KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE

OF

ECONOMIC RESEARCH

KYOTO UNIVERSITY

KYOTO, JAPAN

Export sophistication: A dynamic panel data approach

Evžen Kočenda[a] and Karen Poghosyan[b]

Abstract

In this paper we analyze export sophistication based on a large panel dataset (2001–2015; 101 countries) and using various 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. Based on our analysis we document that GDP per capita and the size of the economy exhibit significant and positive effects on export sophistication; weak institutional quality exhibits negative effect. We also show that export sophistication is path-dependent and stable even during a major economic crisis, which is especially important for emerging and developing economies.

Keywords: international trade; export sophistication; emerging and developing economies; specialization; dynamic panel data; Monte Carlo simulation; panel data estimators

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

a Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Czech Republic. Address: IES, Opletalova 26, 110 00 Prague. Institute of Theory of Information and Automation, Czech Academy of Sciences, Prague; CES, Munich; IOS, Regensburg. E-mail: evzen.kocenda@fsv.cuni.cz

b Central Bank of Armenia, Economic Research Department, Yerevan, Armenia. E-mail: karen.poghosyan@cba.am

1. Introduction

We analyze links between several country characteristics and export sophistication, a concept that aims to encapsulate the productivity level coupled with a country's production; export sophistication is then empirically reflected in exports data. Our set of country characteristics is identical to that employed by Hausman et al. (2007) who show that higher export sophistication is a robust predictor of higher economic growth. They also 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 factors are added, the results are mixed, and we believe that the estimation procedure matters as well. In this paper we take the inspiration from Hausmann et al. (2007) and extend the topic in two ways. First, we modify the specification of export sophistication to account for path dependency in the structure of exports; in doing so we cover 101 (developed, emerging and developing) economies over 2001–2015. Second, we employ several dynamic panel data estimators and perform Monte Carlo simulations to obtain the most accurate and effective estimates.

Our analysis is motivated by the following relevant issues. First, Analyzing export sophistication is particularly important from the perspective of emerging markets and developing economies. Based on the level of export sophistication, Felipe et al. (2014) empirically show that excessive number of countries worldwide is in a "bad product" trap as they export mostly unsophisticated and unconnected products. Policy interventions, targeting the market failures that are widespread in many developing countries, are required to escape the trap. Along with them, the accent is on countries' production structures in terms of specialization and sophistication. Fortunato and Razo (2014) argue that successful developing countries progressively change their production structure, replacing low value added activities and unsophisticated goods with higher value added activities and more sophisticated products; this way they are able to climb a sophistication ladder. Hausmann et al. (2007) have shown that it is not the specialization alone, but the sophistication of goods exports that matters for growth. In accord with the above, the degree of export sophistication differs chiefly due to the effect of emerging markets that import primary goods from developing countries but export manufacturing goods to developed economies (Hanson, 2012).

Second, our paper deals with aggregate exports and we argue that they are likely to exhibit path-dependency, i.e., aggregate exports in various products tend to last once they have started. From a broader perspective, the literature on path dependence related to economic activity argues that long-term aggregate output behavior is affected by the sample path realization of an economy (Durlauf, 1994). More specific case for path dependence can be based on production. Reinstaller and Reschenhofer (2015) bring evidence that path-dependencies in systems of production have a dual role as they are not only a source of structural lock-in, but also a potential starting point for new developments. They show that factors causing path dependence in systems of production (technological relatedness and local capabilities) are also an important source of competitiveness for traded commodities. More productive firms typically select themselves into exports (Melitz, 2003) and that opens a way from production to path-dependence in aggregate exports. Such export path-dependence is evidenced by Egger and Pfaffermayr (2011) for 120 major countries over 1995-2004 period. Path dependence in international trade has also been recognized to depend on foregoing investments (Teece at al., 2007) as investments enlarge the production base (installed capital), increase production capacity, and affect subsequent export performance, often in a form of positive spillovers (Görg and Strobl, 2001; Greenaway et al., 2004). Finally, Nunn (2009) surveys a growing body of empirical evidence that shows how historic events impact economic development, including the evolution of international trade, via channels enabled by institutions, culture, knowledge, and technology. Hence, to account for the path-dependence we modify the specification of the export sophistication model by including the lagged value of the dependent variable.

Third, from an econometric perspective, inclusion of the lagged value of the dependent variable helps 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 rather than strictly exogenous ones. Further, and more importantly, 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 by Anderson and Hsiao (1982) and the GMM-type estimators proposed by Arellano and Bond (1991) and Blundell and Bond (1998).[1] 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 Arellano-Bond two-step GMM estimator (GMM2) performs well in comparison to fixed-effects, IV-type and two step system GMM estimators; it 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.

Thus in our analysis we use a new panel dataset and relatively more accurate estimators to effectively evaluate several important determinants of export sophistication and their parameters. Based on our analysis we show that export sophistication is path-dependent and we confirm that GDP per capita and the size of the economy exercise significant and positive effects on export sophistication; institutional quality exhibits negative effect. Our results could be of interest to practical macroeconomic policymakers, especially from emerging and developing countries. This is because for those countries there is a large scope for the 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 paper is organized as follows. In section 2 we briefly review the literature related to the researched topic. In section 3, we introduce two export sophistication measures, and our specification for the export sophistication regression that accounts for path dependence. In section 4 we present the data and their descriptive statistics. We present results from our estimations and robustness checks in section 5. Brief conclusions are summarized in section 6. Design and results of the Monte Carlo simulations are presented in the Technical Appendix.

---

[1] 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.

2. Literature review

The link between the nature of exports and the performance of open economies is an important empirical question that gains even more currency for emerging and developing economies. Lin and Sim (2013) analyze exports of the Least Developed Countries (LDCs) and persuasively show that quantitative significance of the trade expansion leading to the GDP per capita increases emphasizes the importance of trade with respect to the economic development of low income countries. Further, Hu et al. (2016) and Lin (2015) show that expansion in exports raises firm total factor productivity in China along with a learning effect due to more refined exports. 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 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 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. (2009) 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. In this respect, Lin et al. (2017) analyze whether export sophistication contributes to the income improvement in sub-Saharan Africa and show that within-country variations in export sophistication lead to income growth in the region over the long run. Further,

Cabral and Veiga (2010) 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.

