two sub-samples using the median of the rule of law index. Similarly as in Weldemicael (2014), we have divided countries into two groups according to the rule of law index level to show the effect of institutions on EXPY dynamics. This is because there is a prevailing hypothesis that weaker institutions are associated with slow growth or poor economic performance and vice versa (Mauro, 1995). From Table 3 we see that for the whole sample of countries the effect of institutional quality does not have a significant effect but it has a negative sign. This is because in the whole sample there are many countries with high EXPY but a relatively small institutional quality score. That is why in order to investigate the effect of institutions on the export sophistication level, we have divided the whole sample into two subsamples using the median of the rule of law index and then examined the effect of export path-dependency, GDP per capita, and the size of the economy on the export sophistication.

Table 3: Estimation results (Dependent variable Ln(EXPY))

	Fixed effects	Anderson-Hsiao (Level)	Anderson-Hsiao (Difference)	GMM step one	System GMM step one
$Ln(EXPY)_{it-1}$	0.579***	1.022***	0.249*	0.382***	0.717***
	(24.46)	(3.32)	(1.62)	(10.70)	(28.76)
$Ln(GDPpc)_{it}$	0.247***	0.289***	0.439***	0.469***	0.211***
	(11.53)	(2.77)	(6.34)	(13.96)	(8.61)
HC_{it}	0.001	–0.002	–0.001	–0.001	–0.001
	(1.10)	(–1.31)	(–1.27)	(–1.31)	(–1.31)
$RofL_{it}$	–0.008	–0.032	–0.004	0.001	–0.119***
	(–0.46)	(–0.73)	(–0.13)	(0.01)	(–4.50)
$Ln(POP)_{it}$	0.262***	0.298	0.343**	0.346***	0.039**
	(7.28)	(1.22)	(1.99)	(6.88)	(2.18)
$Ln(AREA)_{it}$	–1.274	–3.117	–2.100	–2.355	–0.086***
	(–1.49)	(–0.78)	(–0.71)	(–1.25)	(–4.47)
Number of obs.	1414	1414	1414	1414	1414
Number of groups	101	101	101	101	101

Note: t-statistics are in parentheses. ***, **, and * indicate significance level at the 1%, 5%, and 10% levels, respectively. Subscript i represents the country, t denotes the period, and *Ln* represents the natural logarithm of the corresponding variables. The variable EXPY refers to export sophistication, GDPpc represents GDP per capita, HC is a measure of human capital, RofL is the rule of law index, POP is population, and AREA is land area.

For the low institutional quality country group, the estimated coefficient of the lagged value of EXPY, GDP per capita, and population are significant and positive, while the coefficients of HC, the rule of law index, and land area are not statistically significant (Table 4). We should mention that for the countries that show a positive effect of the rule of law index on export sophistication, the coefficient is not statistically significant. For the high institutional quality group (Table 5) we see that the lagged value of EXPY, GDP per capita, and the size of the economy have significant and positive effects on export sophistication. Thus, we conclude that these three variables play important roles in export sophistication both for low and high institutional quality countries. Therefore, those variables can be considered potential determinants of export sophistication. Concerning HC, we can see that similarly as for low institutional quality countries, this variable also has no effect on export sophistication for high institutional quality countries and its value is almost equal to zero. On the other hand, it is interesting that for this group of countries the rule of law index has a significant and negative effect on export sophistication (Table 5). This is because high institutional quality level countries have already reached a relatively high level of governance and the future improvement of the rule of law might not represent an improvement per se, but rather increased bureaucracy with a potentially negative effect on trade. The land area exhibits a negative but statistically insignificant effect on the current value of EXPY. This is because this variable is stable over time, while EXPY varies from year to year.

Table 4: Estimation results, countries with low institutional quality (Dependent variable Ln(EXPY))

	Fixed effects	Anderson-Hsiao (Level)	Anderson-Hsiao (Difference)	GMM step one	System GMM step one
$Ln(EXPY)_{it-1}$	0.527^{***}	0.988^{**}	0.253	0.333^{***}	0.658^{***}
	(14.58)	(2.46)	(1.33)	(6.36)	(17.43)
$Ln(GDPpc)_{it}$	0.219^{***}	0.222	0.382^{***}	0.428^{***}	0.254^{***}
	(6.85)	(1.26)	(3.19)	(8.54)	(6.89)
HC_{it}	0.001	–0.003	–0.002	–0.001	–0.002
	(1.13)	(–0.95)	(–0.87)	(–0.85)	(–1.24)
$RofL_{it}$	0.045^{*}	–0.005	0.025	0.031	0.045
	(1.68)	(–0.07)	(0.49)	(0.72)	(0.98)
$Ln(POP)_{it}$	0.571^{***}	0.527	0.625	0.636^{***}	0.027
	(7.22)	(0.92)	(1.50)	(6.02)	(0.79)
$Ln(AREA)_{it}$	$–22.072^{*}$	–10.282	–5.669	–23.851	–0.024
	(–1.63)	(–0.29)	(–0.21)	(–1.03)	(–1.24)
Number of obs.	714	714	714	714	714
Number of groups	51	51	51	51	51

Note: t-statistics are in parentheses. ***, **, and * indicate significance level at the 1%, 5%, and 10% levels, respectively. Subscript i represents the country, t denotes the period, and *Ln* represents the natural logarithm of the corresponding variables. The variable EXPY refers to export sophistication, GDPpc represents GDP per capita, HC is a measure of human capital, RofL is the rule of law index, POP is population, and AREA is land area.

Thus, based on the estimation results for both the whole sample and two sub-samples we conclude that the sign and significance of the estimated parameters almost coincide with those presented in Hausmann et al. (2007). However, the application of several estimation algorithms for the export sophistication regression does not provide information whether one estimation algorithm is superior. In other words, based just on the estimation results we cannot conclude which algorithm delivers the most accurate estimates. To answer this question we need a Monte Carlo-based comparison among the employed estimation algorithms where we know the data-generating process (DGP) and can therefore relate the estimates to the truth.

Table 5: Estimation results, countries with high institutional quality (Dependent variable Ln(EXPY))

	Fixed effects	Anderson-Hsiao (Level)	Anderson-Hsiao (Difference)	GMM step one	System GMM step one
$Ln(EXPY)_{it-1}$	0.604^{***}	0.546^{***}	0.275	0.405^{***}	0.721^{***}
	(21.49)	(2.81)	(0.63)	(12.73)	(26.08)
$Ln(GDPpc)_{it}$	0.268^{***}	0.450^{***}	0.511^{***}	0.450^{***}	0.173^{***}
	(9.68)	(6.08)	(5.56)	(11.82)	(6.36)
HC_{it}	0.001	–0.001	–0.001	–0.001	0.001
	(0.62)	(–0.93)	(–0.84)	(–1.05)	(1.17)
$RofL_{it}$	$–0.064^{***}$	–0.050	–0.042	$–0.088^{***}$	$–0.085^{***}$
	(–2.91)	(–1.43)	(–1.07)	(–2.67)	(–2.68)
$Ln(POP)_{it}$	0.142^{***}	0.215	0.234^{*}	0.155^{***}	–0.001
	(4.83)	(1.49)	(1.71)	(4.22)	(–0.06)
$Ln(AREA)_{it}$	–0.705	–2.153	–1.584	–0.660	–0.040
	(–1.21)	(–1.06)	(–0.78)	(–0.64)	(–4.35)
Number of obs.	700	700	700	700	700
Number of groups	50	50	50	50	50

Note: t-statistics are in parentheses. ***, **, and * indicate significance level at the 1%, 5%, and 10% levels, respectively. Subscript i represents the country, t denotes the period, and *Ln* represents the natural logarithm of the corresponding variables. The variable EXPY refers to export sophistication, GDPpc represents GDP per capita, HC is a measure of human capital, RofL is the rule of law index, POP is population, and AREA is land area.

6 Monte Carlo simulations

While the previous results are of practical interest, a proper comparison of the above-mentioned estimation algorithms can be done using a Monte Carlo simulation. The purpose of the Monte Carlo simulations is to investigate the performance of the GMM-based estimators versus instrumental variables and fixed-effects estimators within a realistic set-up. The DGP that we have chosen closely follows the model estimated in the previous section:

$$\ln(EXPY_{it}) = \beta_1 \ln(EXPY_{it-1}) + \beta_2 \ln(GDPpc_{it}) + \beta_3(HC_{it}) + \beta_4(RofL_{it})$$
$$+ \beta_5 \ln(POP_{it}) + \beta_6 \ln(AREA_{it}) + \eta_i + \nu_{it}$$

$$\eta_i \sim N(0, \sigma_\eta^2),\ \nu_{it} \sim N(0, \sigma_\nu^2)$$

$$\ln(GDPpc_{it}) = \mu(GDPpc_{it-1}) + \xi_{it}^{GDPpc_{it}}$$

$$\ldots$$

$$\ln(AREA_{it}) = \mu(AREA_{it-1}) + \xi_{it}^{AREA_{it}}$$

$\xi_{it}^c \sim N(0,1)$, where $c = (GDPpc, HC, RofL, POP, AREA)$, $\mu = 0.90$.

In order to design simulation experiments we closely follow the methodological approach of Santos and Barrios (2012). In our study the number of the cross-section is $N = 101$, while the time dimension is $T = 14$. In total, we have 1414 observations; we work with one dependent variable (EXPY) and five explanatory variables. We assume that we know the true values of the parameters $\beta = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, \beta_6^*)$. In Table 6 we summarize our Monte Carlo design and present the parameter values in the data-generating process for the export sophistication model.

As we can see from Table 6 we always keep parameters $\beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, \beta_6^*$ fixed while the value of parameter β_1 is allowed to vary between 0.1 and 0.9, in our case $\beta_1 = \{0.1,\ 0.2,\ 0.5,\ 0.8,\ 0.9\}$. The variation of the parameter on the lagged dependent variable is chosen based on the strategy in Santos and Barrios (2012). We allow only the parameter of the lagged dependent variable (EXPY) to vary because we are interested in finding an unbiased estimate for this parameter. Instead of analyzing the effect of σ_ν^2 and σ_η^2 separately, we focus on the variance ratio $\sigma_\eta^2 / \sigma_\nu^2$. When $\sigma_\eta^2 = 0$, the values for the variance of the individual effects account for the fixed effects; when $\sigma_\eta^2 \neq 0$ then the values of the variance of individual effects account for random effects (Santos and Barrios, 2012). The restrictions above can be explained as follows. To be close to one of the opposite estimates depends on the parameter θ , which is equal to $\theta = 1 - \sqrt{\dfrac{\sigma_\nu^2}{\sigma_\nu^2 + T\sigma_\eta^2}}$. Thus, if $\sigma_\eta^2 = 0$, then $\theta = 0$ and the overall estimate will be close to the fixed effects estimate; if $\sigma_\eta^2 \to \infty$, then $\theta \to 1$ and the overall estimate will be close to the OLS estimate. Therefore, when $\sigma_\eta^2 \neq 0$ then our estimate moves from a fixed- to a random-effects estimate. Also from Table 6 we can see that for each value of the ratio $\sigma_\eta^2 / \sigma_\nu^2$ we have five possible combinations of $\mathrm{B} = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, \beta_6^*)$. This is because the parameters $\beta_2^*, \beta_3^*, \beta_4^*, \beta_5^*, \beta_6^*$ are always fixed, while the parameters for β_1 can take the following five possible values: $\beta_1 = \{0.1,\ 0.2,\ 0.5,\ 0.8,\ 0.9\}$.

Table 6: Monte Carlo designs

$\sigma_\eta^2 / \sigma_\nu^2$	N	T	β1	β2	β3	β4	β5	β6
0.0	101	14	0.1	0.3	0.05	0.03	0.15	0.05
0.0	101	14	0.2	0.3	0.05	0.03	0.15	0.05
0.0	101	14	0.5	0.3	0.05	0.03	0.15	0.05
0.0	101	14	0.8	0.3	0.05	0.03	0.15	0.05
0.0	101	14	0.9	0.3	0.05	0.03	0.15	0.05
1.0	101	14	0.1	0.3	0.05	0.03	0.15	0.05
1.0	101	14	0.2	0.3	0.05	0.03	0.15	0.05
1.0	101	14	0.5	0.3	0.05	0.03	0.15	0.05
1.0	101	14	0.8	0.3	0.05	0.03	0.15	0.05
1.0	101	14	0.9	0.3	0.05	0.03	0.15	0.05
2.0	101	14	0.1	0.3	0.05	0.03	0.15	0.05
2.0	101	14	0.2	0.3	0.05	0.03	0.15	0.05
2.0	101	14	0.5	0.3	0.05	0.03	0.15	0.05
2.0	101	14	0.8	0.3	0.05	0.03	0.15	0.05
2.0	101	14	0.9	0.3	0.05	0.03	0.15	0.05
3.0	101	14	0.1	0.3	0.05	0.03	0.15	0.05
3.0	101	14	0.2	0.3	0.05	0.03	0.15	0.05
3.0	101	14	0.5	0.3	0.05	0.03	0.15	0.05
3.0	101	14	0.8	0.3	0.05	0.03	0.15	0.05
3.0	101	14	0.9	0.3	0.05	0.03	0.15	0.05
5.0	101	14	0.1	0.3	0.05	0.03	0.15	0.05
5.0	101	14	0.2	0.3	0.05	0.03	0.15	0.05
5.0	101	14	0.5	0.3	0.05	0.03	0.15	0.05
5.0	101	14	0.8	0.3	0.05	0.03	0.15	0.05
5.0	101	14	0.9	0.3	0.05	0.03	0.15	0.05

Note: For each possible value of $\sigma_\eta^2 / \sigma_\nu^2$ we have five scenarios for parameter β_1. Thus, in total we have 25 possible scenarios for β_1.