Another contribution is Weldemicael (2012), 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, 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 this paper we modify the econometric specification of export sophistication model to account for path dependency in the structure of exports in a large set of 101 countries. Second, we employ several estimators and perform Monte Carlo simulations to obtain the most accurate estimates of the determinants of export sophistication.

3. Export sophistication, model specification, and estimation algorithms

In this paper we analyze the set of determinants introduced by Hausmann et al. (2007) and explore whether the determinants have the same effect (in terms of sign and significance) on export sophistication after modifying the model specification and covering a recent time span. To do this, we aim to obtain the most accurate estimates by performing estimations via suitable estimation procedures. As a second step using the actual panel dataset, we conduct Monte Carlo simulations to investigate the performance of the different panel data estimation algorithms in terms of minimizing bias and root mean square error (RMSE).

3.1 Measures of export sophistication

The measure of export sophistication (denoted as EXPY) is defined as the average income associated with a country's export bundle. We follow Hausmann 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:

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

where Y_i denotes the GDP per capita of country i.

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

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

As a robustness check we further employ an alternative diversification measure. Similarly as Lin et al. (2017), we employ a modification of the Export Similarity Index (ESI) from Schott (2008) and Finger and Kreinin (1979). The ESI is grounded in the idea, that one country (or a group of countries) can be identified as exporting goods of high productivity when compared to other countries. Hence, this particular country (or a group of countries) exhibits the export sophistication benchmark. For our purpose we use the USA as a benchmark. The ESI between country A and the USA for product p is defined as:

$$ESI(A - USA, p) = \sum_p \min \left[S_{A_p}, S_{USA_p} \right],$$

where S_{USA_p} is the share of product p in the total exports from the USA. Higher value of the ESI indicates that exports of country A exhibit more similarity with the U.S. exports. Thus, higher

value of the ESI indicates higher degree of the country A export sophistication. The ESI is bounded by zero and unity.

We calculate the export sophistication measures based on the export data that are described in detail in section 4.

3.2 Econometric model

An empirical observation based on Hausmann 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 their 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 their 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 assumption is in line with our arguments (presented in the Introduction) on the propagation of the path dependence in international trade that (i) arises via production (Reinstaller and Reschenhofer, 2015), investments (Teece at al., 2007) or via channels of institutions, culture, knowledge, and technology (Nunn, 2009), and (ii) its existence is documented in Egger and Pfaffermayr (2011).

We account for path dependency by including the lagged value of the dependent variable (export sophistication). Hence, the specification of export sophistication model can be formalized in the following way:

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

where $\ln(EXPY)$ is the logarithm of the export sophistication index that is formally defined in full detail in section 3.1, $\ln(GDPpc)$ is the logarithm of GDP per capita, HC is a measure of human capital, $RofL$ is the rule-of-law index, $\ln(POP)$ is the logarithm of population, and $\ln(AREA)$ 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.

4. Data and descriptive statistics

We construct a balanced panel dataset consisting of six macroeconomic variables on a yearly frequency from 2001 to 2015; in total we have 1515 observations per each macroeconomic variable (101 countries times 15 years). Our panel includes both the pre- and post-crisis periods.[2] The data include a measure of export sophistication (the dependent variable defined in section 3) and five macroeconomic variables: GDP per capita, human capital, a rule-of-law index, population, and land area. All the variables, with the exception of human capital and the rule-of-law index, are in logarithms. All variables correspond to those used in Hausmann et al. (2007).

We first introduce the macroeconomic determinants. (i) 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. (ii) Human capital (*HC*) 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.[3] (iii) Data on population (*POP*), as a proxy for the size of the economy/market, are reported in millions of inhabitants. (iv) data on the area (*AREA*) are reported in square kilometers. All four independent variables above were obtained from the World Development Indicators (WDI) database of the World Bank. (v) 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 2000–2015. The rule-of-law index is commonly used to capture the degree of institutional quality. The index ranges from -2.5 to 2.5 and a higher value represents better governance and a higher quality of institutions.

---

[2] We formally check for the presence of a structural break and provide more details on its absence in section 5.
[3] For some countries and some years the yearly gross tertiary enrolment ratio data are missing. We fill the missing non-critical number of gaps via interpolation. We do not proxy human capital by average years of schooling from Barro and Lee (2000)'s international education attainment dataset because those data are only available for every five years, while our panel dataset is in a yearly frequency.

The measure of export sophistication (our dependent variable denoted as *EXPY*) is calculated based on the export data from the International Trade Center database (ITC). 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 each year during 2001–2015.[4] Hence, our analysis deals with aggregate exports and not with bilateral trade.

Our data coverage is quite representative as our balanced panel data set covers 101 countries that in terms of GDP represent about 92 % of the global economy and in terms of exports about 86 % of the global exports.[5] According to the International Monetary Fund classification (2017), our 101 countries contain 20 emerging markets, 49 developing countries, and 32 developed economies. In Table 1 we report the 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). The EXPY value increased steadily from 13,756.4 USD in 2001 to 23,827.3 USD in 2015, 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 22,874.3 USD in 2001 to 39,231.0 USD in 2015 (Table 1, column 7).

Using data from columns 5 and 6 of Table 1, we calculate that the minimum value of EXPY increased on average by 7.6%, while the maximum value of EXPY increased on average by 4.4% (both calculated by geometric mean). This is an indication that the EXPY value for some low-income countries during 2001–2015 grows more rapidly than in high- and middle-income countries. This observation conveys optimism 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.

---

[4] The balanced nature of our panel is just a matter of convenience. The estimators work fine with unbalanced panels and our previous estimations were also performed on unbalanced panels. The results were not materially different.
[5] The proportions are calculated based on 2001-2015 data span and represent average values for the total of 15 years. Source: https://data.worldbank.org/.

5. Estimation results

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 2001–2015. The results (not reported but available upon request) suggest that there is no structural break in export sophistication dynamics. Hence, we estimate specification (1) without adjustments for a structural break.

5.1 Results of the two-stage instrumental variable estimation

In our analysis we assume that variables GDP per capita (*GDPpc*) and rule-of law index (*RofL*) are endogenous while population, human capital, and land area are strictly exogenous ones. Variables *GDPpc* and *RofL* are assumed to be endogenous, because specification (1) might involve presence of reverse causality between export sophistication and both endogenous variables; these might be then correlated with the error term. In the first step we employ the fixed-effects instrumental variables estimation algorithm where we employ two instrumental variables (IV).

First, following Bruckner (2012) we use a country-specific international commodity export price index (ICEPI) for agricultural and natural resources commodities (*ComPI*) as a proxy for the GDP per capita.[6] The country-specific international commodity export prices are exogenous for most commodities and countries. Specifically, it has to be noted that developing and emerging economies, that comprise majority in our dataset, are price takers on the international commodity markets (Bruckner, 2012). The data to construct the ICEPI over 2001-2015 are from the IndexMundi database (international commodity prices), the International Trade Centre (values of commodity exports) and the World Bank's World Development Indicators (GDP data).

Second, following Easterly (2007) and Bennett and Nikolaev (2016) we employ a measure of a factor endowment as an IV to proxy for the institutional quality, i.e. rule-of law index. Factor endowment in a form of land availability is considered to be an initial point for the formation of property rights and corresponding institutions that serve to define and defend

---

[6] Similar to Bruckner (2012), we calculate the agricultural commodity export price index with the following commodities: beef, coffee, cocoa, cotton, maize, rice, rubber, sugar, tea, wheat and wood. Further, the commodities included in the natural resource export price index are aluminum, copper, gold, iron and crude oil. The ICEPI then combines both commodity categories; see Bruckner (2012) for details.

property rights (Bennett and Nikolaev, 2016). The data to construct the land availability endowment measure (*WheatAgrLand*) over 2001-2015 are from the Food and Agriculture Organization (FAO) of the United Nations (wheat area harvested) and the World Development Indicators of the World Bank (agricultural land area).[7]

We use the two exogenous instrumental variables above to estimate the model (1); at this stage it is estimated without lagged dependent variable, though. We employ the two-stage instrumental variable estimation (FE TSLS) in the following way. In the first stage, we regress *GDPpc* and the rule-of-law index on the rest of explanatory variables along with two instrumental variables. This stage is formally captured as:

$$\ln(GDPpc_{it}) = \beta_1 HC_{it} + \beta_2 \ln\left(POP_{it}\right) + \beta_3 \ln(AREA_{it}) + \beta_4 ComPI_{it} + \beta_5 WheatAgrLand_{it} + \alpha_i + u_{it}$$
(2a)

$$RofL_{it} = \beta_1 HC_{it} + \beta_2 \ln\left(POP_{it}\right) + \beta_3 \ln(AREA_{it}) + \beta_4 ComPI_{it} + \beta_5 WheatAgrLand_{it} + \beta_i + z_{it} \ \ (2b)$$

In the second stage, we regress export sophistication on the set of our explanatory variables. However, the *GDPpc* and the rule-of-law index are replaced by their estimated values from the first stage ($GDPpc^{est}$, $RofL^{est}$). The second stage is formally captured as:

$$\ln(EXPY_{it}) = \beta_1 \ln(GDPpc_{it}^{est}) + \beta_2 HC_{it} + \beta_3 RofL_{it}^{est} + \beta_4 \ln(POP_{it}) + \beta_5 \ln(AREA_{it}) + \gamma_i + e_{it} \ \ (3)$$

All variables in specifications (2) and (3) are defined in the same way as in specification (1) plus *ComPI* is the country-specific international commodity export price index and *WheatAgrLand* is the ratio of the wheat growing land to the total agricultural land area. Subscript *i* denotes countries and subscript *t* denotes time periods (years). Superscript *est* denotes estimated variables based on the two IVs.

The first stage TSLS results are reported in columns 2 and 3 of the Table 2. The results show that the instrumental variables *ComPI* and *WheatAgrLand* have positive and significant effects on both *GDPpc* and *RofL*. The coefficients of the rest of explanatory variables are also statistically significant.

In column 4 of Table 2 we present the second stage results where we use the predicted values of the $GDPpc^{est}$ and $RofL^{est}$ to estimate the export sophistication model. We see that output and size (*GDPpc* and *POP*) exhibit positive and significant effects on export

---

[7] Easterly (2007) and Bennett and Nikolaev (2016) provide details on constructing the ratio of the land endowments suitable to grow wheat relative to sugarcane. Since sugarcane data are unavailable or impractical for number of countries in our dataset, we modify the endowment ratio to measure the proportion of the wheat area harvested endowment to the total agricultural land in a country.

sophistication (*EXPY*), while quality of institutions (*RofL*) has negative and significant effect. Other explanatory variables (human capital (*HC*) and geographical size (*AREA*)) are statistically insignificant. The results of the output and size are intuitively sensible. The effect of the institutional quality requires additional assessment that we perform later.

In terms of the estimation validity, based on the Kliebergen-Paap LM statistics (reported in Table 2) we reject the null hypothesis of under-identification and conclude that our model is just identified. Further, the Cragg-Donald and Kliebergen-Paap weak instruments identification test values exceed the 10% maximal IV size critical value, which indicates that inputed distortion in the t-test size is less than 10%, and therefore our two exogenous variables (*ComPI*, *WheatAgrLand*) are valid instruments. Thus, using two external exogenous regressors we were able to overcome possible reverse causality problem and put forth the evidence that correlates with earlier research.

5.2 Results of the GMM2 estimation

In this section we account for the path-dependency in export sophistication and estimate full specification (1). Inclusion of lagged dependent variable in the model may involve autocorrelation of the disturbances in panel estimation and some explanatory variables may become predetermined and not strictly exogenous. Hence, to estimate the model (1), we employ a broader set of instruments: (i) the external exogenous IVs introduced in Section 5.1, and (ii) lagged value of the dependent variable and lagged values of the endogenous regressors (*GDPpc* and *RofL*).