As was mentioned above, our experiments are based on actual data and not on generated data. Based on the actual database we estimated a preliminary static fixed-effects model and then calculated the estimated residual variance. Thus, having a distribution for ν_{it} as

$\nu_{it} \sim N(0,\sigma_\nu^2)$, we are able to generate ν_{it} residuals of size $N \times T$. Also, we can generate N individual effects, because we know the distribution for $\eta_i \sim N(0,\sigma_\eta^2)$. We can generate N individual effects $\eta_i \sim N(0,\sigma_\eta^2)$ by choosing one possible value for the ratio $\sigma_\eta^2/\sigma_\nu^2$ from Table 6. Then given the data-generating process and the values of the regressors, we generated a set of dynamic panel data for EXPY. As an initial value for each cross-section we use the first actual value of the dependent variable of each cross-section.

To evaluate the bias properties of the above-mentioned five estimators we perform 1000 Monte Carlo replications in such a way that we create 1000 panel datasets for each 25 parameter combinations separately. Then, we compute the means of the resulting estimates and compare them to the known true parameters. The difference between the mean estimates and the corresponding true values gives us a measure of (in)accuracy for each estimate of the slope parameters of β_1. We evaluate the accuracy with two criteria—Bias and RMSE—that are defined as:

$$BIAS(\beta_i) = \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_i^r - \beta_i^{true}\right)$$

and

$$RMSE(\beta_i) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\beta}_i^r - \beta_i^{true}\right)^2},$$

where r denotes the number of replications.

In our simulation experiments we allow both explanatory variables and their coefficients to vary. This means that each time we re-estimate the parameters for the explanatory variables. In this way we can check the accuracy properties of the five estimators not only for the lagged dependent variable, but also for all of the parameters that we want to estimate. All simulations were carried out using estimation routines written in the MATLAB (2013a) package.[5] The results of the simulations are presented in Tables 7–11.

As mentioned above we are interested in finding the unbiased estimate for parameter β_1. We report the Monte Carlo simulations on the FE, AH-L, AH-D, GMM1, and GMM1_SYS estimators for a wide range of values of β_1 and $\sigma_\eta^2/\sigma_\nu^2$. The main focus of the analysis is on the bias of the FE, AH-L, AH-D, GMM1, and GMM1_SYS estimators, when we increase the variance of the individual effects of the countries. That is why in our analysis we vary the ratio $\sigma_\eta^2/\sigma_\nu^2$ from 0 (fixed effects) to 5 (random effects). From the other side, in order to see the change in bias as the true value of parameter β_1 varies, we allow the values of β_1 to vary between 0.1 and 0.9, at the same time keeping the value of the ratio $\sigma_\eta^2/\sigma_\nu^2$ fixed. Now let's explain the results of the Monte Carlo simulations presented in Tables 7–11.

[5] The MATLAB codes for Monte-Carlo simulations can be provided upon request.

Table 7: Simulation results, T = 14, N = 101, $\sigma_\eta^2/\sigma_\nu^2 = 0.00$, $\mu = 0.9$

Parameters true values	Estimation algorithm									
	FE		AH-L		AH-D		GMM1		GMM1-SYS	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1 = 0.10	–0.023	*0.024*	–0.067	*0.067*	–0.135	*0.138*	–0.021	*0.022*	–0.032	*0.033*
β_2 =0.30	0.020	*0.020*	0.007	*0.008*	–0.001	*0.005*	0.021	*0.021*	0.024	*0.025*
β_3 =0.05	0.009	*0.010*	0.001	*0.004*	0.001	*0.004*	0.011	*0.012*	0.017	*0.017*
β_4 =0.03	0.001	*0.003*	0.000	*0.003*	–0.001	*0.003*	0.000	*0.003*	0.000	*0.004*
β_5 =0.15	0.006	*0.007*	0.003	*0.005*	–0.001	*0.004*	0.007	*0.008*	0.008	*0.009*
β_6 =0.05	0.000	*0.003*	0.001	*0.004*	–0.001	*0.004*	0.000	*0.004*	–0.002	*0.005*
β_1 =0.20	–0.039	*0.039*	–0.122	*0.122*	–0.280	*0.281*	–0.036	*0.037*	–0.054	*0.055*
β_2 =0.30	0.033	*0.034*	0.016	*0.016*	–0.004	*0.007*	0.037	*0.037*	0.042	*0.042*
β_3 =0.05	0.015	*0.016*	0.003	*0.005*	0.002	*0.004*	0.018	*0.019*	0.028	*0.029*
β_4 =0.03	0.002	*0.003*	0.000	*0.004*	–0.001	*0.004*	0.000	*0.004*	0.000	*0.006*
β_5 =0.15	0.009	*0.010*	0.008	*0.009*	–0.002	*0.005*	0.011	*0.012*	0.011	*0.013*
β_6 =0.05	–0.001	*0.004*	0.002	*0.004*	–0.002	*0.005*	–0.001	*0.005*	–0.004	*0.008*
β_1 =0.50	–0.125	*0.125*	–0.226	*0.227*	–0.759	*0.761*	–0.123	*0.124*	–0.139	*0.140*
β_2 =0.30	0.060	*0.060*	0.048	*0.048*	–0.016	*0.018*	0.069	*0.069*	0.069	*0.070*
β_3 =0.05	0.025	*0.026*	0.008	*0.010*	0.006	*0.008*	0.030	*0.031*	0.046	*0.047*
β_4 =0.03	0.004	*0.006*	0.001	*0.006*	–0.004	*0.006*	0.001	*0.008*	0.001	*0.010*
β_5 =0.15	0.013	*0.014*	0.022	*0.023*	–0.007	*0.009*	0.016	*0.018*	0.010	*0.015*
β_6 =0.05	–0.003	*0.007*	0.006	*0.009*	–0.007	*0.009*	0.000	*0.008*	–0.008	*0.014*
β_1 =0.80	–0.300	*0.300*	–0.293	*0.294*	–1.313	*1.317*	–0.305	*0.305*	–0.317	*0.318*
β_2 =0.30	0.086	*0.086*	0.084	*0.085*	–0.033	*0.036*	0.104	*0.104*	0.094	*0.095*
β_3 =0.05	0.037	*0.038*	0.012	*0.016*	0.011	*0.014*	0.045	*0.047*	0.064	*0.066*
β_4 =0.03	0.008	*0.011*	0.002	*0.010*	–0.005	*0.009*	0.006	*0.012*	0.005	*0.015*
β_5 =0.15	0.019	*0.021*	0.039	*0.041*	–0.013	*0.016*	0.026	*0.028*	0.012	*0.020*
β_6 =0.05	–0.004	*0.010*	0.012	*0.016*	–0.015	*0.017*	0.002	*0.012*	–0.011	*0.020*
β_1 =0.90	–0.374	*0.374*	–0.322	*0.323*	–1.512	*1.517*	–0.381	*0.382*	–0.393	*0.393*
β_2 =0.30	0.096	*0.097*	0.095	*0.096*	–0.039	*0.042*	0.117	*0.118*	0.106	*0.107*
β_3 =0.05	0.043	*0.044*	0.014	*0.018*	0.014	*0.017*	0.053	*0.054*	0.073	*0.075*
β_4 =0.03	0.010	*0.013*	0.003	*0.011*	–0.005	*0.010*	0.007	*0.014*	0.007	*0.017*
β_5 =0.15	0.023	*0.024*	0.044	*0.046*	–0.016	*0.019*	0.031	*0.034*	0.016	*0.024*
β_6 =0.05	–0.005	*0.011*	0.013	*0.019*	–0.018	*0.020*	0.002	*0.014*	–0.012	*0.022*