Our results are based on the estimation of model (1) with the GMM2 estimator. The GMM2 estimator was selected as the most fitting algorithm via Monte Carlo simulations that we detail in the Technical Appendix. The model (1) estimation results are presented in Table 3. In terms of model adequacy, the Hansen test shows that for all estimated versions the instruments are jointly valid as the null hypothesis is not rejected; employed instruments are exogenous at 5% significance level. In addition, we provide difference-in-Hansen tests of exogeneity of instrument subsets. The null hypothesis that the subsets of applied instruments are exogenous is not rejected at 5% significance level. Thus, we conclude that there is no evidence of model misspecification.

Based on the result presented in Table 3 (column 2) we see that the signs of the coefficients of the estimated parameters correspond with our priors. Specifically, the lagged value of EXPY shows a statistically significant and positive effect on the current value of EXPY. The estimated coefficients of the GDP per capita and country size are positive and statistically significant as well. This is consistent with the findings of Hausmann et al. (2007), who also proxy country size with its 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; Schott, 2008; Hummels and Klenow, 2005). The coefficients of the human capital and land area are statistically insignificant.

From Table 3 we also detect that for the whole sample of countries (column 2) the effect of institutional quality is negative. We hypothesize that the result is because the whole sample contains many countries with high EXPY but a relatively small institutional quality score. Therefore, we investigate the effect of institutions on the export sophistication in greater detail. We divide the countries into two sub-samples based on the median value of the rule-of-law index (RofL). This step is grounded in the hypothesis of Mauro (1995) that weaker institutions are associated with slower growth or worse economic performance and vice versa. Lower quality of institutions are intuitively associated mostly with emerging and developing economies. Hence, division into sub-samples is further supported by the fact that large proportion of developing economies (36 countries) belongs to the group with low institutional quality. Further, 15 emerging markets are in the group with low institutional quality level and only 5 of them (Chile, Hungary, Malaysia, Poland and South Africa) are in a group with high institutional level.

We re-estimated the model (1) separately for low and high institutional quality countries. The results are reported in Table 3 (columns 3 and 4). Based on the Hansen test of over-identification restrictions and difference-in-Hansesn test of exogeneity of instruments subsets we conclude that there is no evidence of model misspecification. Further, for the low institutional quality country group (Table 3, column 3), the coefficients of the lagged value of EXPY, GDP per capita, and population are statistically significant and positive, while the coefficients of human capital and land area are statistically insignificant. In addition, the rule-of-law index coefficient is statistically significant and negatively impacts the export sophistication in the sub-set of low institutional quality countries. We conjecture that low-quality of institutions is driving this result. The low institutional quality group includes mainly emerging markets and developing economies

where the level of formal institutions does not correspond to that in developed economies and changes only slowly (Bennett and Nikolaev, 2016). These might exhibit some inertia and resistance with respect to social and economic changes and negatively impact the export sophistication.

Finally, estimates for the high institutional quality group (Table 3 column 4) show that the values of EXPY (lagged), GDP per capita, and the size of the economy have statistically significant and positive effects on export sophistication. Thus, in accord with earlier results, we conclude that these three variables exhibit important impact on export sophistication both for low and high institutional quality countries. Human capital and land area are statistically insignificant. For the high institutional quality sub-sample the rule-of-law index is also statistically insignificant (Table 3, column 4), unlike in case of the low institutional quality countries (Table 3, column 3). We conjecture that high institutional quality countries have already reached a relatively high level of governance and further improvement of the rule-of-law might rather increase bureaucracy with a potentially negative effect on trade.

To conclude, based on the estimation results for the whole sample of all countries as well as for two sub-samples we conclude that output and country size positively impact export sophistication that also exhibits path dependence. Quality of institutions, or the lack of it, negatively impact export sophistication in those countries where the rule-of-law is below a sample median level. In those countries path dependence of export sophistication is also somewhat stronger than in countries with higher quality of institutions.

5.3 Robustness checks

In Section 5.2 we presented results from the model (1) where the key difference was the effect in low and high institutional quality countries. Therefore, we perform the following robustness check. The countries that record large GDP per capita but possess low level of institutional quality are the countries with a large ratio of crude oil exports to their total exports. Thus, in the spirit of Weldemicael (2012), we exclude several countries with high proportion of the crude oil exports (more than 50%) and then re-estimate the model (1).[8] The results are presented in Table 3 (column 5). Comparison with our earlier results (column 2) shows their similarity in terms of

---

[8] According to this criterion the following countries are eliminated from the sample: Azerbaijan, Kazakhstan, Ecuador, Oman and Saudi Arabia. The adjusted sample contains 96 countries.

signs and statistical significance: path dependence in export sophistication is present; output and economy size exhibit positive effect; human capital and land area are statistically insignificant. Hence, this set of results is robust to the exclusion of major oil exporting countries from the whole sample. Statistically insignificant impact of the rule-of-law is the single difference in results between full sample and adjusted sample.

Second robustness check is based in employing a different sophistication measure than that proposed by Hausman et al. (2007). Following Lin et al (2017) we use the Export Similarity Index (ESI) that was formally introduced in Section 3.1. Again, we estimate the model (1) for the alternative ESI measure with the GMM2 estimator and we employ a broad set of instruments: (i) the external exogenous IVs introduced in Section 5.1 and further (ii) lagged values of the dependent variable and lagged values of the endogenous regressors (*GDPpc* and *RofL*). The results are presented in Table 4. In terms of model adequacy, the Hansen test shows that the instruments are jointly valid and the difference-in-Hansen tests support exogeneity of instrument subsets at 5% significance level. Thus, we conclude that there is no evidence of model misspecification.

Similarly as in Section 5.2 we estimated the model (1) for the full sample, then separately for low and high institutional quality countries, and we also excluded countries with high proportion of crude oil exports. Based on the result presented in Table 4 we see that the signs of the coefficients as well as their statistical significance correspond to those shown in Table 3 for estimations based on our key sophistication measure due to Hausman et al. (2007). We conclude that output and country size positively impact export sophistication measured by the alternative ESI indicator. The extent of the path dependence in export sophistication is also present and is even stronger than that found with the measure of Hausman et al. (2007). In countries with higher quality of institutions export sophistication path dependence is weaker than in countries with low level of the rule-of-law. Lack of quality of institutions negatively impacts export sophistication in low institutional quality countries and seems to drive the negative effect for the full sample.

Hence, comparison of the results presented in Tables 3 and 4 indicates that our findings are robust with respect to different types of the export sophistication measures employed for estimation.