We begin our inference for the case when the value of the ratio $\sigma_\eta^2/\sigma_\nu^2$ is equal to zero (so a pure fixed-effect model). As we can see from Table 7, when the value of parameter β_1 fluctuates around 0.1–0.5, then the lowest bias was achieved in the case of the GMM1 estimator. But when the true values of parameter β_1 equals to 0.8 or 0.9 the lowest bias was achieved in the case of AH-L estimates. Then we increase the value of the ratio $\sigma_\eta^2/\sigma_\nu^2$ (Table 8) up to 1.0 and again allow the true values of parameter β_1 to vary between 0.1 and 0.9. First of all, from Table 8 we can see that when β_1 fluctuates around 0.1–0.5, then the lowest bias was achieved in the case of the GMM1 estimator, but when β_1 equals 0.8 or 0.9 the lowest bias was achieved in the case of the AH-L estimator. But when we compare the results in Table 8 with the same results in Table 7 we can see that the estimates obtained by the GMM1 and GMM1-SYS algorithms become less biased than the estimates obtained by FE and AH-L. Relating these to the AH-D estimates from Table 8, we can see that the bias was decreased compared with the same results in Table 7. But the starting values of the bias are so large that the AH-D estimator is not able to compete with the GMM1 and GMM1-SYS estimators' corresponding results. That is why we do not take into account the AH-D estimator results in our future explanations. Thus, based on these two tables, we see that when we increase the variance of the individual effects of the countries, the estimates obtained with the GMM1 and GMM1-SYS estimators are more accurate.

Table 8: Simulation results, T = 14, N = 101, $\sigma_\eta^2/\sigma_\nu^2 = 1.00$, $\mu = 0.9$

Parameters true values	Estimation algorithm									
	FE		AH-L		AH-D		GMM1		GMM1-SYS	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1 = 0.10	–0.023	0.024	–0.066	0.067	–0.134	0.137	–0.019	0.021	–0.031	0.032
β_2 =0.30	0.019	*0.020*	0.007	*0.008*	–0.001	*0.005*	0.021	*0.021*	0.023	*0.024*
β_3 =0.05	0.009	*0.010*	0.002	*0.004*	0.001	*0.004*	0.011	*0.012*	0.016	*0.017*
β_4 =0.03	0.001	*0.003*	0.000	*0.003*	–0.001	*0.003*	0.000	*0.003*	0.000	*0.004*
β_5 =0.15	0.006	*0.007*	0.004	*0.005*	0.000	*0.004*	0.008	*0.009*	0.009	*0.010*
β_6 =0.05	–0.001	*0.003*	0.001	*0.004*	–0.001	*0.004*	0.000	*0.004*	–0.002	*0.005*
β_1 =0.20	–0.039	*0.040*	–0.122	*0.122*	–0.279	*0.281*	–0.033	*0.034*	–0.052	*0.053*
β_2 =0.30	0.034	*0.034*	0.016	*0.016*	–0.004	*0.007*	0.037	*0.037*	0.041	*0.041*
β_3 =0.05	0.015	*0.016*	0.003	*0.005*	0.002	*0.004*	0.018	*0.018*	0.027	*0.028*
β_4 =0.03	0.002	*0.003*	0.000	*0.004*	–0.002	*0.004*	0.000	*0.004*	0.000	*0.006*
β_5 =0.15	0.010	*0.010*	0.007	*0.008*	–0.002	*0.005*	0.012	*0.013*	0.013	*0.014*
β_6 =0.05	–0.002	*0.004*	0.001	*0.004*	–0.003	*0.005*	–0.001	*0.005*	–0.004	*0.008*
β_1 =0.50	–0.125	*0.125*	–0.227	*0.228*	–0.758	*0.760*	–0.120	*0.120*	–0.137	*0.137*
β_2 =0.30	0.060	*0.060*	0.048	*0.049*	–0.015	*0.018*	0.069	*0.070*	0.069	*0.070*
β_3 =0.05	0.025	*0.026*	0.007	*0.010*	0.005	*0.008*	0.028	*0.029*	0.044	*0.046*
β_4 =0.03	0.004	*0.006*	0.001	*0.006*	–0.003	*0.006*	0.002	*0.007*	0.001	*0.010*
β_5 =0.15	0.014	*0.015*	0.022	*0.023*	–0.007	*0.009*	0.018	*0.020*	0.013	*0.017*
β_6 =0.05	–0.003	*0.007*	0.006	*0.009*	–0.007	*0.009*	0.001	*0.009*	–0.008	*0.015*
β_1 =0.80	–0.300	*0.301*	–0.294	*0.295*	–1.311	*1.315*	–0.302	*0.303*	–0.314	*0.315*
β_2 =0.30	0.086	*0.087*	0.084	*0.085*	–0.033	*0.035*	0.104	*0.104*	0.094	*0.095*
β_3 =0.05	0.038	*0.039*	0.012	*0.016*	0.011	*0.014*	0.044	*0.046*	0.064	*0.066*
β_4 =0.03	0.007	*0.010*	0.002	*0.009*	–0.005	*0.009*	0.005	*0.012*	0.004	*0.014*
β_5 =0.15	0.019	*0.021*	0.039	*0.040*	–0.014	*0.017*	0.027	*0.029*	0.014	*0.021*
β_6 =0.05	–0.004	*0.010*	0.012	*0.016*	–0.014	*0.017*	0.002	*0.012*	–0.010	*0.019*
β_1 =0.90	–0.375	*0.375*	–0.322	*0.323*	–1.503	*1.508*	–0.380	*0.380*	–0.391	*0.391*
β_2 =0.30	0.097	*0.097*	0.095	*0.096*	–0.039	*0.041*	0.117	*0.118*	0.105	*0.106*
β_3 =0.05	0.044	*0.045*	0.014	*0.019*	0.014	*0.017*	0.053	*0.054*	0.073	*0.075*
β_4 =0.03	0.009	*0.012*	0.003	*0.012*	–0.005	*0.010*	0.007	*0.014*	0.007	*0.016*
β_5 =0.15	0.023	*0.024*	0.044	*0.046*	–0.016	*0.019*	0.033	*0.035*	0.018	*0.025*
β_6 =0.05	–0.005	*0.011*	0.013	*0.018*	–0.017	*0.020*	0.002	*0.014*	–0.013	*0.022*

In Table 9 we continue to increase the variance of the individual effects of the countries up to 2.0. From Table 9 we see that again when the true values of parameter β_1 fluctuate around 0.1–0.5, then the lowest biases are achieved in the case of the GMM1 estimator, and when the true parameter of β_1 is equal to 0.8 or 0.9 then the lowest biases are achieved in the case of the AH-L estimator. Compared with Tables 7 and 8 we see that when we increase the variance of the individual effects of the countries, the results obtained with the GMM1 and GMM-SYS1 estimators are more accurate, because the bias becomes smaller than the same results in Tables 7 and 8.

Table 9: Simulation results, T = 14, N = 101, $\sigma_\eta^2/\sigma_\nu^2 = 2.00$**,** $\mu = 0.9$

Parameters true values	Estimation algorithm									
	FE		AH-L		AH-D		GMM1		GMM1-SYS	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1 = 0.10	–0.024	*0.024*	–0.067	*0.067*	–0.137	*0.139*	–0.019	*0.020*	–0.030	*0.031*
β_2 =0.30	0.020	*0.020*	0.007	*0.008*	–0.002	*0.005*	0.021	*0.021*	0.023	*0.023*
β_3 =0.05	0.010	*0.010*	0.002	*0.004*	0.001	*0.004*	0.011	*0.012*	0.016	*0.016*
β_4 =0.03	0.001	*0.003*	0.000	*0.003*	–0.001	*0.003*	0.000	*0.003*	0.000	*0.004*
β_5 =0.15	0.006	*0.007*	0.003	*0.005*	–0.001	*0.004*	0.008	*0.009*	0.009	*0.010*
β_6 =0.05	–0.001	*0.003*	0.000	*0.003*	–0.001	*0.004*	–0.001	*0.004*	–0.003	*0.005*
β_1 =0.20	–0.039	*0.040*	–0.122	*0.122*	–0.280	*0.282*	–0.032	*0.033*	–0.051	*0.051*
β_2 =0.30	0.033	*0.034*	0.016	*0.016*	–0.004	*0.007*	0.036	*0.037*	0.040	*0.040*
β_3 =0.05	0.015	*0.016*	0.003	*0.005*	0.002	*0.004*	0.018	*0.018*	0.027	*0.027*
β_4 =0.03	0.002	*0.003*	0.000	*0.003*	–0.001	*0.004*	0.000	*0.004*	0.000	*0.006*
β_5 =0.15	0.010	*0.010*	0.007	*0.008*	–0.002	*0.004*	0.013	*0.013*	0.014	*0.015*
β_6 =0.05	–0.002	*0.004*	0.002	*0.004*	–0.002	*0.005*	0.000	*0.005*	–0.004	*0.008*
β_1 =0.50	–0.126	*0.127*	–0.227	*0.227*	–0.754	*0.756*	–0.118	*0.119*	–0.135	*0.135*
β_2 =0.30	0.061	*0.061*	0.048	*0.049*	–0.015	*0.017*	0.069	*0.070*	0.068	*0.069*
β_3 =0.05	0.026	*0.026*	0.008	*0.010*	0.006	*0.008*	0.028	*0.029*	0.044	*0.045*
β_4 =0.03	0.004	*0.007*	0.001	*0.006*	–0.003	*0.006*	0.002	*0.008*	0.001	*0.010*
β_5 =0.15	0.013	*0.015*	0.022	*0.023*	–0.006	*0.009*	0.019	*0.021*	0.014	*0.018*
β_6 =0.05	–0.003	*0.007*	0.005	*0.009*	–0.008	*0.010*	0.001	*0.008*	–0.007	*0.014*
β_1 =0.80	–0.301	*0.302*	–0.295	*0.296*	–1.307	*1.310*	–0.301	*0.302*	–0.312	*0.313*
β_2 =0.30	0.087	*0.087*	0.084	*0.084*	–0.032	*0.035*	0.103	*0.104*	0.093	*0.094*
β_3 =0.05	0.038	*0.039*	0.013	*0.017*	0.012	*0.014*	0.045	*0.046*	0.063	*0.065*
β_4 =0.03	0.008	*0.011*	0.003	*0.010*	–0.005	*0.009*	0.006	*0.013*	0.005	*0.016*
β_5 =0.15	0.019	*0.021*	0.039	*0.040*	–0.013	*0.016*	0.029	*0.031*	0.016	*0.023*
β_6 =0.05	–0.004	*0.010*	0.012	*0.016*	–0.014	*0.017*	0.002	*0.013*	–0.010	*0.020*
β_1 =0.90	–0.374	*0.374*	–0.320	*0.322*	–1.511	*1.516*	–0.377	*0.377*	–0.387	*0.388*
β_2 =0.30	0.096	*0.097*	0.095	*0.095*	–0.039	*0.042*	0.116	*0.117*	0.103	*0.105*
β_3 =0.05	0.043	*0.044*	0.013	*0.018*	0.014	*0.017*	0.051	*0.053*	0.072	*0.073*
β_4 =0.03	0.010	*0.013*	0.003	*0.012*	–0.005	*0.010*	0.008	*0.014*	0.007	*0.016*
β_5 =0.15	0.023	*0.025*	0.045	*0.047*	–0.016	*0.020*	0.035	*0.038*	0.020	*0.027*
β_6 =0.05	–0.004	*0.011*	0.013	*0.018*	–0.018	*0.021*	0.003	*0.014*	–0.010	*0.021*

In Table 10 we continue to increase the variance of the individual effect of the countries up to 3.0. Then, as we can see from Table 10, all of the previous conclusions can be applied also to this table. Finally in Table 11 we increase the variance of the individual effect of the countries to 5.0. From Table 11 we can see that all the above conclusions are still the same. The one difference is that when the true values of parameter β_1 fluctuate around 0.8–0.9, then the bias obtained with GMM1 becomes smaller than the same bias obtained with FE. This means that with continued increasing of the variance of the individual effect of the countries, the behavior of the GMM1 estimator becomes better and better.

Table 10: Simulation results, T = 14, N = 101, $\sigma_\eta^2/\sigma_\nu^2 = 3.00$**,** $\mu = 0.9$