6. Synthesis of results and conclusions

We analyze the determinants of export sophistication in a dynamic panel data estimation set-up on a large sample of 101 countries over 15 year period (2001-2015). Our set contains 20 emerging markets, 49 developing countries, and 32 developed economies. We also perform Monte Carlo simulations to select an estimator best fitting our estimation strategy and single out the Arellano-Bond two-step GMM (GMM2) estimator. We also perform several robustness checks and show that our results are robust with respect to alternative export sophistication measure as well as to modifications of the country sample composition.

We show that output and country size (population) positively impact export sophistication. From this point of view, our results are consistent with the findings of Hausmann et al. (2007) that are, however, obtained from a cross-section assessment. On the other hand, human capital and land area (territorial size) do not exhibit statistically significant effects.

Quality of institutions, or the lack of it, negatively impact export sophistication. From our detailed assessment we see that the negative impact is associated with those countries where the rule-of-law is below a sample median level. Since export sophistication index is, by its construction, only indirectly linked to economic growth and performance, we can only indirectly take the above result as support of the Mauro (1995) hypothesis that weaker institutions are associated with slow growth or poor economic performance and vice versa.

When we modify the export sophistication model with a lagged value of the dependent variable we obtain a significant and positive effect. This finding suggests a strong persistence of the behavior of export sophistication. This is a sign that there exists not only path dependence in aggregate export activities (documented by Egger and Pfaffermayr, 2011) but also path dependence in terms of export sophistication itself. The fact that export sophistication exhibits strong path dependence is also likely the reason for the absence of structural breaks in export sophistication in our sample of countries. This result is also in line with empirical evidence that during the global financial crisis the qualitative structure of exports did not change and remained 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.

Finally, our results confirm (earlier documented) importance of the country's output and size on export sophistication and bring new evidence on its path-dependent property. Our findings further accentuate the importance to improve and cultivate quality of formal institutions, especially in emerging and developing economies where weak institutions negatively affect export sophistication.

References

Anand, R., S. Mishra, and N. Spatafora. 2012. Structural Transformation and the Sophistication of Production. Working Paper No. 12/59, International Monetary Fund, Washington, DC.

Anderson, T., and C. Hsiao. 1981. Estimation of dynamic models with error components. Journal of the American Statistical Association 76: 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 58 (2): 277-297.

Baum, C., M. Schaffer, and S. Stillman. 2007. Enhanced routines for instrumental variables/GMM estimation and testing. Stata Journal 7 (4): 465-506.

Barro, R. and J. Lee. 2000. International dataset on educational attainment. Available online at: http://www.cid.harvard.edu/ciddata/ciddata.html.

Bennett, D., and B. Nikolaev. 2016. Factor endowments, the rule of law and structural inequality. Journal of Institutional Economics 12 (4): 773-795.

Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87 (1): 115-143.

Bun, M., and M. Carree. 2005. Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 23 (2): 200-210.

Bruckner, M. 2012. Economic growth, size of the agricultural sector and urbanization in Africa. Journal of Urban Economics 71 (1): 26-36.

Cabral, M., and P. Veiga. 2010. Determinants of Export Diversification and Sophistication in Sub-Saharan Africa. Social Science Research Network, FEUNL Working Paper Series N. 550.

Da Costa Neto, N., and R. Romeu. 2011. Did Export Diversification Soften the Impact of the Global Financial Crisis? Working Paper No. 11/99, International Monetary Fund, Washington, DC.

Durlauf, S. 1994. Path Dependence in Aggregate Output. Industrial and Corporate Change 1: 149-172.

Easterly, W. 2007. Inequality Does Cause Underdevelopment: Insights from a New Instrument. Journal of Development Economics 84 (2): 755–776.

Egger, P., and M. Pfaffermayr. 2011. Structural Estimation of Gravity Models with Path-dependent Market Entry. CEPR Discussion Paper 8458, London: Centre for Economic Policy Research.

Felipe, J., U. Kumar, and A. Abdon. 2014. How rich countries became rich and why poor countries remain poor: It's the economic structure... duh!. Japan and the World Economy 29: 46–58.

Finger J., and M. Kreinin. 1979. A measure of 'export similarity' and its possible uses. Economic Journal 89 (356): 905-912.

Fortunato, P., and C. Razo. 2014. Export sophistication, growth and the middle-income trap. Transforming Economies: Making Industrial Policy Work for Growth, Jobs and Development. ILO: Geneva: 267-287.

Görg, H., and E. Strobl. 2001. Multinational companies and productivity spillovers: A meta-analysis with a test for publication bias. Economic Journal 111 (475): 723-739.

Greenaway, D., N. Sousa, and K. Wakelin. 2004. Do domestic firms learn to export from multinationals? European Journal of Political Economy 20 (4): 1027–1044.

Hansen, G. 2001. A bias-corrected least squares estimator of dynamic panel models. Allgemeines Statistisches Archiv 85: 127-140.

Hanson, G. 2012. The rise of middle kingdoms: Emerging economies in global trade. The Journal of Economic Perspectives 26 (2): 41-64.

Hausmann, R., J. Hwang, and D. Rodric. 2007. What you export matters. Journal of Economic Growth 12 (1): 1-25.

Hu, C., F. Lin, and X. Wang. 2016. Learning from exporting in China: A firm-specific instrumental approach. Economics of Transition 24 (2), 299-334.

Hummels, D. and P. Klenow. 2005. The variate and quality of a Nation's export. American Economic Review 95 (3): 704-723.

Kiviet, J. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68 (1): 53-78.

Krugman, P. 1980. Scale economies, product differentiation, and the pattern of trade. American Economic Review 70 (5): 950-959.

Lin, F. 2015. Learning by exporting effect in China revisited: An instrumental Approach. China Economic Review 36: 1-13.

Lin, F., and N. Sim. 2013. Trade, income and Baltic Dry Index. European Economic Review 59: 1–18.

Lin, F., E. Weldemicael, and X. Wang. 2017. Export sophistication increases income in Sub-Saharan Africa: Evidence from 1981-2000. Empirical Economics 52 (4): 1627-1649.

Melitz, M. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6): 1695-1725.