Parameters true values	Estimation algorithm									
	FE		AH-L		AH-D		GMM1		GMM1-SYS	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1 = 0.10	–0.024	*0.024*	–0.066	*0.067*	–0.135	*0.138*	–0.017	*0.019*	–0.029	*0.030*
β_2 =0.30	0.019	*0.020*	0.007	*0.008*	–0.002	*0.005*	0.021	*0.021*	0.022	*0.022*
β_3 =0.05	0.009	*0.010*	0.001	*0.004*	0.001	*0.004*	0.011	*0.011*	0.015	*0.016*
β_4 =0.03	0.001	*0.003*	0.000	*0.003*	–0.001	*0.003*	0.000	*0.003*	0.000	*0.004*
β_5 =0.15	0.006	*0.007*	0.003	*0.005*	–0.001	*0.004*	0.008	*0.009*	0.009	*0.010*
β_6 =0.05	–0.001	*0.003*	0.001	*0.004*	–0.001	*0.004*	0.000	*0.004*	–0.002	*0.005*
β_1 =0.20	–0.040	*0.040*	–0.122	*0.122*	–0.279	*0.281*	–0.031	*0.032*	–0.050	*0.050*
β_2 =0.30	0.034	*0.034*	0.016	*0.016*	–0.004	*0.007*	0.036	*0.037*	0.039	*0.040*
β_3 =0.05	0.015	*0.016*	0.003	*0.005*	0.002	*0.004*	0.017	*0.018*	0.026	*0.027*
β_4 =0.03	0.002	*0.004*	0.000	*0.004*	–0.001	*0.004*	0.000	*0.004*	–0.001	*0.006*
β_5 =0.15	0.010	*0.010*	0.007	*0.008*	–0.002	*0.005*	0.013	*0.014*	0.014	*0.016*
β_6 =0.05	–0.001	*0.004*	0.002	*0.004*	–0.002	*0.005*	0.000	*0.005*	–0.004	*0.007*
β_1 =0.50	–0.126	*0.127*	–0.227	*0.227*	–0.760	*0.762*	–0.116	*0.117*	–0.132	*0.133*
β_2 =0.30	0.060	*0.061*	0.048	*0.048*	–0.016	*0.018*	0.069	*0.069*	0.067	*0.067*
β_3 =0.05	0.026	*0.026*	0.007	*0.010*	0.006	*0.008*	0.028	*0.029*	0.043	*0.044*
β_4 =0.03	0.004	*0.006*	0.001	*0.006*	–0.003	*0.006*	0.001	*0.007*	0.001	*0.010*
β_5 =0.15	0.013	*0.015*	0.022	*0.023*	–0.007	*0.009*	0.020	*0.022*	0.015	*0.019*
β_6 =0.05	–0.003	*0.007*	0.006	*0.009*	–0.007	*0.009*	0.002	*0.009*	–0.007	*0.014*
β_1 =0.80	–0.302	*0.302*	–0.295	*0.296*	–1.313	*1.316*	–0.300	*0.300*	–0.311	*0.311*
β_2 =0.30	0.087	*0.087*	0.084	*0.085*	–0.033	*0.035*	0.103	*0.104*	0.092	*0.093*
β_3 =0.05	0.038	*0.039*	0.012	*0.016*	0.012	*0.014*	0.044	*0.045*	0.062	*0.064*
β_4 =0.03	0.008	*0.010*	0.002	*0.010*	–0.005	*0.009*	0.006	*0.012*	0.005	*0.015*
β_5 =0.15	0.020	*0.021*	0.039	*0.040*	–0.014	*0.017*	0.030	*0.032*	0.018	*0.024*
β_6 =0.05	–0.004	*0.010*	0.011	*0.015*	–0.015	*0.017*	0.003	*0.013*	–0.009	*0.019*
β_1 =0.90	–0.376	*0.376*	–0.321	*0.322*	–1.508	*1.513*	–0.377	*0.377*	–0.387	*0.387*
β_2 =0.30	0.097	*0.098*	0.096	*0.096*	–0.039	*0.042*	0.117	*0.118*	0.104	*0.105*
β_3 =0.05	0.043	*0.044*	0.014	*0.019*	0.014	*0.017*	0.051	*0.053*	0.071	*0.073*
β_4 =0.03	0.010	*0.012*	0.003	*0.011*	–0.005	*0.010*	0.007	*0.014*	0.007	*0.017*
β_5 =0.15	0.023	*0.025*	0.044	*0.046*	–0.016	*0.019*	0.035	*0.038*	0.021	*0.027*
β_6 =0.05	–0.005	*0.011*	0.013	*0.018*	–0.017	*0.020*	0.003	*0.014*	–0.011	*0.021*

Therefore, based on the inferences from Tables 7–11 we conclude that GMM-type estimators (Arellano-Bond and Blundell-Bond) perform well compared with the fixed effects and instrumental variables estimators, particularly when we increase the coefficient on the lagged dependent variables and the variance of the individual effects of the countries. Thus, based on the simulation results, we conclude that the GMM-type estimators are able to give more reasonable and accurate estimation results than fixed-effects or instrumental variable estimators.

Table 11: Simulation results, T = 14, N = 101, $\sigma_\eta^2/\sigma_\nu^2 = 5.00$, $\mu = 0.9$

Parameters true values	Estimation algorithm									
	FE		AH-L		AH-D		GMM1		GMM1-SYS	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_1 = 0.10	–0.024	*0.025*	–0.067	*0.068*	–0.137	*0.140*	–0.017	*0.018*	–0.028	*0.029*
β_2 =0.30	0.020	*0.020*	0.007	*0.008*	–0.002	*0.005*	0.021	*0.021*	0.021	*0.021*
β_3 =0.05	0.009	*0.010*	0.001	*0.004*	0.001	*0.003*	0.011	*0.011*	0.014	*0.015*
β_4 =0.03	0.001	*0.003*	0.000	*0.003*	–0.001	*0.003*	0.000	*0.003*	–0.001	*0.004*
β_5 =0.15	0.006	*0.007*	0.003	*0.005*	–0.001	*0.004*	0.009	*0.009*	0.009	*0.011*
β_6 =0.05	–0.001	*0.003*	0.001	*0.004*	–0.001	*0.004*	0.000	*0.004*	–0.002	*0.005*
β_1 =0.20	–0.041	*0.041*	–0.122	*0.123*	–0.280	*0.282*	–0.030	*0.031*	–0.048	*0.049*
β_2 =0.30	0.034	*0.034*	0.016	*0.016*	–0.004	*0.007*	0.036	*0.037*	0.038	*0.038*
β_3 =0.05	0.016	*0.016*	0.003	*0.005*	0.002	*0.004*	0.017	*0.018*	0.025	*0.026*
β_4 =0.03	0.002	*0.003*	0.000	*0.003*	–0.001	*0.004*	0.000	*0.004*	–0.001	*0.006*
β_5 =0.15	0.010	*0.010*	0.007	*0.008*	–0.002	*0.005*	0.014	*0.015*	0.015	*0.016*
β_6 =0.05	–0.002	*0.004*	0.001	*0.004*	–0.003	*0.005*	0.000	*0.005*	–0.004	*0.007*
β_1 =0.50	–0.128	*0.129*	–0.227	*0.228*	–0.758	*0.760*	–0.115	*0.116*	–0.130	*0.131*
β_2 =0.30	0.061	*0.061*	0.049	*0.049*	–0.015	*0.017*	0.069	*0.070*	0.066	*0.066*
β_3 =0.05	0.026	*0.027*	0.008	*0.010*	0.006	*0.008*	0.027	*0.028*	0.042	*0.043*
β_4 =0.03	0.004	*0.007*	0.001	*0.006*	–0.003	*0.006*	0.001	*0.007*	0.000	*0.010*
β_5 =0.15	0.014	*0.015*	0.023	*0.024*	–0.007	*0.009*	0.022	*0.024*	0.018	*0.022*
β_6 =0.05	–0.003	*0.007*	0.006	*0.009*	–0.007	*0.009*	0.002	*0.009*	–0.005	*0.013*
β_1 =0.80	–0.303	*0.303*	–0.294	*0.295*	–1.306	*1.310*	–0.298	*0.299*	–0.308	*0.308*
β_2 =0.30	0.087	*0.088*	0.084	*0.085*	–0.032	*0.035*	0.103	*0.104*	0.091	*0.092*
β_3 =0.05	0.039	*0.039*	0.012	*0.016*	0.012	*0.014*	0.043	*0.045*	0.062	*0.064*
β_4 =0.03	0.008	*0.010*	0.002	*0.010*	–0.005	*0.009*	0.005	*0.012*	0.005	*0.014*
β_5 =0.15	0.020	*0.021*	0.039	*0.040*	–0.013	*0.016*	0.032	*0.034*	0.020	*0.025*
β_6 =0.05	–0.004	*0.010*	0.012	*0.016*	–0.015	*0.017*	0.004	*0.013*	–0.009	*0.018*
β_1 =0.90	–0.375	*0.376*	–0.320	*0.321*	–1.512	*1.517*	–0.374	*0.374*	–0.383	*0.383*
β_2 =0.30	0.097	*0.097*	0.095	*0.095*	–0.041	*0.043*	0.116	*0.117*	0.103	*0.104*
β_3 =0.05	0.044	*0.045*	0.013	*0.018*	0.015	*0.018*	0.050	*0.052*	0.070	*0.072*
β_4 =0.03	0.010	*0.012*	0.003	*0.012*	–0.005	*0.010*	0.007	*0.014*	0.007	*0.017*
β_5 =0.15	0.023	*0.025*	0.044	*0.046*	–0.016	*0.019*	0.038	*0.040*	0.023	*0.030*
β_6 =0.05	–0.004	*0.011*	0.013	*0.018*	–0.017	*0.020*	0.004	*0.014*	–0.009	*0.020*

7 Synthesis of results and conclusions

We analyze the determinants of export sophistication in a dynamic panel data estimation set-up and discriminate among several estimation procedures. A comparison of the different estimation algorithms, namely fixed effects, Anderson-Hsiao (level and difference instruments), Arellano-Bond, and Blundell-Bond estimators show that the GMM-type estimators (Arellano-Bond and Blundell-Bond) outperform instrumental-variable and fixed-effects estimators. Based on Monte Carlo simulations we conclude that the GMM-type estimators behave much better when the true coefficient of the lagged dependent variable is close to one. This is because, as was shown by Monte Carlo experiments, when we increase the value of the dependent lagged parameter and the variance of the individual effects, the bias of the GMM-type estimators decreases.

Thus, based on the Arellano-Bond GMM estimation results, we are able to note that these estimates are the closest estimates to the true values. Hence, from this point of view the most accurate estimation results presented in Tables 3–5 are those obtained via the Arellano-Bond estimation. Now we can conclude that GDP per capita and population are statistically significant and have the expected sign. These two variables are strong determinants of export sophistication. From this point of view, our finding is consistent with the findings of Hausmann et al. (2007). But the difference is that for the dynamic panel data model the impact of GDP per capita and population on the export sophistication is much higher.

Further, human capital is insignificant and its value almost equals zero, which means that human capital is not a strong determinant for export sophistication and we conclude that a causal effect may go from export sophistication to human capital and not vice versa. The impact of institutional quality on the export sophistication is also close to zero and its value is not significant. But when we conduct estimation for two different groups of countries we see that the estimate for the rule of law index has a positive but still insignificant effect for low institutional quality countries and it has negative and significant effect for high institutional quality countries. This means that the rule of law index can be considered as a determinant for export sophistication in high institutional quality countries, albeit its impact is negative. Finally, the territorial size of the countries has a negative impact on export sophistication, which means that small countries could more easily diversify production and export than larger countries.

Finally, we added to the export sophistication model a lagged value of this dependent variable. The lagged value of export sophistication has a significant and positive effect on its current value, and the estimated coefficient suggests a strong persistence of the behavior of export sophistication. This is a sign of the path dependency of the export activities in terms of export sophistication. Further, strong persistence is also likely the reason for the absence of structural breaks in export sophistication for a large sample of countries and explains why export sophistication is not affected by the crisis.

Our results show that when we apply the panel data over a different period and different country set the findings of Hausmann et al. (2007) remain robust. Our findings using panel data from 2001–2014 confirm that export path-dependency, GDP per capita, and the size of the economy exert significant and positive effects on export sophistication. On the other hand, we find that the impact of institutional quality has a positive effect only for low institutional quality countries, but has a negative and significant effect for countries with high institutional quality. The high persistence of export sophistication forms a basis for the policy to strongly support export diversification that promotes not only productivity and sustainable economic growth but also resistance during economic downturns.

Technical Appendix

In this Technical Appendix we provide a brief overview of the basic estimation algorithms employed in the paper.

3.1 The instrumental variables approach (Anderson-Hsio estimator)

The IV-type estimator in a dynamic panel data model was first proposed by Anderson and Hsiao (1982). In order to summarize this algorithm let's consider the following model.

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \eta_i + \varepsilon_{it}\ ,\ i=1,...,N,\ \ t=1,...,T\ ,\ \varepsilon_{it} \sim N(0,\sigma_\varepsilon^2) \text{ and } |\rho|<1$$

x_{it} is a vector of the explanatory variable, η_i denotes the unobserved time-invariant individual effect, and ε_{it} is the error term with $E(\varepsilon_{it})=0$ and $E(\varepsilon_{it}\varepsilon_{js})=\sigma_\varepsilon^2$ if $i=j$ and $t=s$, $E(\varepsilon_{it}\varepsilon_{js})=0$ otherwise. We assume that:

$$E(\eta_i)=0\,, E(\eta_i x_{it})\neq 0\,, E(\varepsilon_{it}x_{it})=0\,.$$

The most popular way to obtain consistent estimates for ρ is to eliminate the individual effects. Hence, instead of the within-estimator we apply first differencing and obtain:

$$\begin{aligned}(y_{it}-y_{i,t-1}) &= \rho(y_{i,t-1}-y_{i,t-2})+\beta(x_{it}-x_{i,t-1})+(\varepsilon_{it}-\varepsilon_{i,t-1})\\ \Delta y_{it} &= \rho\Delta y_{i,t-1}+\beta\Delta x_{it}+\Delta\varepsilon_{it}\end{aligned}.$$

In the model above, $y_{i,t-1}$ is correlated by construction with $\varepsilon_{i,t-1}$ and $E[y_{i,t-1}(\varepsilon_{i,t} - \varepsilon_{i,t-1})] = E[\varepsilon_{i,t-1}(\varepsilon_{i,t} - \varepsilon_{i,t-1})] = -E(\varepsilon_{i,t-1}^2) = -\sigma_\varepsilon^2$. Therefore, we need instruments that are uncorrelated with $(\varepsilon_{it} - \varepsilon_{i,t-1})$, but correlated with $(y_{it-1} - y_{i,t-2})$. Anderson and Hsiao suggest to use level instruments $y_{i,t-2}$ or the lagged difference $(y_{it-2} - y_{i,t-3})$, because $E[y_{i,t-2}(\varepsilon_{it} - \varepsilon_{i,t-1})] = E[\varepsilon_{i,t-2}(\varepsilon_{i,t} - \varepsilon_{i,t-1})] = 0$ and $E[(y_{i,t-2} - y_{i,t-3})(\varepsilon_{it} - \varepsilon_{i,t-1})] = E[(\varepsilon_{i,t-2} - \varepsilon_{i,t-3})(\varepsilon_{it} - \varepsilon_{i,t-1})] = 0$. Using these instruments the above model can be estimated in the traditional instrumental variable way (Behr, 2003).