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

Nickel, S. 1981. Biases in Dynamic models with fixed effects. Econometrica 49 (6): 1417-1426.

Nunn, N. 2009. The Importance of History for Economic Development. Annual Review of Economics 1 (1): 65-92.

Reinstaller, A., and P. Reschenhofer. 2015. Path dependence in national systems of production and "self discovery" of environmental technologies in the EU 28 countries. Working Papers series no. 106, WWWforEurope.

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

Schott, P. 2008. The relative sophistication of Chines exports. Economic Policy 53: 5-49.

Shelburne, R. 2010. The global financial crisis and its impact on trade: The world and the European emerging economies. United Nations Economic Commission for Europe, Discussion Paper Series No. 2010.2.

Stock, J., and M. Yogo. 2002. Testing for Weak Instruments in Linear IV Regression. Working Paper No. 284, The National Bureau of Economic Research.

Teece, D. J., G. Pisano, and A. Shuen. 1997. Dynamic capabilities and strategic management. Strategic Management Journal 18 (7): 509–533.

Weldemicael, E. 2012. Determinants of Export Sophistication. The University of Melbourne, mimeo.

Zhu, S., X. Fu, M. Lai, and J. Xuan. 2009. What drives the export sophistication of countries? SLPTMD Working Paper Series No. 033, University of Oxford.

Table 1. Descriptive statistics for EXPY (USD)

Year	Obs.	Mean	SD	Minimum	Maximum	Range
1	2	3	4	5	6	7
2001	101	13756.4	4864.0	3022.4	25896.7	22874.3
2002	101	14295.2	5080.5	2941.3	27965.3	25024.0
2003	101	14766.7	5221.0	2792.7	25733.2	22940.5
2004	101	15774.7	5543.5	3982.9	28105.3	24122.5
2005	101	16591.6	5543.5	4005.8	27537.2	23531.4
2006	101	18090.2	5644.1	4487.9	29551.2	25063.3
2007	101	19316.0	5987.0	5480.5	31398.5	25918.1
2008	101	20227.6	6347.1	4274.9	33774.8	29499.9
2009	101	19780.8	6092.3	4150.7	34981.4	30830.7
2010	101	20544.4	6230.2	5335.3	36170.2	30834.9
2011	101	21546.2	6609.5	6636.6	41152.7	34516.0
2012	101	22120.2	6507.3	5129.9	41522.7	36392.8
2013	101	22871.1	6536.5	5942.3	42391.4	36449.0
2014	101	23408.1	6579.5	8803.0	45414.9	36611.9
2015	101	23827.3	6996.4	8405.8	47636.8	39231.0

Dependent variable	First stage		Second stage
	Ln(GDPpc)	RofL	Ln(EXPY)
(1)	(2)	(3)	(4)
ComPI$_{it}$	0.342***	0.079***	--
	(0.068)	(0.023)	--
WheatAgrLand$_{it}$	0.034***	0.034***	--
	(0.004)	(0.004)	--
HC$_{it}$	0.015***	0.002***	0.007
	(0.001)	(0.001)	(0.012)
Ln(POP$_{it}$)	0.619***	0.025***	0.790***
	(0.072)	(0.066)	(0.095)
ln(AREA$_{it}$)	-5.063***	-3.203***	-3.527
	(1.076)	(1.344)	(3.156)
Ln(GDPpc$_{it}$)^{est}	--	--	0.419***
	--	--	(0.066)
RofL$_{it}$^{est}	--	--	-0.266***
	--	--	(0.102)
Observations	1515	1515	1515
Number of countries	101	101	101
Number of periods	15	15	15
Number of regressors	4	4	5
Number of endogenous regressors	1	1	2
Number of instruments	5	5	5
Number of excluded instruments	2	2	2
Kleibergen-Paap LM test (p-value)	0.00	0.00	0.00
Cragg-Donald Wald F stat (excluded inst.)	150.67	80.56	44.96
Kleibergen-Paap Wald stat (excluded inst.)	60.48	51.49	19.41
Stock and Yogo critical values:			
10 % maximal IV size	19.93	19.93	7.03
15 % maximal IV size	11.59	11.59	4.58
20 % maximal IV size	8.75	8.75	3.95
25 % maximal IV size	7.25	7.25	3.63

Note: Robust standard errors are reported in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. *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. Ln(GDPpc)^{est} and RofL^{est} are predicted values of Ln(GDPpc) and RofL from the first stage. ComPI is international commodity price index for agricultural and natural resources.WheatAgrLand represents ratio of the wheat harvested land to the agricultural land. We apply critical values obtained from Stock and Yogo (2005). Although critical values do not exist for the Kliebergen-Paap statistic, we follow the approach suggested in Baum et al. (2007) and apply the Stock and Yogo (2005) critical values initially tabulated for the Cragg-Donald statistic. The null hyphothesis of the Kliebergen-Paap LM test is that the structural equation is under-identified (i.e. the rank condition fails). In columns (2) and (3) we report a single endogenous regressors first step regression and their corresponding test statistics. For single endogenous regressors the Cragg-Donald and Kliebergen-Paap statistics reduce to the standard non-robust F statistics and heteroscedastisity-robust first stage F statistics, respectively.

Table 3. Estimation results when ln(GDPpc) and RofL are predetermined variables (Dependent variable Ln(EXPY))

Variables	Whole sample GMM2	Low institutional quality GMM2	High institutional quality GMM2	Excluding major oil exporting countries GMM2
(1)	(2)	(3)	(4)	(5)
$Ln(EXPY)_{it-1}$	0.102***	0.245***	0.180***	0.104***
	(0.032)	(0.027)	(0.009)	(0.008)
$Ln(GDPpc)_{it}$	0.719***	0.489***	0.758***	0.691***
	(0.104)	(0.131)	(0.155)	(0.099)
HC_{it}	0.000	-0.001	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
$RofL_{it}$	-0.181***	-0.120*	-0.137	-0.120
	(0.058)	(0.094)	(0.082)	(0.103)
$Ln(POP)_{it}$	0.327***	0.664***	0.317***	0.467***
	(0.073)	(0.230)	(0.118)	(0.160)
$Ln(AREA)_{it}$	-0.805	-21.798	-3.372	-0.266
	(1.348)	(19.657)	(2.481)	(2.802)
No. of obs.	1313	663	650	1248
No. of groups	101	51	50	96
Number of periods	13	13	13	13
No. of instruments	44	44	44	82
AB test for AR(2) in first diff. (p-value)	0.650	0.882	0.621	0.595
Hansen test of overidentification restrictions (p-value)	0.060	0.233	0.155	0.106
Diff.-in-Hansen tests of exogeneity of inst. subsets:				
Hansen test excluding group (p-value)	0.066	0.226	0.071	0.067
Difference (null H = exogenous) (p-value)	0.266	0.388	0.946	0.834

Note: The Windmeijer corrected standard errors are reported in parentheses. ***, **, and * indicate statistical significance 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 this model we assume that *Ln(GDPpc)* and *RofL* are predetermined variables, while *HC, Ln(POP)* and *Ln(AREA)* and two external variables *ComPI* and *WheatAgrLand* are strictly exogenous.

The set of instruments used for GMM2:

GMM-type: $Ln(EXPY)_{i,t-2}, Ln(GDPpc_{i,t-2}), Ln(GDPpc_{i,t-1}), RofL_{it-2}, RofL_{it-1}$ up to 1 lag.

Standard: $\Delta HC_{it}, \Delta Ln(POP_{it}), \Delta Ln(AREA_{it}), \Delta ComPI_{it}, \Delta WheatAgrLand_{it},$

Table 4. Estimation results when ln(GDPpc) and RofL are predetermined variables (Dependent variable ESI)

Variables	Whole sample GMM2	Low institutional quality GMM2	High institutional quality GMM2	Excluding major oil exporting countries GMM2
(1)	(2)	(3)	(4)	(5)
ESI_{it-1}	0.623***	0.706***	0.143***	0.592***
	(0.061)	(0.077)	(0.021)	(0.067)
$Ln(GDPpc)_{it}$	0.124***	0.094***	0.303***	0.136***
	(0.022)	(0.026)	(0.046)	(0.025)
HC_{it}	0.000	-0.001	0.000	0.000
	(0.000)	(0.001)	(0.000)	(0.000)
$RofL_{it}$	-0.076**	-0.052*	-0.023	-0.064
	(0.029)	(0.029)	(0.036)	(0.042)
$Ln(POP)_{it}$	0.061***	0.082***	0.072**	0.071***
	(0.022)	(0.019)	(0.031)	(0.026)
$Ln(AREA)_{it}$	0.049	-5.714	0.701	-0.144
	(0.593)	(4.778)	(0.984	(0.647)
No. of obs.	1313	663	650	1248
No. of groups	101	51	50	96
No. of periods	13	13	13	13
No. of instruments	82	44	44	82
AB test for AR(2) in first diff. (p-value)	0.408	0.729	0.292	0.415
Hansen test of overidentification restrictions (p-value)	0.119	0.191	0.227	0.161
Difference-in-Hansen tests of exogeneity of instrument subsets:	--	--	--	--
Hansen test excluding group (p-value)	0.101	0.221	0.126	0.138
Difference (null H = exogenous) (p-value)	0.520	0.261	0.882	0.534

Note: The Windmeijer corrected standard errors are reported in parentheses. ***, **, and * indicate statistical significance 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 ESI refers to the export similarity index, 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 this model we assume that *Ln(GDPpc)* and *RofL* are predetermined variables, while *HC, Ln(POP)* and *Ln(AREA)* and two external variables *ComPI* and *WheatAgrLand* are strictly exogenous.

The set of instruments used for GMM2 are:

GMM-type: $Ln(ESI)_{i,t-2}, Ln(GDPpc_{i,t-2}), Ln(GDPpc_{i,t-1}), RofL_{t-2}, RofL_{t-1}$ up to 1 lag.

Standard: $\Delta HC_{it}, \Delta Ln(POP_{it}), \Delta Ln(AREA_{it}), \Delta ComPI_{it}, \Delta WheatAgrLand_{it}$

Technical Appendix

In this appendix we present the main steps of the Monte-Carlo experimental design. The purpose of the Monte Carlo simulations is to assess performance of various estimators. Further, we present the results, based on which to choose an estimator that is best suited for our data set and provides most accurate estimates for the export sophistication model. All simulations were carried out using estimation routines written in the MATLAB (2013a) package.[9]

Out of a number of estimators that have been developed we consider a static panel data estimation algorithm with fixed effects (FE) and various dynamic panel data estimation algorithms, in particular, Anderson-Hsiao instrumental variables estimators with level and difference instruments (AH-L and AH-D), Arellano-Bond two-step GMM (GMM2), and Blundell-Bond two-step system GMM (SYS-GMM2) estimators.

We assume that our data generation process (DGP) closely follows the model estimated in the section 5:

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

$$\eta_i \sim N(0, \sigma_\eta^2), \ v_{it} \sim N(0, \sigma_v^2).$$

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 S1 we summarize the Monte Carlo design and present the parameter values of the data-generating process. As we can see from Table S1, 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_v^2 and σ_η^2 separately, we focus on the variance ratio $\sigma_\eta^2 / \sigma_v^2$ as in Santos and Barrios (2012). 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.

Taking into account that our experiments are based on the actual data, first we estimate a preliminary static fixed-effects model and then calculate the estimated residual variance. However, the estimated residual variance might change when different values of β_1 are applied to the actual dataset. Therefore, it is sensible to allow for a change in the residual variance

---

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

together with changing value of the lagged parameter. Otherwise, keeping the residual variance constant while increasing lagged parameter value would increase an estimate bias. Thus, having a distribution for v_{it} as $v_{it} \sim N(0, \sigma_v^2)$, we are able to generate v_{it} residuals of size $N \times T$ based on actual values of σ_v^2. 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 $\sigma_\eta^2 / \sigma_v^2$ from Table S1. Then, given the data-generating process and the values of the regressors, we generate a set of possible 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 coefficient β_1 we perform 1000 Monte Carlo replications in such a way that we create 1000 panel datasets for each 20 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:

$$BIAS(\beta_1) = \frac{1}{R} \sum_{r=1}^{R} \left(\hat{\beta}_1^r - \beta_1^{true} \right)$$

and

$$RMSE(\beta_1) = \sqrt{\frac{1}{R} \sum_{r=1}^{R} \left(\hat{\beta}_1^r - \beta_1^{true} \right)^2},$$

where r denotes the number of replications.