3.2 The GMM approach (Arellano-Bond estimator)

Let us consider the same dynamic panel data model as in section 3.1. The GMM estimation algorithm is also based on the first differencing of the above model. As noted by Arellano and Bond (1991), all $y_{i,t-2-j}$, $j=0,1,...,$ satisfy the conditions

$$\begin{aligned}E(y_{i,t-2-j}(y_{i,t-1}-y_{i,t-2})) &\neq 0\\ E(y_{i,t-2-j}(\varepsilon_{i,t}-\varepsilon_{i,t-1})) &= 0\end{aligned}.$$

Therefore, they all are legitimate instruments for $(y_{i,t-1} - y_{i,t-2})$. For example in period t = 3, variable y_{i1} is a valid instrument, since it is correlated with $(y_{i,2} - y_{i,1})$ and not correlated with $(\varepsilon_{i,3} - \varepsilon_{i,2})$, as long as ε_{it} are not serially correlated. When t = 4, the variable y_{i1} and y_{i2} are valid instruments, since they are correlated with $(y_{i,3} - y_{i,2})$ and not correlated with $(\varepsilon_{i,4} - \varepsilon_{i,3})$. In result we will have the following matrix of all possible instruments:

$$W_i = \begin{bmatrix} y_{i1} & 0 & \dots & 0 & \Delta x_{i3} \\ 0 & y_{i1}, y_{i2} & \dots & 0 & \Delta x_{i4} \\ 0 & 0 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & y_{i1}, y_{i2}, \dots, y_{i,T-2} & \Delta x_{iT} \end{bmatrix}.$$

Thus we have $E[W_i \Delta\varepsilon_{it}] = 0$. If ε_{it} is homoscedastic then as an initial weighting we can use the following $(T-2)(T-2)$ size matrix:

$$H = \begin{bmatrix} 2 & -1 & 0 & \dots & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \vdots & \vdots & 0 & -1 & 2 \end{bmatrix}.$$

Using the set of all possible instruments and the initial weighting matrix the model can be estimated in the traditional GMM way (Behr, 2003).

3.3 The system-GMM approach (Blundell and Bond estimator)

The above GMM estimator suggested by Arellano-Bond is known to be rather inefficient when instruments are weak. Blundell and Bond (1998) suggest using both level and differencing instruments simultaneously. We will not present all the details about the Blundell-Bond algorithm, because all the details related to the instruments matrix and initial weighting matrix are provided in Behr (2003). Thus, having a matrix of all available instruments (level and differenced instruments) and an initial weighting matrix, the remaining part of the estimation can proceed in the standard GMM manner.

References

Anand, R., S. Mishra, and N. Spatafora. (2012). Structural Transformation and the Sophistication of Production. IMF Working Paper No. 59.

Anderson, T., and C. Hsiao (1981). Estimation of dynamic models with error components. Journal of American Statistical Association, Vol. 76, No. 375, pp. 598–606.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies, Vol. 58, No. 2, pp. 277–297.

Baltagi, B. (2013). Econometric analysis of panel data, Fifth Edition, John Wiley & Sons, Chichester.

Behr, A. (2003). A comparison of dynamic panel data estimators: Monte Carlo evidence and an application to the investment function. Discussion paper 05/03, Economic Research Centre of the Deutsche Bundesbank.

Blundell, R. and S. Bond (1998). Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics, vol. 87, No. 1, pp. 115–143.

Bun, M., and Carree, M. (2005). Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 3(2), 200–210 (11).

Cabral, M., and P. Veiga (2010). Determinants of Export Diversification and Sophistication in Sub-Saharan Africa. FEUNL Working Paper Series N. 550, Lisbon: Universidade Nova de Lisboa.

Córcoles, D., Díaz-Mora, C., Gandoy, R. (2014). Product sophistication: A tie that binds partners in international trade. Economic Modelling, vol. 44(S1), pp. S33–S41.

Da Costa Neto and R. Romeu (2011). Did Export Diversification Soften the Impact of the Global Financial Crisis? IMF Working Paper, WP/11/99.

Gaduh, Arya B. (2002). Properties of Fixed Effects Dynamic Panel Data Estimators for a Typical Growth Dataset. CSIS Economics Working Paper Series 062. Jakarta, Indonesia: Centre for Strategic and International Studies.

Hansen, G. (2001), A bias-corrected least squares estimator of dynamic panel models, Allgemeines Statistisches Archiv, Vol. 85, pp. 127–140.

Hausmann, R. and B. Klinger (2006). Structural transformation and patterns of comparative advantage in the product space. CID Working paper No. 128 August (Cambridge, Mass.: Center for International Development).

Hausmann, Ricardo, J. Hwang, and D. Rodric (2007). What you export matters. Journal of Economic Growth, Springer, Vol. 12 (1), pp. 1–25.

Hausmann, R., C. Hidalgo, S. Bustos, M. Coscia, S. Chung, J. Jimenes, A. Simoes, M.A. Yildirim (2011). The atlas of economic complexity: Mapping paths to prosperity. Center for International Development. Cambtodge MA Harvard University.

Hidalgo, C.A., Klinger B., Barabási A.L., & Hausmann R. (2007) The product space conditions the development of nations. Science. Vol. 317, pp. 482–487.

Hummels, D. and P.J. Klenow (2005). The variate and quality of a Nation’s export. American Economic Review, Vol. 95, No. 3, pp. 704–723.

Kiviet, J., (2012). Monte Carlo simulation for econometricians. Foundations and Trends® in Econometrics, now publisher, Vol. 5(1–2), pp. 1–181.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. American Economic Review 70, pp. 950–959.

Lall, Sanjaya, John Weiss, and Jinkang Zhang, (2005). The Sophistication of Exports: A New Measure of Product Characteristics. Queen Elisabeth House Working Paper Series No. 123, Oxford University.

Liang, N., Parkhe, A. (1997). Importer Behavior: The Neglected Counterpart of International Exchange. Journal of International Business Studies, 28 (3), 495–530.

Mauro, P. (1995). Corruption and Growth. The Quarterly Journal of Economics, Vol. 110 (3): pp.~681–712.

Mehta, A & J. Felipe (2014). Education and the journey to the core: Path-dependence of leapfrogging?. ADB Economics working paper series 395, Asian Development Bank.

Michaely, Michael, (1984). Trade, Income Levels, and Dependence, North-Holland, Amsterdam.

Nickel, S.J. (1981). Biases in Dynamic models with fixed effects. Econometrica, Vol. 49, pp. 1417–1426.

Santos, L. and E. Barrios (2011). Small Sample Estimation in Dynamic Panel Data Models: A Simulation Study. Open Journal of Statistics. 1 (2) pp. 58–73.

Scott, P. (2008). The relative sophistication of Chines exports. Economic Policy, 1, pp. 5–49.

Shelburne, R. (2010). The global financial crisis and its impact on trade: The world and the European emerging economies. Discussion Paper Series Number 2010.2, United Nation Economic Commission for Europe, Switzerland.

Weldemicael, E. (2014). Technology, Trade Costs and Export Sophistication. The World Economy, 37(1), 14–41.

Zhu S, Fu X., Lai M., Ji Xuan (2010). What drives the export sophistication of countries? J World Econ 4: pp. 28–43.