Monte Carlo results are presented in Table S2. We start with the case when the value of the ratio $\sigma_\eta^2 / \sigma_v^2$ is equal to zero. As we can see from Table S2, when the value of parameter β1 fluctuates around 0.1–0.9, then the lowest bias and RMSE are achieved in the case of the GMM2 estimator. Then we increase the value of the ratio $\sigma_\eta^2 / \sigma_v^2$ up to 1.0 and again allow the true values of parameter β1 to vary between 0.1 and 0.9. Again, from Table S2 we see that the lowest bias and RMSE is again achieved in the case of the GMM2 estimator. However, when we

compare these results with the previous ones (when $\sigma_\eta^2 / \sigma_\nu^2 = 0.0$), we see that the estimates obtained by the GMM2 and GMM2-SYS estimators become less biased than the estimates obtained by FE. When we relate previous results to the AH-L and AH-D estimates (Table S2), we can see that the bias is about the same when compared with the same results for $\sigma_\eta^2 / \sigma_\nu^2 = 0.0$. But the starting values of the bias are so large that the AH-L and AH-D estimator is not able to compete with the GMM2 and GMM2-SYS estimators' corresponding results. That is why we do not take into account the AH-L and AH-D estimator results in our subsequent explanations. We again continue to increase the variance of the individual effect of the countries up to 3.0 and then up to 5.0. From Table S2 we see that all previous conclusions still hold. The one difference is that when the true values of parameter β₁ equal to 0.9, then the bias obtained with GMM2 and GMM2-SYS becomes smaller than the bias obtained with FE. This means that with further increase of the variance of the individual effect of the countries, the behavior of the GMM2 and SYS-GMM2 estimators further improves.

Hence, based on the our experiment results reported in Table S2 we conclude that GMM-type estimators (GMM2 and SYS-GMM2) 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.

Finally, based on the comprehensive assessment of the results presented in Table S2 we chose the GMM2 estimator to estimate model (1) to obtain the estimation results on which we base our inferences.

Table S1. Monte Carlo designs

$\sigma_\eta^2 / \sigma_\nu^2$	N	T	β_1	β_2	β_3	β_4	β_5	β_6
0.0	101	15	0.10	0.30	0.05	0.03	0.15	0.05
0.0	101	15	0.50	0.30	0.05	0.03	0.15	0.05
0.0	101	15	0.90	0.30	0.05	0.03	0.15	0.05
1.0	101	15	0.10	0.30	0.05	0.03	0.15	0.05
1.0	101	15	0.50	0.30	0.05	0.03	0.15	0.05
1.0	101	15	0.90	0.30	0.05	0.03	0.15	0.05
3.0	101	15	0.10	0.30	0.05	0.03	0.15	0.05
3.0	101	15	0.50	0.30	0.05	0.03	0.15	0.05
3.0	101	15	0.90	0.30	0.05	0.03	0.15	0.05
5.0	101	15	0.10	0.30	0.05	0.03	0.15	0.05
5.0	101	15	0.50	0.30	0.05	0.03	0.15	0.05
5.0	101	15	0.90	0.30	0.05	0.03	0.15	0.05

Note: N denotes number of countries, T denotes number of years. For each possible value of $\sigma_\eta^2 / \sigma_\nu^2$ we have three scenarios for parameter β_1. Thus, in total we have 12 possible scenarios for β_1.

Table S2. Simulation results, T = 15, N = 101

Variance ratio	Parameter true values	FE		AH-L		AH-D		GMM2		SYS-GMM2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$\sigma_\eta^2 / \sigma_\nu^2 = 0.0$	$\beta_1 = 0.10$	-0.040	0.050	-0.084	0.093	-0.110	0.130	0.032	0.045	-0.017	0.036
	$\beta_1 = 0.50$	-0.091	0.094	-0.394	0.397	-0.539	0.544	-0.014	0.028	-0.120	0.123
	$\beta_1 = 0.90$	-0.057	0.058	-0.256	0.261	-1.227	1.231	-0.024	0.026	-0.066	0.067
$\sigma_\eta^2 / \sigma_\nu^2 = 1.0$	$\beta_1 = 0.10$	-0.041	0.049	-0.084	0.094	-0.110	0.131	0.046	0.057	0.007	0.034
	$\beta_1 = 0.50$	-0.105	0.108	-0.394	0.396	-0.541	0.545	0.029	0.039	-0.079	0.084
	$\beta_1 = 0.90$	-0.080	0.081	-0.259	0.268	-1.235	1.239	0.003	0.009	-0.035	0.036
$\sigma_\eta^2 / \sigma_\nu^2 = 3.0$	$\beta_1 = 0.10$	-0.046	0.053	-0.082	0.092	-0.108	0.128	0.051	0.062	0.058	0.068
	$\beta_1 = 0.50$	-0.134	0.136	-0.396	0.400	-0.540	0.545	0.039	0.047	-0.032	0.045
	$\beta_1 = 0.90$	-0.123	0.124	-0.258	0.274	-1.232	1.236	0.005	0.010	-0.023	0.025
$\sigma_\eta^2 / \sigma_\nu^2 = 5.0$	$\beta_1 = 0.10$	-0.049	0.056	-0.082	0.094	-0.107	0.128	0.053	0.064	0.107	0.113
	$\beta_1 = 0.50$	-0.159	0.161	-0.396	0.401	-0.543	0.549	0.042	0.050	0.008	0.033
	$\beta_1 = 0.90$	-0.163	0.163	-0.265	0.292	-1.233	1.239	0.006	0.011	-0.015	0.019

Note: N denotes number of countries, T denotes number of years. The following abbreviations are used to label estimators used in the Monte Carlo simulations. Static panel data estimation algorithm with fixed effect (FE); Anderson-Hsiao instrumental variables estimators with level and difference instruments (AH-L and AH-D); Arellano-Bond two-step GMM estimator (GMM2); Blundell-Bond two-step system GMM estimator (SYS-GMM2).